JSM 2004 ABSTRACTS

Toronto, Ontario, Canada
August 8–12

sponsored by:
American Statistical Association
Institute of Mathematical Statistics
International Biometrics Society
Eastern North American Region
Western North American Region
Statistical Society of Canada

JOINT STATISTICAL MEETINGS

TORONTO, ONTARIO, CANADA

2004

Advanced Statistical Software for Efficient, Accurate and Effective Problem Solving

Essential Software Tools for Statisticians

StatXact®
StatXact® PROCs for SAS® Users

- Over 130 Exact Procedures Covering the Most Frequently Used Statistical Tests, Confidence Intervals and Sample Size Requirements
- Stratified 2x2, 2xC and Stratified RxC Tables
- Exact Tests for Superiority, Equivalence and Non-Inferiority of Paired Binomial Data
- Cytel Studio: Summary Statistics, T-Tests, ANOVA, and Linear Regression, Expanded Case and Table Editors

East®

- Design, Simulation and Interim Monitoring of Group Sequential and Adaptive Clinical Trials
- New Families of Spending Functions Allow for Futility-Only Based Trials
- Data Can Be Viewed on Multiple Scales: P-Value, Z-Scale, Mean and Others

LogXact®
LogXact® PROCs for SAS® Users

- Exact Logistic Regression
- Powerful Monte Carlo Procedures Enable Fast Exact Inference for Much Larger Data Sets
- LogXact PROCs Work seamlessly in Your SAS Environment
- Exact Inference for Poisson Regression

ToxTools®

- Software for Dose Response Analysis, Benchmark Dose Estimation and Risk Assessment

Egret®

- Modeling Tools for Biomedical and Epidemiological Case-control Studies

Call today to learn more about the Cytel solution.

Do Advanced Exact Analyses and Basic Statistics in One Place Instead of Using Several Software Packages

Distributed with each purchase of StatXact and LogXact, Cytel Studio consists of all the features that StatXact and LogXact share:

- the data editors
- the script language
- graphics capabilities
- basic regression
- ANOVA and T-Tests

Further enhancements, by customer request, are in developments such as more data manipulation capabilities and more graphics. Stop by for a sneak peak.

Visit JSM booths #114, #116, #118!

Cytel, StatXact, StatXact PROCs, LogXact, LogXact PROCs, East, and Egret are trademarks or registered trademarks of Cytel Software Corporation. All other company and product names are the property of their respective owners. Copyright © 2004 Cytel Software Corporation. All rights reserved.

Cytel
STATISTICAL SOFTWARE

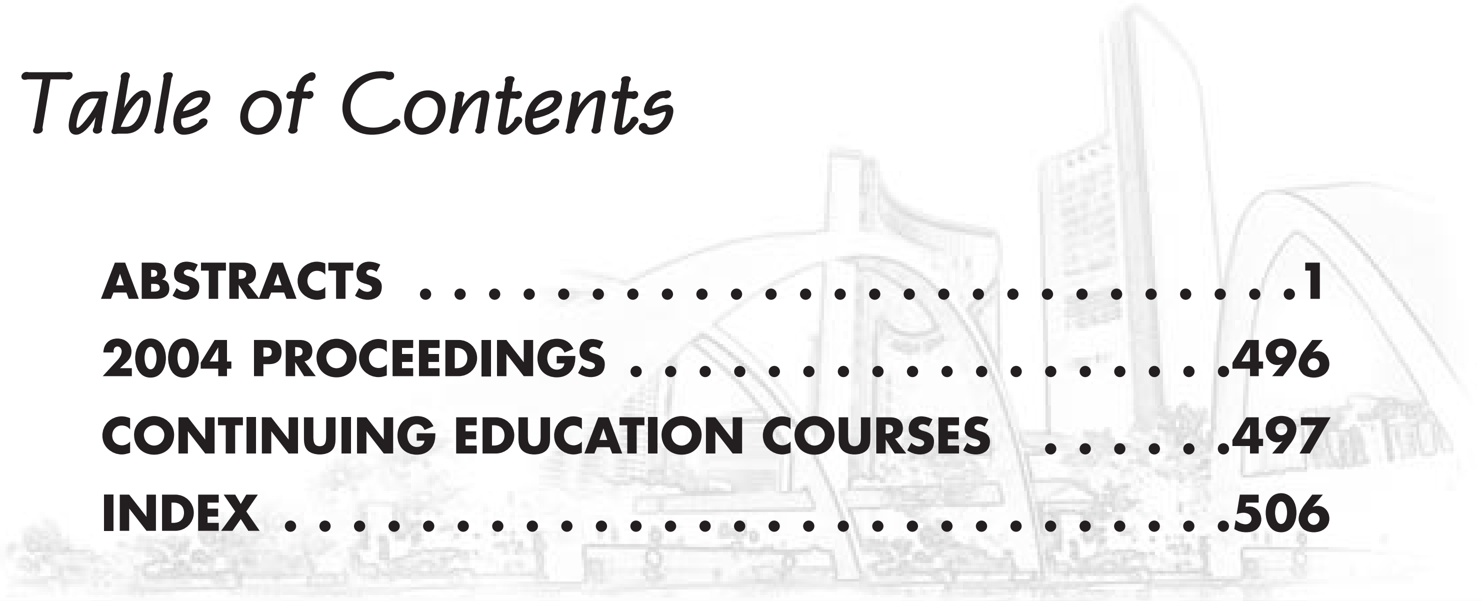
www.cytel.com
sales@cytel.com
T: +1.617.661.2011

The

2004
Joint Statistical Meetings

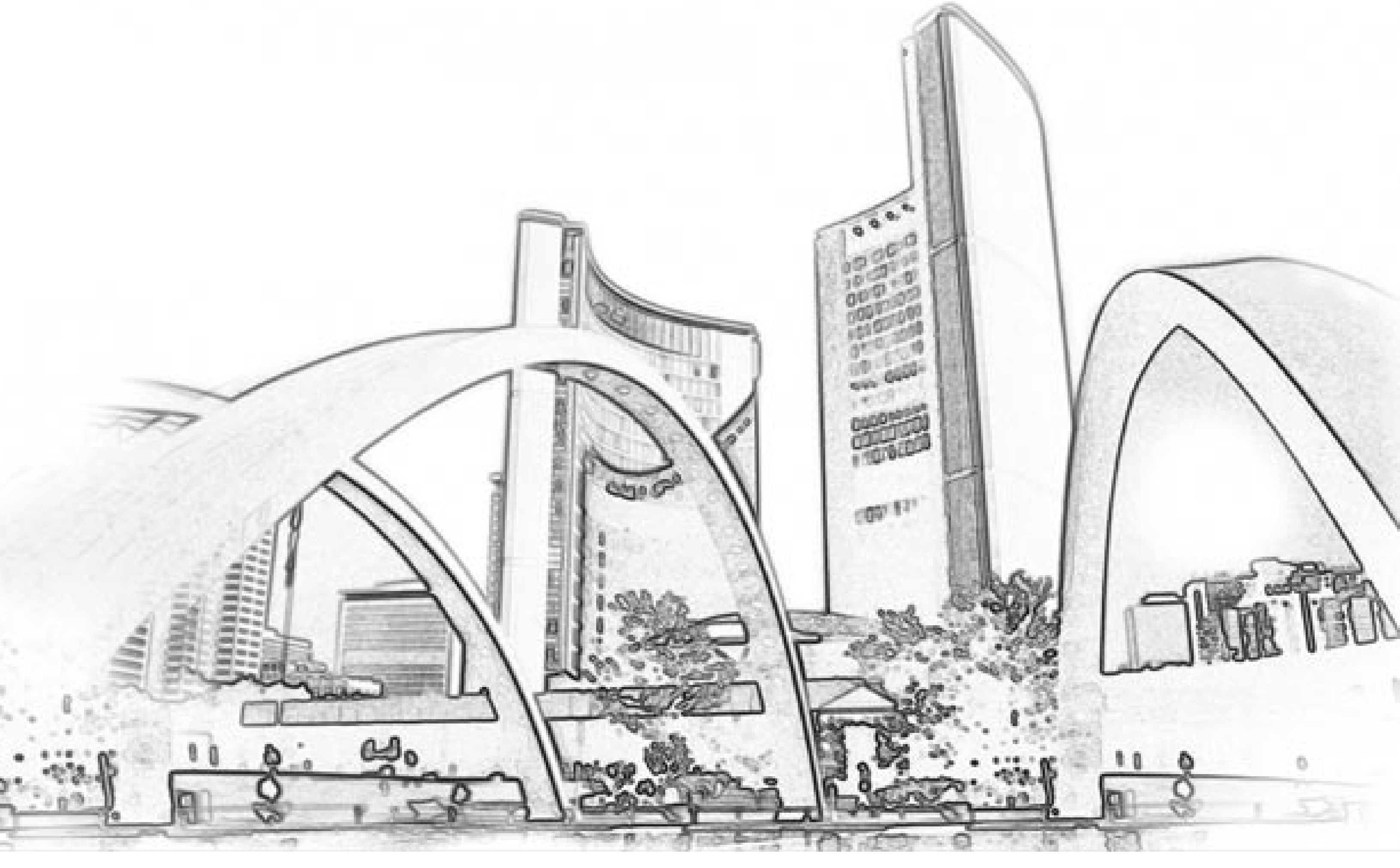
August 8–12, 2004
Toronto, Ontario, Canada

Table of Contents



STATISTICS AS A UNIFIED DISCIPLINE
AUGUST 8-12
JSM
TORONTO
2004

SESSIONS



1 Bridging the Gap: Issues and Methods for Transitioning Between the 1977 and 1997 OMB Standards for Collecting Data on Race and Ethnicity ◬

Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section

Sunday, August 8, 2:00 pm–3:50 pm

Overview of Race-bridging Models for 2000 and Development of Revised Models for Post-censal Years

◆ Jennifer Parker, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 6415,
Hyattsville, MD 20782

jdparker@cdc.gov

Key Words: *population estimates*

Early in 2003, the National Center for Health Statistics (NCHS) released population files for 2000 and 2001, based on the Census Modified Race Summary File, where multiple race counts on the Census file were distributed amongst the single-race counts. The objective of this talk is to describe the initial models used to distribute the multiple race counts, discuss issues associated with the original models, and present preliminary findings from ongoing work that may be used for future population counts. The National Health Interview Survey (NHIS) has allowed multiple race responses for many years and asks multiple race respondents to identify one of their responses as a primary race. Using the multiple-race responses in the NHIS, the primary race responses were used as the outcome variable in categorical regression models; individual- and county-level variables were included as covariates. Models and predictions differed markedly between multiple-race groups. Revisions to these models will include additional years of NHIS and new county-level variables from the 2000 Census.

Replicating the Race-bridging Models Using the Census Quality Survey

◆ Deborah D. Ingram, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., MS6208,
Hyattsville, MD 20782

ddingram@cdc.gov

Key Words: *bridging models, vital rates, logistic regression*

The Census Quality Survey to Evaluate Responses to the Census 2000 Question on Race (CQS) is used to evaluate the bridging models developed by NCHS to bridge the Census 2000 multiple-race population to single-race categories (American Indian or Alaska Native, Asian or Pacific Islander, Black, and White). The Census Quality Survey used a split panel design to collect race data using both the 1990 census race question and the Census 2000 race question from about 50,000 households. To assess the robustness of the bridging models obtained using the 1997-2000 NHIS, the models are fit to the CQS data. The resulting parameter estimates are compared to those obtained using the NHIS data. Further, county-age-sex-Hispanic origin-specific probabilities of selecting each of the possible single-race categories are obtained from the CQS parameter estimates and applied to the Census 2000 Summary Data file to obtain bridged-race population counts. These population counts are compared to those obtained using the NHIS bridging proportions to determine how close the agreement is. The impact of using additional Census 2000 contextual variables (i.e. segregation indexes) is examined.

Assessing Variability Due to Race Bridging

◆ Nathaniel Schenker, National Center for Health Statistics

National Center for Health Statistics, CDC, 3311 Toledo Rd.,
Hyattsville, MD 20782

nschenker@cdc.gov

Key Words: *Bayesian, missing data, multiple imputation, Taylor series, vital statistics*

Whereas census counts are traditionally treated as nonrandom population quantities in many applications, bridged census counts have random variability, because they are estimates. Techniques for assessing variability due to race bridging are discussed. Methods developed by Schafer and Schenker for inference with imputed conditional means, which can be considered a first-order approximation to multiple imputation, are adapted to the bridging problem and applied to bridged 2000 census counts as well as to selected vital rates for 2000 computed using bridged census counts as denominators. The relative standard errors of the bridged census counts by race under the 1977 standards tend to be higher for finer geographic levels and lower for coarser geographic levels. For each state (or the District of Columbia), the relative standard error of the count for a given race is no greater than .05. For birth and death rates by age group and 1977 race at the national level, on an absolute basis, use of bridged counts in the denominators does not add substantially to the relative standard errors.

2 Analysis of Intensively Collected Data

WNAR, Section on Physical and Engineering Sciences, ENAR

Sunday, August 8, 2:00 pm–3:50 pm

Normalization and Significance Analysis of cDNA Microarrays

◆ Jianqing Fan, Princeton University

Princeton University, Dept. of Operations Res. & Fin. Eng.,
Princeton, NJ 08544

jqfan@princeton.edu

The quantitative comparison of two or more microarrays can reveal, for example, the distinct patterns of gene expression that define different cellular phenotypes or the genes that are induced in the cellular response to certain stimulations. Normalization of the measured intensities is a prerequisite of such comparisons. However, a fundamental problem in cDNA microarray analysis is the lack of a common standard to compare the expression levels of different samples. Several normalization protocols have been proposed to overcome the variabilities inherent in this technology. We have developed a normalization procedure based on within-array replications via a Semi-Linear In-slide Model (SLIM), which adjusts objectively experimental variations without making critical biological assumptions. The significant analysis of gene expressions is based on a newly developed weighted t statistic, which accounts for the heteroscedasticity of the observed log-ratios of expressions, and a balanced sign permutation test.

Functional Regression for Longitudinal Data

◆ Hans-Georg Mueller, University of California, Davis

University of California, Davis, Department of Statistics, One Shields Ave., Davis, CA 95616

mueller@wald.ucdavis.edu

Key Words: *functional data analysis, sparse data, smoothing*

We review functional regression models with emphasis on situations where the response is a random trajectory. In longitudinal data analysis, observation times of trajectories are typically irregular and sparse. We present nonparametric methods that are based on several local least squares smoothing steps and can handle such sparse data. By borrowing strength from the entire sample of observed curves, nonparametric estimation becomes feasible in situations with sparse data that are normally not conducive to applications of smoothing. In addition to asymptotic results including rates of convergence, we discuss the concept of a residual process that replaces the ordinary residuals in classical regression and is useful in studying goodness of fit for functional regression. The methods are illustrated with examples from biomedical studies. This presentation is based on joint work with Jeng-Min Chiou, Peter Hall, Jane-Ling Wang and Fang Yao.

Accounting for Response Correlation in Varying-coefficient Models

◆ Naisyin Wang, Texas A&M University; Jeng-Min Chiou, National Health Research Institute

Texas A&M University, Blocker 432, College Station, TX 77840-3143

nwang@stat.tamu.edu

Key Words: *asymptotic property, nonparametric estimation, longitudinal data analysis*

This paper considers a nonparametric varying coefficient model in a longitudinal data setting. An estimating procedure which accounts for the dependency of within subject responses is presented. The major gain is in the improvement of estimation efficiency. Asymptotic properties of this estimator are provided. The numerical efficacy of the proposed methodology is demonstrated through a small simulation study and the analysis of a real dataset.

Mixed-effects State-space Models for Longitudinal Data Analysis

◆ Hulin Wu, University of Rochester Medical Center; Dacheng Liu, University of Rochester Medical Center; Xu-Feng Niu, Florida State University

University of Rochester Medical Center, Dept. of Biostatistics and Computational Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

hwu@bst.rochester.edu

Key Words: *state-space model, longitudinal data, mixed-effects model, differential equations, Kalman filter, HIV/AIDS*

Recent development of nonparametric smoothing/regression methods for longitudinal data allows us to robustly analyze the longitudinal data with minimum assumptions on the model. However, another important direction has not been explored, which is how to efficiently analyze the longitudinal data when the mechanisms generating the data are known and can be described by complicated differential or difference equations. We propose a novel class of mixed-effects state-space models for longitudinal data. If the dynamics of individual observations can be formulated in a system of differential or difference equations, then state-space modeling is straightforward and appealing. Three methods are developed for estimating unknown parameters, i.e. the global two-stage (GTS) method, the EM-based maximum likelihood method (MLE), and the Bayesian approach. Simulation results indicate that all the three methods perform well. Finally, we apply the mixed-effects state-space model to a dataset from an AIDS clinical trial to illustrate the proposed methodology.

3 Training of Government Statisticians ⌘

Section on Statistical Education, International Statistical Institute, Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section, International Association for Statistical Education

Sunday, August 8, 2:00 pm–3:50 pm

Teaching a Course on Government Statistics in a University Statistics Department

◆ Patrick Murphy, University College Dublin

University College Dublin, Belfield, Dublin 4, Ireland

Patrick.Murphy@ucd.ie

Key Words: *government, statistics, education*

Most official statisticians around the world would receive their training “on the job,” or more formally by international organizations such as the Training of European Statisticians Institute in Europe or the International Monetary Fund in Washington, D.C. The Department of Statistics in University College, Dublin, Ireland, recently introduced a new course on official statistics as an integrated part of its curriculum. It was first delivered in the academic year 1999-2000 to a class of final year undergraduate, graduate diploma, and master's students. This paper will describe the environment in which the course was delivered, and will contain descriptions of the development of the course, and of the course itself. It will also illustrate how students in the course were assessed and what the students' views were regarding the course. We will point out some changes that were made when teaching the course for the second and third times. Finally, we will summarize our experiences with this course, provide some indications about where students who have taken this course are now employed, and outline our views for the future of the course.

The Joint Program in Survey Methodology and Its Impact on the Federal Statistical Agency Workforce

◆ Cynthia Z.F. Clark, U.S. Census Bureau; Roger Tourangeau, University of Maryland; Gia F. Donnalley, U.S. Census Bureau

U.S. Census Bureau, Room 2031-3, Washington , DC 20233-9000

cclark@census.gov

Key Words: *training, survey, government, master's, sampling*

The Joint Program in Survey Methodology, initiated in 1993, has had a major impact on the workforce at federal statistical agencies. This paper discusses history and structure of the program and impact of the program on the careers of its graduates. In particular, it examines how employees at one of these agencies—the U.S. Census Bureau—supported in the master's degree program, have contributed to survey methodological research. It explores the impact contributions from these employees on the official statistics programs conducted by the Census Bureau.

The Challenges of Providing International Statistical Training to Government Statisticians

◆ Elizabeth A. Taylor, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 2190, Washington, DC 20212

taylor.elizabeth@bls.gov

Key Words: *assessment of trainers in statistics, development program for trainers in statistics*

Providing statistical training to groups composed of government statisticians of different countries poses many challenges. The U.S. Bureau of Labor Statistics has been conducting international seminars since 1945. This paper discusses the challenges that we encounter. It presents some of the ways in which we deal with the challenges and our thoughts on what further steps we must take in order to make the training as effective as possible.

4 Bootstrap Methods and Time Series Data

Business and Economics Statistics Section

Sunday, August 8, 2:00 pm–3:50 pm

Applications of the Fast Double Bootstrap

◆ James G. MacKinnon, Queen's University

Queen's University, Dept. of Economics, 94 University Ave., Kingston, ON K7L 3N6 Canada

jgm@qed.econ.queensu.ca

Key Words: *bootstrap test, double bootstrap*

The fast double bootstrap is an approximation to the double bootstrap that can provide more accurate results than ordinary bootstrap tests. For it to be valid, the test statistic must be independent of the random parts of the bootstrap data-generating process. This paper presents simulation evidence on the performance of fast double bootstrap tests in several cases of interest to econometricians. The procedure appears to work well in some cases but not all.

Semiparametric Markov Bootstrap

◆ Bruce E. Hansen, University of Wisconsin

University of Wisconsin, Dept. of Economics, 1180 Observatory Dr., Madison, WI 53706

bhansen@ssc.wisc.edu

Key Words: *time series, kernel estimation, nonparametric, conditional distribution*

The time-series Markov bootstrap is based on a nonparametric estimate of the one-step-ahead conditional distribution. When the conditioning set is moderately large, nonparametric estimation is not feasible. We propose a semiparametric specification of the conditional distribution. A parametric reduces the series to one with reduced serial dependence, and then the remainder is treated nonparametrically. The reduction in the nonparametric dimension improves the rate of convergence of the estimator, and therefore improves the refinement rate of the bootstrap.

Infinite Order Kernels and Nonparametric Regression

◆ Tim McMurray, University of California, San Diego; Dimitris Politis, University of California, San Diego

University of California, San Diego

tmcmurry@math.ucsd.edu

Key Words: *integrated time series, tapered bootstrap, unit root testing*

The usefulness of infinite order kernels in nonparametric regression and autoregression will be discussed. Smoothing with infinite order kernels leads to estimators with minimal bias and an optimal rate of convergence over an appropriate class of smooth functions.

In addition, it allows for a bandwidth selection process which adapts to the smoothness of the underlying function. Aided by infinite order kernels, applications of the bootstrap to constructing improved confidence bands for nonparametric regression will also be presented.

5 Bayesian Solutions to Challenging Problems in Ecology ⟁

Section on Bayesian Statistical Science, Section on Statistics and the Environment, Section on Statistical Computing

Sunday, August 8, 2:00 pm–3:50 pm

Hierarchical Bayesian Models for Spatio-temporal Processes in Ecology

◆ Christopher K. Wikle, University of Missouri, Columbia

University of Missouri, Columbia, Dept. of Statistics, 146 Middlebush, Columbia, MO 65211

wiklec@missouri.edu

Key Words: *spatial, partial differential equations*

Ecological processes often encompass a very extensive range of spatial and temporal scales of variability, and include complicated interactions across domains, variables, and systems. To understand and eventually predict such complicated processes, we must make use of available scientific knowledge, as well as honestly account for uncertainties in that knowledge. Furthermore, we must account for the uncertainty in our observations. Hierarchical Bayesian spatio-temporal models can accommodate these uncertainties and provide rich models for ecological processes. In particular, we will discuss such models motivated by partial differential equation priors.

Model-based Combination of Spatial Information for Stream Networks

◆ Mark S. Handcock, University of Washington

University of Washington, Box 354320, Seattle, WA 98195-4320

handcock@stat.washington.edu

Key Words: *spatial-temporal, hierarchical model, environmental, directed acyclic graph, network*

We describe an approach to improve understanding of the biological integrity of stream and river systems in the United States mid-Atlantic Region by combining information from separate spatial-temporal monitoring surveys, available contextual information on hydrologic units and remote-sensing information. We develop hierarchical spatial statistical models for environmental indicators on the streams and rivers that capture the spatial variation in the measures. These models have been used to estimate the indicators through the riverine system based on the information from multiple sources and aggregate scales. We also quantify the uncertainty in the estimates and develop methods to visualize the resulting estimates and uncertainties. This research illustrates how statistical methodology can be used to leverage the information in scattered monitoring surveys.

State-space Models for Biological Monitoring Data

◆ Devin S. Johnson, University of Alaska, Fairbanks; Jennifer A. Hoeting, Colorado State University

University of Alaska, Fairbanks, Dept. of Mathematical Sciences, Fairbanks, AK 99775-6660

ffdsj@uaf.edu

Key Words: *compositional data, graphical models, logistic normal, random effects, species assemblage, conditional independence*

An emerging area of research in ecology is the analysis of functional species assemblages. In essence, the analysis of functional assemblages is concerned with determining and predicting the composition of individuals categorized using different life history traits instead of strict taxa names. We propose a state-space model for the analysis of multiple trait compositions along with site-specific covariate information. A site-specific random effects term allows for modeling extra variability including spatial variability in trait compositions. This approach has several advantages over the traditional logistic normal model used in the analysis of similar compositional data. The model can also be considered in terms of a chain graph model. If there are no structural zeros in the space of possible trait combinations (combinations of traits that are impossible), we show that the model parameters correspond to conditional independence relationships. Using a Gibbs sampling approach, we illustrate application of the model on a dataset of fish species richness in the mid-Atlantic region of the U.S.

6 gR Graphical Model in R

Section on Bayesian Statistical Science, Section on Statistical Computing, Section on Statistical Graphics

Sunday, August 8, 2:00 pm–3:50 pm

GRbase: A Common Platform for Graphical Models in R

◆ Claus Dethlefsen, Aalborg University

Aalborg University, Dept. of Mathematics, Fredrik Bajers Vej 7G, Aalborg, DK-9220 Denmark

dethlef@math.auc.dk

Key Words: *R, graphical models, software*

We have developed a package, gRbase, consisting of a number of classes and associated methods to support the analysis of data using graphical models. It is developed for the open source language, R, and is thereby available for several platforms. The package is intended to be widely extendible and flexible so that developers may implement further types of graphical models using the available methods. Grbase contains methods for representing data, a representation of models using graphs, specification of models using a formal language and an interactive graphical user interface for working with models. We show how these building blocks can be combined and integrated with inference engines in some special cases of directed and undirected graphical models.

R Functions for Computing Graphs Induced from a DAG after Marginalization and Conditioning

◆ Giovanni Marchetti, University di Firenze

University di Firenze, Dipartimento di Statistica, viale Morgagni, 59, I 50134 Firenze, Italy

gmm@ds.unifi.it

I describe a new R package, which extends the ggm package, for deriving from a given DAG model some graphs induced after marginalizing over and conditioning on some variables. The package has a few basic functions that find the essential graph, the induced concentration and covariance graphs, the chain graph implied by the DAG after grouping and reordering the variables, the ancestral graph, and the summary graph. The new package has the same format of ggm and can be useful to explore the impact of latent variables or of selection effects on a chosen data-generating model.

Bayesian Graphical Models for Locating Mobile Wireless Users

◆ David Madigan, Rutgers University

Rutgers University, Dept. of Statistics, 574 Hill Center, Piscataway, NJ 08855

madigan@stat.rutgers.edu

The increasing use of wireless networking, especially 802.11 ("Wi-Fi")-based wireless systems in enterprise networks, is generating interest in techniques to accurately estimate user location. Applications include location-aware content delivery, emergency location, and access control. This talk will discuss an approach using hierarchical Bayesian models that offers some surprising advantages. The talk will also describe some implementation details involving Perl and scripted WinBUGS.

7 Statistics in Finance & Marketing ▲

ASA, Southern Ontario Chapter (Toronto), Business and Economics Statistics Section, Section on Statistics and Marketing

Sunday, August 8, 2:00 pm–3:50 pm

Predictive Modeling for Consumer Purchase Decisions

◆ Zhen Mei, Manifold Data Mining, Inc.

Manifold Data Mining, Inc., 501 Alliance Ave. Suite 205, Toronto, ON M6N 2J1 Canada

zhen@manifolddatamining.com

Key Words: *predictive modeling, cluster analysis, consumer behavior*

Consumer purchase decision is the process of selecting, purchasing, using, or disposing of products, services, ideas, or experiences to satisfy the consumers needs and desires. As the economy has been shifting from product-driven to customer-focused, understanding consumer behaviors has become an integrated part of business and marketing operations for many companies. Consumer purchase behavior is influenced by many factors, such as culture, social class, reference groups and family, and psychological influences, and so on. Demographics, lifestyle, and household expenditures are the key quantifiable variables for describing and understanding consumer behaviors. We will introduce a hybrid algorithm of cluster analysis and parametric modeling to predict the likelihood of consumers in responding favorably to a given product. By leveraging the predictive power of Manifold comprehensive micro-marketing databases and an adaptive dimension reduction technique we are able to link consumer behavior with the demographic data: age, family, education, dwelling, occupation, employment, income, ethnicity, etc., and household expenditure data on food, clothing, shelter, recreation, etc.

Integrating Structural Modeling, Financial Linkages, and Initiative Development: Techniques to Bring Research Back to Earth

◆ Patrick J. Barton, Synovate Symmetrics; Bruce A. Corner Synovate Symmetrics

8700 E. Via de Ventura, Suite 310, Scottsdale, Arizona 85258

pbarton@symmetrics.com

Key Words: *customer relationships, customer loyalty, structural equation modeling, causal modeling, marketing analytics*

Many of Synovate Symmetrics' clients struggle with the perennial challenge of integrating "pure" research results with more tactical initiative development, particularly when initiative selection is driven by financial considerations that lie outside the scope of the research. Some of Synovate Symmetrics' current research involves analytical integration of structural equation modeling with financial linkage analysis coupled within a highly accessible dynamic simulation environment. This presentation describes our general integration methodology and provides examples of its application to specific clients.

8 Topics in Statistical Genetics: From Linkage to Gene Discovery ▲

Biometrics Section, Caucus for Women in Statistics, WNAR

Sunday, August 8, 2:00 pm–3:50 pm

A Nonparametric Confidence Set Inference Procedure for Gene-mapping

◆ Shili Lin, The Ohio State University

The Ohio State University, Dept. of Statistics, 1958 Neil Ave., Columbus, OH 43210-1247

shili@stat.ohio-state.edu

Key Words: *coverage probability, relative risks, disease gene localization, linkage analysis, multiple tests*

In genome scan studies, tests are often performed to search for signals of linkage throughout the genome for hundreds or even thousands of genetic markers. This practice has raised several

important statistical questions, including multiplicity adjustment, confidence inference, and asymptotic properties. I will present an alternative approach for gene mapping based on Confidence Set Inference (CSI). This procedure constructs a confidence set for the location of a disease locus directly. The confidence set is constructed in such a way that multiplicity adjustment is unnecessary, no matter how many markers are tested. Furthermore, our formulation enables us to localize the disease gene to a small genomic region, an attractive feature for fine mapping. Simulation studies are carried out to demonstrate the advantages of CSI and to evaluate its performance for several nonparametric test statistics. The effects of violations of two key assumptions are also evaluated. An application of CSI to a dataset made available by the Genetic Analysis Workshop 13 confirms its practical utility, as the results compare favorably to those obtained from a standard nonparametric approach.

Genomic Convergence: Combining Genetic and Genomic Information to Identify Genes for Complex Traits

◆ Elizabeth R. Hauser, Duke University Medical Center

Duke University Medical Center, 595 LaSalle St., Box 3445, Center for Human Genetics, Durham, NC 27710-0001

elizabeth.hauser@duke.edu

Key Words: *gene expression, association studies, linkage analysis*

With the completion of the human genome sequence, new resources may be brought to bear on the problem of gene identification for complex disease. As the sources of information have become richer, the complexity of genetic analysis has increased. We have coined the term “Genomic Convergence” to describe the process of combining multiple types of genomic information to identify susceptibility genes for complex genetic disease. Sources of information may include genetic linkage analysis in families and genetic association studies, both in families and in case-control datasets. These data may be augmented by information on gene expression in affected tissues, locations of genes and ESTs in the sequence database, as well as regions of evolutionary conservation across species. In addition, phenotypic and clinical covariates may also help clarify the role of individual genes in disease etiology. I will discuss methods for combining information across a variety of sources with the goal of speeding identification of susceptibility genes for complex traits. These concepts will be illustrated with examples from studies of several complex diseases currently under way at the Duke CHG.

Some Recent Developments for Massive Multiple Comparisons and Clustering of Microarray Gene Expression Data

◆ Kim-Anh Do, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, Dept. of Biostatistics, 1515 Holcombe Blvd., Houston, TX 77030-4009

kim@mdanderson.org

Key Words: *microarray, gene expression*

We discuss two new methodology/software developments for the analysis of microarray gene expression: (1) Model-based inference is proposed for differential gene expression, using a nonparametric Bayesian probability model for the distribution of gene intensities under different conditions. The probability model is a variation of traditional Dirichlet process (DP) mixture models. We illustrate the proposed method in a simulation study and a microarray experiment in colon cancer versus normal tissue. We will discuss the ease of making joint inference about a subgroup of genes being differentially expressed and of estimating the total number of significantly expressing genes. Further, the control of false positive rates can be automatically incorporated into this approach. (2) “Geneshaving” and the related program GeneClust (developed at M.D. Anderson Cancer Center), can be used for either supervised or unsupervised clustering of microarray gene expression data. The method is discussed and applied to the analysis of some well-known datasets: the colon data of Alon, et al. (2000), the leukemia data of Golub, et al. (2000), and the NCI60 data.

9 Record Linkage Applications ▲ ⌘

Section on Survey Research Methods, Social Statistics Section, Section on Government Statistics

Sunday, August 8, 2:00 pm–3:50 pm

Use of Probabilistic Record Linkage for the Canadian Cancer Registry

◆ Brad W. Thomas, Statistics Canada; Jocelyne Marion, Statistics Canada

Statistics Canada, R.H. Coats Bldg., 16th Floor, Tunney’s Pasture, Ottawa, ON K1A 0T6 Canada

thombra@statcan.ca

Key Words: *probabilistic record linkage, internal linkage, unduplication, external linkage*

Cancer is a major cause of deaths in Canada. Both its detection and treatment place great burdens on the Canadian Health Care system. It is vital for the Canadian government to be able to estimate accurately the incidence of cancer in Canada each year, as well as be able to detect trends which indicate increases or decreases, in order for sufficient health care funds to be made available. Cases of cancer incidence in Canada are reported to the Provincial/Territorial Cancer Registries (PTCRs), who in turn pass this information onto Statistics Canada, for inclusion onto the Canadian Cancer Registry (CCR), a patient-oriented database that has cancer incidence in Canada recorded since 1992. This presentation focuses on estimating the incidence of cancer in Canada and the use of probabilistic record linkage to reduce cases of both underestimation and overestimation of cancer incidence. There are two linkages done in annual CCR production: (1) an internal linkage of the CCR to itself to detect possible duplicates of patients or tumours; (2) an external linkage of the CCR to the Canadian Mortality DataBase (CMDB) to find cases of cancer that were not detected until later.

Research Use of Restricted Data: The HRS Experience

◆ Michael A. Nolte, University of Michigan; Janet J. Keller, University of Michigan

University of Michigan, Institute for Social Research, 426 Thompson St. Room 3100 ISR, Ann Arbor, MI 48104

manolte@isr.umich.edu

Key Words: *microdata, record linkage, secure environment, virtual private network, disclosure limitation*

The University of Michigan Health and Retirement Study (HRS) surveys more than 22,000 Americans over the age of 50 every two years. Supported by the National Institute on Aging (U01 AGO 9740), the study paints an emerging portrait of an aging America's physical and mental health, insurance coverage, financial status, family support systems, labor market status, and retirement planning. Most HRS data products are available without cost to researchers and analysts, although user registration is required in order to download files. Other HRS datasets are restricted, and are available only under specific contractual conditions that severely limit their uses. This paper describes the steps taken by HRS to allow researcher access to restricted data while at the same time maintaining respondent confidentiality. The following topics related to HRS data linkages are addressed: content, researcher eligibility, distribution procedures, technical issues involved in implementing a data protection plan, alternative distribution methods, disclosure limitation.

National Record Linkage of UK Hospital Records and Death Registrations

◆ Leicester E. Gill, University of Oxford

University of Oxford, Dept. of Public Health, Old Rd. Campus, Headington, Oxford, OX3 7LF UK

leicester.gill@dphpc.ox.ac.uk

Key Words: *record linkage, national files, false positives, false negatives*

We describe some of the innovative features of the Oxford Record Linkage Study (ORLS), the developments in probabilistic matching methods, and the use of intelligent and data-mining methodologies to select potential links between pairs of records. The ORLS person linked file contains a collection of linkable abstracts that comprise a health region in the United Kingdom. The ORLS file contains 12 million records for six million people and spans 39 years. This dataset is used for the preparation of person-linked health services statistics, and for epidemiological and health services research. The policy of the ORLS is to comprehensively link all the records rather than prepare links on an ad hoc basis. The ORLS have developed improved techniques for deterministic and probabilistic linkage and also methods for reducing the amount of clerical review, which is time consuming, expensive, and of variable quality. The methodology has been extended and refined for matching and linking large UK government datasets, in particular the National Health Service Central Register (60+ million records), cancer registry, and other disease registers, and local authority registers.

Child Immigrants as Citizens

◆ Karen A. Woodrow-Lafield, University of Notre Dame

University of Notre Dame, Institute for Latino Studies, PO Box 764, 250 McKenna Hall, Notre Dame, IN 46556-0764

KarenWLafield@cs.com

Key Words: *immigrants, citizens, children, derivative, administrative, parents*

This study investigates naturalization among child immigrants. Parents' actions in naturalizing simultaneously convey U.S. citizenship to minor children, and parents need only request certificates of citizenship. U.S.-born parents may similarly convey citizenship to adopted children. Adult sons and daughters of naturalized parents may later request a U.S. passport on the basis of derivative citizenship status through the Department of State. Individuals who immigrated as children may naturalize independently as adults. Less than 5% of foreign-born children are naturalized and probably derived that status at parental naturalization. The older a child at immigration, likelihood of derivative status is lesser simply because older children would be more likely to age out of eligibility as parents met residency requirements to naturalize. Children aged 0 to 10 years at immigration would be more likely to derive citizenship. This analysis draws on linked records for immigrants (1978-1991) and naturalizations (1978-1996). Despite limitations, preliminary analyses illustrate this approach adds to understanding of naturalization outcomes for child immigrants.

Approximate String Comparator Search Strategies for Very Large Administrative Lists

◆ William E. Winkler, U.S. Census Bureau

U.S. Census Bureau, Statistical Research, Room 3000-4, 4700 Silver Hill Rd., Stop 9100, Washington, DC 20233-9100

william.e.winkler@census.gov

Key Words: *search mechanisms, approximate string comparison, computer-matching*

Rather than collect data from a variety of surveys, it is often more efficient to merge information from administrative lists. Matching of person files might be done using name and date-of-birth as the primary identifying information. There are obvious difficulties with entities having a commonly occurring name such as John Smith that may occur 30,000+ times (1.5 for each date-of-birth). If there are 5% typographical error in each field, then using fast character-by-character searches can miss 20% of true matches among uncommonly occurring records where name plus date-of-birth might be unique. This paper describes some existing solutions and current research directions.

10 Recent Developments in Statistics of Income Research ▲ ⌘

Section on Government Statistics

Sunday, August 8, 2:00 pm–3:50 pm

The 1979-2002 SOI Continuous Work History Sample Individual Income Tax Panel

◆ Michael E. Weber, Internal Revenue Service

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

michael.e.weber@irs.gov

Each year, SOI produces a cross-sectional microdata file of individual income tax returns. Since 1979, multiple CWHS Social Security Number (SSN) series have been included as part of the cross-sectional sampling process. When the CWHS portion of these yearly cross-sectional samples are linked together using the primary SSN listed on the tax return a longitudinal sample of individual income tax returns can be created. In the past, a limited dataset of such longitudinal income tax data was released in a public-use format. Unfortunately, this public-use file was discontinued in 1990 due to disclosure concerns. However, the yearly cross-sectional files have continued to accumulate and the SOI Individual Income Tax Return CWHS Panel file now spans 1979-2002. In addition, the current version includes information on gender and age. This paper will explain how the SOI Individual Income Tax Return CWHS Panel is created and "cleaned," as well as provide some basic tabulations that can guide future users of the panel as well as discuss how economists can use this file for tax policy research.

Further Analysis of the Distribution of Income and Taxes

◆ Michael Strudler, Internal Revenue Service; Thomas B. Petska, Internal Revenue Service; Ryan Petska, Ernst & Young LLP

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

michael.i.strudler@irs.gov

Key Words: *income distribution, tax burden, administrative records*

Administrative records from tax returns are a reliable source of information on the distribution of personal income and tax burdens. This paper is the sixth in a series examining trends in the distribution of individual income and taxes based on a consistent and comprehensive measure of income derived from individual income tax returns. In this analysis, we examine changes in the income distribution and tax burdens between 1979-1999 including the effects of federal income and Social Security payroll taxes on the after-tax distribution of income. In addition to analysis of the basic data, we estimate Lorenz curves and Gini coefficients to see how income inequalities have changed over time. We use a panel of returns in order to study how incomes and taxes change for individuals over time. Using these data, we look at after-tax earnings over the whole time period to get a better understanding of the nature of income inequalities. We conclude with some thoughts on the historical trends in pre- and post-tax distributions and burdens.

Use of Individual Retirement Arrangements to Save for Retirement, 1996-2001—Results from a Matched File of Tax Returns and Information Documents

◆ Peter J. Sailer, Internal Revenue Service; Sarah A. Holden, Investment Company Institute

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

peter.j.sailer@irs.gov

By combining tax returns and information returns in one database, the Statistics of Income (SOI) Division has made it possible to study trends in contributions to Individual Retirement Arrangements (IRAs), as well as the participation in other types of retirement plans, by individual taxpayers. The authors will compare the number of taxpayers using traditional deductible IRAs, nondeductible IRAs, and Roth IRAs to the number taxpayers eligible to use them. For the nonusers, coverage by other types of plans will be shown as well. Trends in retirement savings by various age and income groups will be discussed.

Customer Satisfaction Initiatives within the Statistics of Income of the Internal Revenue Service

◆ Kevin Cecco, Internal Revenue Service

Internal Revenue Service, Statistics of Income, PO Box 2608, Washington, DC 20013

kevin.cecco@irs.gov

Key Words: *customer satisfaction, survey results*

Measuring Customer Satisfaction continues to play an important role for many of the Federal Statistical Agencies. The Statistics of Income (SOI) Division of the Internal Revenue Service relies on results from a number of customer satisfaction surveys to assess customer feedback and improve the information and services provided to our customers. SOI has made a commitment to administer satisfaction surveys to our primary customers at the Office of Tax Analysis, the Joint Committee on Taxation, and the Bureau of Economic Analysis, as well as to internal employees and customers within the IRS. As a critical source of valuable information, the surveys allow SOI to tailor data dissemination efforts more effectively. This paper will focus on highlighting and summarizing results from these customer surveys, document the process for collecting this information, and discuss future plans for expanding customer satisfaction initiatives within SOI.

Back to Basics: Reconciling Definitions of Key Variables

◆ William R. Pratt, Internal Revenue Service

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

william.r.pratt@irs.gov

Key Words: *administrative records, reconciling definitions*

Comparing government statistics across series, even within a single agency, can be quite challenging. One of the key difficulties is the use of similar terminology with different definitions. This paper will take a look at why these differences in definitions exist and how understanding them can prevent flawed analysis. Several terms and their variations will be examined and compared, including business receipts, total receipts, and total income.

11 Imaging and Related Topics

Section on Statistical Computing

Sunday, August 8, 2:00 pm–3:50 pm

Information Scaling Laws in Vision

◆ Ying Nian Wu, University of California, Los Angeles; Cheng-en Guo, University of California, Los Angeles; Song-Chun Zhu, University of California, Los Angeles

ywu@stat.ucla.edu

When we walk in the woods, the leafs and twigs nearby can be perceived individually, whereas the leafs and twigs far away only give us a collective foliage impression. This perceptual transition from individual structures to collective textures is ubiquitous in natural scenes, but it has not been theoretically studied. We explain this phenomenon by two scaling laws: If we move farther away from a pattern, (1) the resulting image looks more random, and (2) the pattern appears less perceptible. We also identify several important concepts, and connect two generic image representations in literature.

Pattern Reproduction Using Likelihood Inference for Cellular Automata

◆ Radu V. Craiu, University of Toronto; Thomas C.M. Lee, Colorado State University

University of Toronto, Dept. of Statistics, 100 St. George St., Toronto, ON M5S 3G3 Canada

craiu@utstat.toronto.edu

Key Words: *stochastic cellular automaton, binary patterns, greyscale patterns, cellular automaton, minimum description length principle*

Cellular automata are discrete dynamical systems evolving in discrete time over a lattice structure with a multitude of cells. Recent computer experiments have shown that cellular automata dynamics can be used to generate a wide variety of patterns, some highly irregular. Here we report initial findings on likelihood methods for the inverse problem. Specifically, for a given pattern we try to estimate the stochastic cellular automaton that may reproduce it. Issues related to model selection are solved using the minimum description length principle. We will illustrate the approach with a series of patterns some of which are successfully reproduced.

Spatio-temporal Wavelet Analysis for Functional MRI

◆ Victor Solo, University of Michigan; Christopher Long, Massachusetts General Hospital; Manoach Dara, MGH/NMR Center; Emery N. Brown, Harvard Medical School

University of Michigan, Dept. EECS, 1301 Beal Ave., Ann Arbor, MI 48109

vsolo@eecs.umich.edu

Key Words: *MRI, spatio-temporal, wavelets*

Characterizing the spatio-temporal behavior of the BOLD signal in functional MRI (fMRI) is a central issue in understanding brain function. While the nature of functional activation clusters is fundamentally heterogeneous, many current analysis approaches use spatially invariant models that can degrade anatomic boundaries and distort the underlying functional activation. Furthermore, few analysis approaches use true spatio-temporal continuity in their statistical formulations. To address these issues, we present a novel spatio-temporal wavelet procedure that uses a stimulus con-volved hemodynamic signal plus correlated noise model. The wavelet fits, computed by `1-constrained maximum-likelihood estimation, provide efficient multiscale representations of heterogeneous brain structures, and give well-identified, parsimonious spatial activation estimates that are modulated by the temporal fMRI dynamics. In a study of both simulated data and actual fMRI memory task experiments, our new method gave lower mean-squared error and more focal activation maps than models using standard wavelet and pre-smoothing techniques.

Using Ultra-high-resolution Spectra to Reconstruct the Temperature and Composition of an Astronomical Source

◆ David A. van Dyk, University of California, Irvine; Hosung Kang, Harvard University; Vinay Kashyap, Harvard-Smithsonian Center for Astrophysics; Alanna Connors, Eurika Scientific

University of California, Irvine, CA 92697-1250

dvd@uci.edu

Key Words: *astronomy, Bayesian methods, missing data, MCMC, EM algorithm, spectral analysis*

An astronomical spectrum is essentially a histogram of the wavelength of the electromagnetic radiation from an astronomical source. In empirical high-energy astrophysics data is collected on the wavelength/energy of each photon that arrives at a detector. High-resolution spectra carry subtle information as to the physical environment of the cosmological source. For example, when an electron jumps down from one quantum state of an atom to another, the energy of the electron decreases. This energy is radiated away from the atom in the form of a photon with energy equal to difference of the energies associated with the two quantum states. These differences are unique to the ion that produced the photon and can be used to identify the ion. If the source is relatively hot, the emission corresponding to more energetic quantum states will be relatively strong. Thus, a spectrum carries information as to the temperature of the source. We describe the model-based statistical methods developed by the California-Harvard Astrostatistics Collaboration that aim to use spectral data to reconstruct the physical environment of an astronomical source.

Feature Representation and Pattern-filtering for Acoustic Detection

◆ Zhiyi Chi, University of Chicago

University of Chicago, 5734 University Ave., Dept. of Statistics, Chicago, IL 60637

chi@galton.uchicago.edu

Key Words: *pattern recognition, point processes, neuroscience, bioacoustics, simulation*

Many biological studies require detection of various constituents of bioacoustic signals. We developed a fast and accurate approach to the avian acoustic detection. The key is to represent signals by points on the spectro-temporal domain, which significantly reduces the dimension of the data and allows us to formulate acoustic detection as detection of certain global patterns of points. Under a Poisson point process model, the latter detection is formulated as classification based on likelihood ratio and shown to be equivalent to linear filtering of point processes. Accuracy and computational efficiency are achieved by applying multiple filters sequentially, so that only ambiguous “hot spots” in the signal are processed in each step. The training of the detector only requires a small sample. It constructs the filters based on structures learned from the sample, and sets the rest of the parameters by simulation. When implemented online, the approach enables neuroscientists to conduct experiments on neurobehavioral interactions with a degree of precision that could not be achieved before.

12 Phase 2/3 Combination Designs to Accelerate Drug Development ▲ ⊞

Biopharmaceutical Section

Sunday, August 8, 2:00 pm–3:50 pm

Issues with Phase II/III Combination Designs

◆ H.M. James Hung, U.S. Food and Drug Administration

U.S. Food and Drug Administration, DB1, CDER, Room 5062, 1451 Rockville Pike, Rockville, MD 20852

hung@cder.fda.gov

Key Words: *efficiency, missing data*

Combining phase II and phase III may improve efficiency for drug development. The extent of the improvement, however, requires extensive evaluation subjecting to many practical constraints. The typical phase II designs that traditionally are relatively small and study narrow patient populations offer information that is often limited in assisting the planning of phase III trials. Combining phase III designs with such phase II designs may not offer much improvement in efficiency. An informative phase II design needs to allow full exploration of the dose range for efficacy and safety of the study drug. It also needs to facilitate exploration of the nature of dropouts or missing data so that the dropouts or missing data may be minimized and a sensible strategy of sensitivity analysis may be planned to handle missing data. Any useful statistical methods, frequentist or Bayesian, should be used to explore the phase II data. Some parts of such informative phase II data may be combined with phase III data. Issues with such combination include those of what Type I (or false positive) error needs to be of concern, how to combine, and more.

Phase 2/3 Combination Designs Comparing Several with a Control

◆ Qing Liu, J&J Pharmaceutical Research and Development

J&J Pharmaceutical Research and Development, 920 Route 202, PO Box 300, Raritan, NJ 08869

qliu2@prdus.jnj.com

Key Words: *adaptive design, adaptive closed-testing procedure, Dunnett test*

We propose a phase 2/3 combination design for late-stage clinical development to identify a more effective treatment than a standard control among several experimental treatments. The design consists of two stages. In the first stage, patients are randomized to one of the several treatments or control. At the end of the first stage, short-term safety and efficacy are examined, upon which promising treatments are selected for further evaluation. The required sample size and method of analysis are also determined. In the second stage, newly enrolled patients are randomized to only one of the selected treatments or control; and patients for the selected treatments or control, including those enrolled in the first stage, are evaluated using a clinical endpoint requiring a longer follow-up. At the end of the second stage, data of both stages from patients randomized to one of the selected treatments or control are integrated for final hypothesis testing via an adaptive closed-test procedure. A notable feature is that the decision rule governing regimen selection, sample size calculation, and method of analysis need not be specified in advance to maintain the validity of the trial.

Optimized Two-stage Adaptive Design

◆ Keaven M. Anderson, Merck & Co., Inc.; Qing Liu, J&J Pharmaceutical Research and Development

Merck & Co., Inc., BL3-2, PO Box 4, West Point, PA 19486-0004

keaven_anderson@merck.com

Key Words: *adaptive design, group sequential design, optimal design, clinical trials*

Proschan and Hunsberger proposed an adaptive two-stage design with the sample size and critical value for the second stage determined by the observed treatment difference and critical value from the first stage. A conditional error function maps the observed critical value in stage 1 to a required critical value for stage 2 to maintain the overall Type I error. The observed treatment difference and desired power are then used to determine the second stage sample size. Liu and Chi modified this procedure by substituting a minimum treatment difference of interest for the observed stage 1 treatment difference. We select an optimized Liu and Chi design by defining a flexible conditional error function family, assuming a prior distribution for the true treatment difference and selecting a loss function (e.g., expected sample size). Examples will be given comparing these optimal two-stage adaptive designs to optimal two (or more)-stage group sequential designs. Cases will be presented where the optimal two-stage design from either class is slightly better. In general, the optimal two-stage adaptive and group sequential designs are comparable.

Nonstop Drug Development: Strategies for a Quick Transition from Phase II to Phase III

◆ Ronald W. Helms, Rho, Inc.; Karen Kesler, Rho, Inc.

Rho, Inc., 100 Eastowne Dr., Chapel Hill, NC 27514

Ron_Helms@RhoWorld.com

Key Words: *clinical trials, accelerated drug development, NonStop, biopharmaceutical*

The NonStop drug development strategy, as introduced by Helms (2001), can be used to make dramatic reductions in the time and cost of phase II, under appropriate circumstances. This presentation focuses on the transition from phase II to phase III, addressing some issues not addressed in the original presentation. Briefly, a phase II implementation of the NonStop strategy is based on one or more phase II studies with the objective of selecting a treatment regimen to be carried forward into phase III. A phase II study design includes a control treatment and a set of "many" (e.g., 4-12) "active" treatment regimens or "treatment arms," covering a spectrum that has a high probability of including at least one that can successfully be carried to phase III. Frequent interim analyses are conducted with the objective of "pruning" most of the treatment arms as quickly as possible, either for low efficacy or a poor safety profile. The goal is quickly to reduce the number of remaining treatment arms to one and, as soon as this happens, immediately transition to phase III. This presentation addresses solutions to issues that arise in the phase II-III transition.

13 Web-based Statistical Tools ▲

Section on Quality and Productivity

Sunday, August 8, 2:00 pm–3:50 pm

Clearing Out the Cobwebs for "Accidental" Statisticians: A Web-based Tool

◆ Cheryl L. Jennings, Motorola Semiconductor Products Sector; Kathryn Kennedy, Arizona State University

Motorola Semiconductor Products Sector, 3841 E Talowa St., Phoenix, AZ 85044

c.jennings@motorola.com

Key Words: *web-based tool, expert knowledge, statistical resource*

Making the utilization of statistical analysis tools approachable to infrequent users is a challenge. Although many have received rudimentary, software-based training, they have not gained skill from repetitive application. They may be unprepared to apply concepts to atypical problems or to even utilize different software. When the availability of human expert assistance is limited, the desktop computer offers a convenient opportunity to fill the void. To help these users decide when and how to use the right tools, we created an intranet site to share expert knowledge and lessons learned. This talk discusses a web-based strategy for assisting infrequent users and provides suggestions for developing a user-friendly statistical resource.

Delivering Statistical Reports to Executives via Web-based Dashboards

◆ Samaradasa Weerahandi, Time Warner; Martin Koschat, Time Warner

Time Warner, 751B, West 50th St., New York, NY 10020

Sam_Weerahandi@timeinc.com

Key Words: *marketing dashbords, SPlus analytics, StatServer, Java*

This presentation addresses some issues in developing and delivering statistical reports meant for executive dashboards. Reports may range from simple reports on company revenue and sales performance to charts that provide insight into company performance in a competitive environment. We will discuss how readily available analytics developed with statistical packages such as SAS and SPlus can be leveraged to such reports and how they can be integrated in web-based dashboards for real-time delivery. The presentation will also include a demo of a Marketing Dashboard. This dashboard is Java applet placed on a web page for easy access. Its reports and charts are developed using SPlus functions.

Implementing Web-based Statistical Tools

◆ Webster West, Integrated Analytics LLC

Integrated Analytics LLC, 200 S. Saluda Ave., Columbia, SC 29205

websterwest@yahoo.com

Key Words: *StatCrunch, web, Java, applet, interactive, software*

There are a number of statistical software tools available for use via the web. A brief survey of the pros and cons associated with the different types of tools will be given. A detailed discussion of StatCrunch will be provided. StatCrunch is a statistical software package freely available at *www.statcrunch.com*. The package, which runs within the context of a web browser, offers a number of standard statistical analysis routines and features some unique components, such as interactive graphics. The discussion will focus on how the StatCrunch tool can be implemented within an industrial setting to provide users with easy access to data and the software to analyze it.

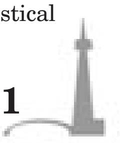
Competitive Advantage with eSTAT

Jose G. Ramirez, W.L. Gore & Associates, Inc.; ◆ Brenda Ramirez, W.L. Gore & Associates, Inc.; Stuart Nagy, W.L. Gore & Associates, Inc.

bramirez@wlgore.com

Key Words: *electronic statistics, web-deployment*

In today's highly competitive environment, it is not enough to collect the right data; it is not enough to use the best statistical approaches with the data; it is not enough to publish results to decisionmakers; it is not enough to train the masses in statistical



thinking. To be truly competitive, a company must be able to turn data into knowledge quickly and reliably, and distribute it to the companies decisionmakers anytime and anyplace. A business's intranet site is the statistician's catalyst to bring truth to the statement "Data + Statistics = Competitive Advantage." We will discuss the power behind web-deployment of statistical ideas and tools and the different ways a statistician can carry out eSTAT (electronic statistics) on their company's intranet sites. We will also describe our experiences using statistical web tools and critical cross-functional allies that can help institutionalize a company's eSTAT program.

14 Spatial and Temporal Models and Applications ▲

Section on Bayesian Statistical Science

Sunday, August 8, 2:00 pm–3:50 pm

Bayesian Spatial Modeling of Air Pollution in 12 Southern California Communities Incorporating Space-time Interaction

◆ Nuoo-Ting Molitor, University of Southern California; John Molitor, University of Southern California

University of Southern California, 1540 Alcazar St., Los Angeles, CA 90089-9011

nuootinl@usc.edu

Key Words: *Bayesian statistics, spatial analysis, space-time interaction, air pollution, lung function*

We present an application of Bayesian spatial methods to the analysis of the effect of air pollution on the lung function of children in Southern California communities. The data were obtained from The Southern California Children's Health Study (CHS), a prospective study in children that was designed to examine long-term effects of air pollution on respiratory health. Bayesian spatial methods incorporating space-time interaction will be employed. Analysis will be performed using the WinBUGS software.

Bayesian Semiparametric Models for Colorectal Cancer Incidence

◆ Song Zhang, University of Missouri, Columbia; Dongchu Sun, University of Missouri, Columbia; Chong He, University of Missouri, Columbia

University of Missouri, Columbia, 510 High St. Apt. 0 326, Columbia, MO 65201

szq22@mizzou.edu

Key Words: *Bayesian analysis, semiparametric models, spatial and temporal effects, colorectal cancer, cancer incidence rates*

It has been known that colorectal cancer incidence rates had been increasing since the middle of 1985 and have been deceasing since then. We have proposed a Bayesian semiparametric model to capture simultaneously the demographic variables, such as age and sex, random geographic variables, such as regional differences, longitudinal variables, such as temporal trends in mortality, and spatial variables, such as the correlation between neighboring regions. The relationship of three different subsites of colorectal cancer is also explored.

Stochastic Search Model Selection for Restricted VAR Models

◆ Shawn Ni, University of Missouri; Dongchu Sun, University of Missouri, Columbia; Edward I. George, University of Pennsylvania

University of Missouri, 118 Professional Bldg., Dept. of Economics, Columbia, MO 65211

nix@missouri.edu

Key Words: *model selection, Bayesian VAR, stochastic search, MCMC*

We conduct Bayesian stochastic variable selection of Vector Autoregressive (VAR) models. We develop a Markov chain Monte Carlo (MCMC) algorithm that selects the elements of both VAR regression coefficients matrix and error variance matrix. Numerical simulations show that our approach is quite effective. We apply the method in selection among models of business cycles with both macroeconomic and industry-specific shocks.

Normal in Econometrics

◆ Daniel Waggoner, Federal Reserve Bank of Atlanta; James Hamilton, University of California, San Diego; Tao Zha, Federal Reserve Bank of Atlanta

Federal Reserve Bank of Atlanta, 1000 Peachtree St. NE, Atlanta, GA 30309

Daniel.F.Waggoner@atl.frb.org

Key Words: *mixture distributions, vector autoregressions, cointegration, regime-switching, numerical Bayesian methods, weak identification*

The issue of normalization arises whenever two different values for a vector of unknown parameters imply the identical economic model. A normalization implies not just a rule for selection which among equivalent points to call the MLE, but also governs the topography of the set of points that go into a small-sample confidence interval associated with that MLE. A poor normalization can lead to multimodal distributions, disjoint confidence intervals, and very misleading characterizations of the true statistical uncertainty. This paper introduces the identification principle as a framework upon which a normalization should be imposed, according to which the boundaries of the allowable parameter space should correspond to loci along which the model is locally unidentified. We illustrate these issues with examples taken form mixture models, structural VARs, and cointegration.

Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach

◆ Jesus Fernandez-Villaverde, University of Pennsylvania

University of Pennsylvania, 160 McNeil, 3718 Locust Walk, Philadelphia, PA 19104

jesusfv@econ.upenn.edu

This paper presents a framework to undertake likelihood-based inference in nonlinear dynamic equilibrium economies. We develop a sequential Monte Carlo algorithm that delivers an estimate of the likelihood function of the model using simulation methods. This likelihood can be used for parameter estimation and for model comparison. The algorithm can deal both with nonlinearities of the economy and with the presence of non-normal shocks. We show consistency of the estimate and its good performance in finite simulations. This new algorithm is important because the existing empirical literature that wanted to follow a likelihood approach was limited to the estimation of linear models with Gaussian innovations. We apply our procedure to estimate the structural parameters of the neoclassical growth model.

15 Biostatistical Methods Using Bayesian Inference ▲

Section on Bayesian Statistical Science

Sunday, August 8, 2:00 pm–3:50 pm

Model Selection for a K-level Categorical Regression Variable: Are All K Levels Necessary?

◆ Susan A. Alber, University of California, Los Angeles; Robert E. Weiss, University of California, Los Angeles

University of California, Los Angeles, 4416 Fulton Ave. 10, Sherman Oaks, CA 91423

alber@ucla.edu

Key Words: *model selection, hierarchical models*

We present a model selection approach for regression with a k-level categorical explanatory variable. The model space (M) contains all possible ways to partition the response means for the k groups into subsets with equal values. A hierarchical model is used with the prior for the k means conditional on M. Several different priors for M are considered. The method will be demonstrated using a six-level categorical treatment variable from a study on the effects of lead exposure.

Modeling Semicontinuous Longitudinal Data

◆ Warren S. Comulada, University of California, Los Angeles; Robert E. Weiss, University of California, Los Angeles

University of California, Los Angeles, School of Public Health, Dept. of Biostatistics, Los Angeles, CA 90095-1772

scomulad@ucla.edu

Key Words: *semicontinuous, longitudinal*

Semicontinuous variables are composed of zero and positive responses. One method to model longitudinal semicontinuous outcome variables has a random effects logistic model for whether a response is positive or zero and a linear random effects model for the positive responses; two correlated random effects are included. We propose a set of models fit using Bayesian methods. One model shares but a single random effect between the models. The other model shares a linear predictor between models and includes two correlated random effects. We discuss computational issues in fitting these models and propose a method to compare models. To illustrate, we apply these models to data from the CLEAR study, a behavioral intervention study of 172 HIV-postive young people. We investigate predictors of recent substance use over time.

Importance Reweighting within Gibbs for Hierarchical Regression Modeling

◆ Li-Jung Liang, University of California, Los Angeles; Robert E. Weiss, University of California, Los Angeles

University of California, Los Angeles, Biostatistics, 3161 Sepulveda Blvd., Apt. 208, Los Angeles, CA 90034

liangl@ucla.edu

Key Words: *hierarchical regression model, mixture of Dirichlet processes, Markov chain Monte Carlo, phylogenetic, HIV*

We develop a statistical model and computational algorithm for combining analyses of a number of datasets. Individual analyses are fit independently using previously written stand-alone software that fits a complex Bayesian model using MCMC simulation. Each individual analysis is computationally intensive and MCMC output from each of these complex Bayesian analyses is available. Constructing a large complex model involving all the original datasets is time consuming and may be difficult. Instead, our strategy is to use the existing MCMC samples of the individual posteriors. We place a hierarchical regression model across the individual analyses for estimating parameters of interest within and across analyses. Our model has two key features. We use a MDP prior for the parameters of interest to relax parametric assumptions and to ensure the prior distribution for the parameters of interest is continuous. We use an importance reweighting algorithm within Gibbs to sample values of the individual parameters. We demonstrate our approach on a set of phylogenetic models of HIV-1 nucleotide sequence data.

Modeling a Mixture of Ordinal and Continuous Repeated Measures

◆ Xiao Zhang, University of California, Los Angeles; Walter J. Boscardin, University of California, Los Angeles; Thomas R. Belin, University of California, Los Angeles

University of California, Los Angeles, 3281 Sepulveda Blvd., #308, Biostatistics Dept., Los Angeles, CA 90034

zxiao@ucla.edu

Key Words: *Gibbs sampler, multivariate probit model, Metropolis-Hastings algorithm, parameter extension*

We study the correlation structure for a mixture of ordinal and continuous repeated measures using a Bayesian approach. We assume a multivariate probit model for the ordinal variables and a normal linear regression for the continuous variables, where latent normal variables underlying the ordinal data are correlated with continuous variables in the model. Due to the probit model assumption, we are required to sample a covariance matrix with some of the diagonal elements equal to one. We use the idea of parameter-extended data augmentation and apply the Metropolis-Hastings algorithm to get a sample from the posterior distribution of the covariance matrix incorporating the relevant restrictions. The methodology is illustrated through a simulated example and through an application to data from the UCLA Brain Injury Research Center.

Bayesian Recombination Identification: New Models and Better Ways of Incorporating Prior Information

◆ Vladimir N. Minin, University of California, Los Angeles; Karin S. Dorman, Iowa State University; Marc A. Suchard, University of California, Los Angeles

University of California, Los Angeles, Dept. of Biomathematics, David Geffen School of Medicine, Los Angeles, CA 90095

vminin@ucla.edu

Key Words: *recombination, HIV evolution, Bayes, MCMC, spatial point process*

We develop a dual Multiple Change Point (MCP) model in a Bayesian framework to improve detection of recombination among aligned nucleotide sequences. We define two a priori independent changepoint processes that describe variation in phylogenetic tree topologies and evolutionary pressures respectively. This approach results in more accurate recombination detection over the single MCP model, where spatial phylogenetic variation is modeled with one changepoint process. We demonstrate the increased accuracy of recombination detection by simulating recombination at different sites in the real dataset of mtDNA sequences from four primates. Sampling from the posterior distribution is accomplished via reversible jump MCMC. To simultaneously increase the precision of the recombination identification, one can use informative priors on recombination locations. We develop a new class of priors using a spatial point process that allows incorporating information about previously detected recombinants a flexible manner. We apply these priors to an HIV nucleotide sequence dataset to estimate the locations of recombination sites.

16 Statistical Methods for Detecting Spatial Clusters ▲ ⊞

Biometrics Section, Section on Statistics in Epidemiology

Sunday, August 8, 2:00 pm–3:50 pm

A Simulation Approach for Detecting Clusters: An Application of Taiwan Cancer Mortality

◆ Ching-Syang J. Yue, National Chengchi University; Foong Ling Chin, National Chengchi University

National Chengchi University, No. 64, Section 2, Chi-Nan Rd., Dept. of Statistics, Taipei, 116 Taiwan (R.O.C.)

csyue@nccu.edu.tw

Key Words: *cluster, rare disease, case-event data, aggregate data, cancer mortality*

In recent years, many statistical methods have been proposed for detecting excesses of rare diseases, i.e., clusters, in space or in space-time. Most of these methods deal with case-event or individual-level data and are designed to detect clusters with shape close to circles. In this study, adapting Choynowski's (1957) idea, a simulation-based approach is proposed to detect noncircular clusters with aggregate or group-level data. The proposed cluster-detection method will be used to compare with Nagarwalla's Spatial Scan Statistic, a frequently used method. Computer simulation is used to illustrate the validity, with respect to Type I and Type II errors, of the proposed approach. In addition, the cancer mortality data in Taiwan area is also used as a demonstration of the proposed test.

A Flexible Scan Statistic for Detecting Arbitrarily Shaped Clusters

◆ Toshiro Tango, National Institute of Public Health, Japan; Kunihiko Takahashi, National Institute of Public Health, Japan

National Institute of Public Health, Japan, 2-3-6 Minami, Wako, International 351-0197 Japan

tango@niph.go.jp

Key Words: *hot-spot clusters, likelihood ratio test, Poisson distribution, spatial epidemiology, power*

Kulldorff's spatial scan statistic has been applied in a wide variety of epidemiological studies for cluster detection. His scan statistic, however, uses a circular window to define the potential cluster areas and thus has a difficulty of correctly identifying actual noncircular clusters. We propose a flexible scan statistic that can identify clusters of any shape. The performance of the proposed procedure is compared with that of Kulldorff's circle-based scan statistic via Monte Carlo simulation. We considered several circular and noncircular hot-spot cluster models and examined the newly introduced bivariate power distribution classified by the number of regions detected as the most likely cluster and the number of hot-spot regions (assumed in the simulation) included in the most likely cluster. The proposed flexible scan statistic is shown to have higher usual powers and also to identify the noncircular hot-spot clusters more accurately than Kulldorff's scan statistic. The proposed procedure is illustrated with some disease maps simulated in the Tokyo Metropolitan area.

How to Evaluate Tests for Identifying Spatial Clusters

◆ Kunihiko Takahashi, National Institute of Public Health, Japan; Toshiro Tango, National Institute of Public Health, Japan

National Institute of Public Health, Japan, 2-3-6 Minami, Wako, International 351-0197 Japan

kunihiko@niph.go.jp

Key Words: *cluster detection, hot-spot clusters, hypothesis testing, power, spatial epidemiology*

Many different tests have been proposed to detect spatial disease clustering without any prior information about their locations. These tests can be classified into two types: CDT (cluster detection tests) such as Kulldorff's special scan statistic, and GCT (global clustering tests) such as Tango's maximized excess events test. To compare the performance of these tests, most of the authors use the power, i.e., the probability of rejecting the null hypothesis of no clustering for whatever reason. However, the power is not always appropriate to evaluate CDT since the purpose of CDT is to both reject the null hypothesis and identify the cluster areas correctly. We shall propose a new performance measure of CDT, which includes the power as a special case. We shall illustrate the proposed measure with two tests, Kulldorff's spatial scan statistic and a flexible scan statistic that we proposed.

A Test to Detect Space-time Clustering and a Comparison with Some Existing Methods

◆ James C. Gear, Monsanto Company; Dana Quade, University of North Carolina

Monsanto Company, Statistics Technology Center - O3B, 800 North Lindbergh St., St. Louis, MO 63167

James.C.Gear@monsanto.com

Key Words: *disease clustering, space-time clustering*

This research introduces a new statistical test for evaluating space-time clustering in data where exact location and time information are available for the points of interest (cases). The test statistic, DP, is defined as the length of the path from X[1] to X[n] when the n cases are ordered by time of occurrence. Significance of the test is most appropriately determined by comparing the directed path length of the data to the empirical distribution of lengths obtained from all possible orderings of the n cases, or a random subset of those orderings when n is large. The properties of this test are investigated using Monte Carlo techniques on clustered and unclustered simulated data. This test is then compared with the commonly used Knox's test and Mantel's Generalized Regression, using Monte Carlo techniques on simulated data. Nearly 20 years of brain cancer data from New Mexico are then examined with these tests, in order to compare the performance of these tests on actual data.

Spatial Cluster Detection Using Bayes Factors from Overparameterized Models

◆ Ronald Gangnon, University of Wisconsin, Madison; Murray Clayton, University of Wisconsin, Madison

University of Wisconsin, Madison, 207 WARF Office Bldg., 610 Walnut St., Madison, WI 23726

ronald@biostat.wisc.edu

Key Words: *spatial clustering, partition models, Bayes factors, Markov chain Monte Carlo*

We consider a partition model for estimation of regional disease rates and for detection of spatial clusters. Formal inference regarding the number of partitions (or clusters) can be obtained using a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm. As an alternative, we consider models with a fixed, but overly large, number of partitions. We explore the ability of these models to provide informal inferences about the number and locations of clusters using localized Bayes factors. We illustrate these two approaches using two datasets from the literature (the New York leukemia data and the North Carolina SIDS data) as well as data on breast cancer incidence in Wisconsin.

17 Analysis of Mortality and Morbidity ▲

Section on Statistics in Epidemiology

Sunday, August 8, 2:00 pm–3:50 pm

Socioeconomic Differentials in Mortality: Does the Myth Prevail in 21st Century?

◆ Jay H. Kim, Centers for Disease Control and Prevention; Jay J. Kim, National Center for Health Statistics; Paul D. Williams, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, NCHS, 3311 Toledo Rd., Room 3106, Hyattsville, MD 20782

jkim@cdc.gov

Key Words: *socioeconomic status, mortality, health service area*

The inverse relationship between socioeconomic status (SES) and mortality rates have been well-known mythology ever since Hauser explored it with 1950 data in Chicago metropolitan area. This subject has been revisited by many researchers over time in many different ways and countries. The authors will use the national death registration of 1999, 2000, and 2001, the most recent data from National Center for Health Statistics, and income and education data by county from Census 2000. After the socioeconomic status indicator for each county is constructed with income and education levels, the mean of SES values of counties that make up a health service area (HAS) will be SES value of that HAS. The 805 HSAs will be assigned to one of the five SES groups. With analysis of variance and concordance test, authors will investigate the mortality rate differentials by SES to explore whether the myth still prevails in the beginning of the 21st century.

Life Expectancy Decomposition by Disease

◆ Charles C. Lin, U.S. Census Bureau; Norman J. Johnson, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 8700 DSMD, 3716-3, Washington, DC 20233-8700

charles.c.lin@census.gov

Key Words: *life expectancy, decomposition, disease-free life-year, mortality, life table*

It is common to use life expectancy to summarize the life-time mortality experience of subgroups of a population of interest. Large differences in life expectancy between certain subgroups are known. For example, Lin et al. have shown that men aged 25 with fewer years of education have a life expectancy of five to seven years less, on average, than those with more years of education. However, in applications such as this example, the major disease contributions to life expectancy differences are difficult to obtain and need to be investigated. We will show a decomposition of life expectancy into disease-free life-years for major diseases. The estimate of each disease-free life-year can be computed backward recursively from the death rates by disease and age. Estimates of variance will be given. The life expectancy difference will include the contributions from major diseases and the difference at older ages which cannot be determined by disease. Life tables with competing risks from major diseases constructed for the National Longitudinal Mortality Study will be used for demonstration purposes. Culprit major diseases underlying life expectancy differences will be identified.

Leukemia Mortality after Fractionated Moderate-dose-rate Ionizing Radiation in the Canadian Fluoroscopy Cohort and a Comparison with Leukemia Mortality among the A-bomb Survivors

◆ Lydia Zablotska, Columbia University

Columbia University, 722 West 168th St., Suite #1100B, New York, NY 10032

lbz7@columbia.edu

Key Words: *leukemia, mortality, chronic lymphocytic leukemia, fluoroscopy, atomic bomb survivors*

Canadian fluoroscopy cohort was used to analyze mortality experience from leukemia due to exposure to fractionated moderate-dose-rate ionizing radiation between 1950 and 1987. A substantial fraction (39%) of subjects received multiple chest x-ray fluoroscopies during the course of pneumothorax for treatment of tuberculosis. Individual bone marrow doses were estimated from combination of the number of fluoroscopies, interviews with physicians who conducted pneumothorax, and data from an experimental human phantom study of organ dose per unit of surface exposure produced by contemporary fluoroscopes. Poisson regression analyses were used to estimate excess relative rates per gray (ERR per Gy) and the associated measures of uncertainty. Cumulative person-year experience was cross-classified by sex, age risk, calendar year at risk, and province. There were 148 deaths from leukemia among 66,464 subjects (13 were chronic lymphocytic leukemia (CLL)). Categorical analyses showed a pattern of increasing risks with increasing dose. Linear dose-response analyses showed increased risks for all leukemia (ERR=0.77) and leukemia excluding CLL (ERR=1.2).

Interaction between BMI and Diabetes in Modeling Mortality

◆ Robert G. Fowler, Florida State University

Florida State University, PO Box 15526, Tallahassee, FL 32317

fowler@stat.fsu.edu

Key Words: *Cox-model, meta-analysis, BMI, diabetes*

This paper investigates the interaction between BMI and diabetes. It is known that both incidence of diabetes and greater BMI increases the relative risk of mortality for most causes of death. Preliminary investigation shows that this interaction is present in several studies and absent in others. Estimates of the relative risk associated with this interaction are calculated using Cox proportional hazard models. The individual-level data for these models is from the diverse population dataset of 27 medical studies that has been compiled at Florida State University. Meta-analysis is used to summarize the differing results from the proportional hazard models. The presence of a BMI-diabetes interaction indicates weight is a greater risk factor for diabetics in comparison to nondiabetics.

Simple Models to Assess the Risk of Developing Diabetes

◆ Ken Williams, University of Texas, San Antonio; Steven E. Stern, Australian National University; Michael P. Stern, University of Texas, San Antonio

University of Texas, San Antonio, Dept. of Med/Epidemiology (stop #7873), Health Science Center, 7703 Floyd Curl, San Antonio, TX 78249

williamsk@uthscsa.edu

Key Words: *diabetes, incidence, prediction, logistic regression, recursive partitioning, score*

Available models which can identify individuals likely to benefit from diabetes preventive treatment are not commonly applied because they require calculations or a two-hour oral glucose tolerance test. We therefore sought ways to easily identify in a clinic, individuals likely to develop diabetes. Of 3,228 subjects examined in the San Antonio Heart Study between 1979 and 1988 and again after seven to eight years, 295 developed diabetes. A logistic regression model from this study estimated the probability of developing diabetes as a function of age, gender, ethnicity, fasting glucose, blood pressure, HDL cholesterol, body mass index, and family history. We dichotomized each continuous predictor and rescaled the coefficients so they totaled from 0 to 100 points to provide a simple risk assessment score sheet. The area under the receiver operating characteristics curve of this score was 0.831 vs. 0.836 (p = 0.51) for the continuous model. We also used recursive partitioning to develop a decision tree model which provided equivalent sensitivity at high specificity from simple decision rules, e.g., treat if fasting glucose > 100 and HDL < 50 mg/dl (sensitivity = 30%, specificity = 97%).

18 How to Write a Book for the *ASA-SIAM Series on Statistics and Applied Probability* ⌘

General Methodology

Sunday, August 8, 2:00 pm–3:50 pm

How to Write a Book for the *ASA-SIAM Series on Statistics and Applied Probability*

◆ Robert N. Rodriguez, SAS Institute Inc.; ◆ Linda C. Thiel, Society for Industrial and Applied Mathematics; ◆ Simon Dickey, Society for Industrial and Applied Mathematics; ◆ Roxy L. Peck, California Polytechnic State University; ◆ Thomas W. O'Gorman, Northern Illinois University

Northern Illinois University, 121 Alden Place, Dekalb, IL 60115-4311

ogorman@math.niu.edu

Key Words: *ASA-SIAM Book Series, book authorship, writing a statistics book, statistics books*

The *ASA-SIAM Series* is the professional book series of the American Statistical Association. It was established as a joint venture between the ASA and the Society for Industrial and Applied Mathematics to publish high-quality, affordable books of general interest to statisticians, biostatisticians, applied mathematicians, scientists, engineers, and other groups of statistical practitioners. An important new focus of the *Series* is short expository overview books (around 150 pages in length) that introduce statisticians to new areas of methodology and application. “How to” books on statistical methods for researchers in other fields are also an excellent fit for the *Series*, as are interdisciplinary works. This discussion will cover the key aspects of writing a successful book for the *Series*—from planning to production. The participants will include experienced authors and editors.

19 General Topics

Section on Nonparametric Statistics

Sunday, August 8, 2:00 pm–3:50 pm

Stochastic Linear Hypotheses for Nonparametric Analysis of Microarrays

◆ Jeanne Kowalski, Johns Hopkins University

Johns Hopkins University, 550 North Broadway, Suite 1103, Baltimore, MD 21205

jkowals1@jhmi.edu

Key Words: *genomics, microarrays, nonparametric, stochastic*

I introduce a class of stochastic linear hypotheses to facilitate high-dimensional comparisons among several groups, based on as few as a single sample per group. This class in particular includes the Mann-Whitney Wilcoxon rank sum test as a special case. I discuss the analytic approach in two stages, within the context of a microarray experiment. In the first part, I estimate a number of candidate genes that characterizes a general comparative criteria by formulating comparisons in terms of stochastic linear hypotheses, which are tested based on developed U-statistic theory. In the second part, I discuss a bioinformatics algorithm for selecting candidate genes by comparing intensity functionals, using inner product and singular value decomposition concepts, in combination. As motivation, I compare intensities among T-cell clones singly exposed to conditions hypothesized as pathways leading to T-cell clonal anergy, followed by the genomic characterization of such conditions.

Nonparametric Estimation of Mass within a Galaxy

◆ Xiao Wang, University of Michigan; Michael Woodroofe, University of Michigan

University of Michigan, 2151 Hubbard Rd. #29, Ann Arbor, MI 48105

wangxiao@umich.edu

Key Words: *kernel-smoothing, isotonic method, mass of a galaxy*

We consider using a nonparametric method to estimate the distribution of mass within a galaxy. The data will consist of positions and spectral information from a sample of a few thousand stars. The data pose interesting statistical questions. The main one is a difficult inverse problem in which one attempts to infer the mass distribution, primarily from the radial velocities of stars (which can be determined from the data). This new procedure is to combine the kernel smoothing techniques with the isotonic methods to obtain a consistent estimate of mass distribution. The use of nonparametric methods is rare with this type of data and so our work has the potential to influence future statistics within astronomy, since only a general model is assumed in this nonparametric approach.

Spatial Statistics and Fast New Algorithms

◆ Alexander G. Gray, Carnegie Mellon University

Carnegie Mellon University, Robotics Institute, 5000 Forbes Ave., Pittsburgh, PA 15213

agray@cs.cmu.edu

Key Words: *spatial statistics, n-point correlation, astronomy, algorithm, point processes*

We present algorithms which dramatically reduce the cost of computing many common spatial statistics. The naive computational cost of estimators of second-order properties of spatial point processes, such as the the K function and the second-order intensity function, is typically quadratic in N, the number of points. For higher-order properties, statistics such as the n-point correlation functions have naive cost N raised to the power n. Our algorithms compute exact answers, yet reduce these costs by several orders of magnitude in practice, and are currently being used by cosmologists on astronomical datasets of unprecedented size.

FCAR Modeling of Vector Nonlinear Time Series

Jane L. Harvill, Mississippi State University; ◆ Nibiao Zhang, Mississippi State University

Mississippi State University, Dept. of Mathematics and Statistics, PO Drawer MA, Mississippi State, MS 39762-9715

nz9@ra.msstate.edu

Key Words: *vector nonlinear time series, nonparametric models, forecasting, testing nonlinearity, bootstrap*

Dynamic systems of multiple variables are often nonlinear. We extend the univariate functional coefficient autoregressive model to vector time series and present a bootstrap test of vector time series nonlinearity based on the FCAR model. Predictive properties of the model are presented for univariate and vector time series. We conclude by applying the results to a multichannel gamma ray burst.

Records of Discrete Distributions

◆ Mohammad Ahsanullah, Rider University

Rider University, Dept. of Management Sciences, Lawrenceville, NJ 08648

ahsan@rider.edu

Key Words: *discrete distributions, records, characterizations*

Normal as well as weak records of a sequence of independent and identically distributed random variables taking values on 0,1,2,... are considered. Several distributional properties of these record values are presented. Based on these distributional properties some characterizations of discrete distributions are given.

On Estimation of Pareto Distribution

◆ Mei Ling Huang, Brock University; Jing Tang, Brock University

Brock University, Dept. of Mathematics, St. Catharines, ON L2S 3A1 Canada

mhuang@brocku.ca

Key Words: *quantile, order statistics, weighted empirical distribution function, truncated Pareto distribution*

The paper studies Pareto distribution and truncated Pareto distribution. Several estimation methods for distribution functions and quantiles are discussed. The efficiency functions of these estimators have been derived. Monte Carlo simulation results confirm the theoretical conclusions. Comparisons and suggestions for different methods are also given.

The Reflection Principle Revisited

◆ John L. Stedl, Chicago State University

Chicago State University, College of Business, 95th and King Dr., Chicago, IL 60628

drstedl@yahoo.com

Key Words: *Reflection Principle, random walk, path*

In his classic book on probability, the great William Feller introduced the Reflection Principle. It states that the number of paths from (x,y) to (u,v) which touch or cross the x-axis is equal to the number of paths from the reflected point (x,-y) to (u,v),where (x,y) and (u,v) are integral points in the positive quadrant: u>x>=0, y>0, v>0 . The purpose of this paper is to obtain the probability distribution of the number of times the x-axis is crossed for all paths from (0,0) to (0,2n). The distribution is derived by repeated applications of the Reflection Principle. The mean and standard deviation of this random variable are derived. Finally, a quality control application using this result is presented.

20 Models and Methods for Censored Data and Extreme Value Prediction ▲

Section on Statistics and the Environment

Sunday, August 8, 2:00 pm–3:50 pm

Analyzing Censored Data Based on Autoregressive Models

◆ Jung Wook Park, North Carolina State University; Sujit K. Ghosh, North Carolina State University; Marc G. Genton, North Carolina State University

jwpark@stat.ncsu.edu

Key Words: *censored data, imputation, time series, truncated data*

Time series measurements are often observed with data irregularities, such as truncation (detection limit) or censoring. Practitioners often disregard censored-data cases which often result into biased estimates. We present an attractive remedy for handling censored or truncated data based on a class of autoregressive models. In particular, we introduce an imputation method particularly well-suited to fitting autoregressive models in the presence of censored data. We demonstrate the effectiveness of the technique for a problem common to many time series data and describe its adaptation to several other frequently encounted situations. For pedagogic purposes, our illustration of the approach based on a simulation study is limited to a simple AR(1) truncated data problem, but its potential for use beyond this problem is apparent.

Estimation of Limits of Detection in the Analysis of Environmental Data

◆ Changjian Jiang, Monsanto

Monsanto, 700 Chesterfield Pkwy. W, Chesterfield, MO 63017

changjian.jiang@monsanto.com

Key Words: *Box-Cox transformation, censored data, maximum likelihood, limit of detection, confidence limits*

We present a statistical model for the estimation of the limit of detection in environmental studies, where the mean and standard deviation of the responses for an analyte almost always depend on the actual concentration and normality of the distribution is questionable. The proposed method applies to a set of fortified samples. It assumes the dependence of the means and standard deviations of the responses on the concentration being linear after the Box-Cox transformation on both the responses and the fortification levels with power selections based on the data. Limit of detection and other environmental study measures, expressed as functions of the model parameters, can be estimated by maximizing the likelihood function using the SAS procedure NLMIXED.

Assessing the Effect of Interventions in the Context of Mixture Distributions with Known Detection Limits

◆ Haitao Chu, Johns Hopkins University; Thomas Kensler, Johns Hopkins University; Alvaro Muñoz, Johns Hopkins University

Johns Hopkins University, Bloomberg School of Public Health, 615 N. Wolfe St., E7139, Baltimore, MD 21205

hchu@jhsph.edu

Key Words: *mixture models, maximum likelihood, left-censoring, model selection, bias, bootstrap*

Many quantitative assay measurements of metabolites of environmental toxicants in clinical investigations are subject to left censoring due to values falling below assay detection limits. Moreover, when observations occur in both unexposed individuals and exposed individuals who reflect a mixture of two distributions due to differences in exposure, metabolism, response to intervention, and other factors, the measurements of these biomarkers can be bimodally distributed with an extra spike below the limit of detection. Therefore, estimating the effect of interventions on these biomarkers becomes an important and challenging problem. We present maximum likelihood methods to estimate the effect of intervention in the context of mixture distributions when a large proportion of observations are below the limit of detection. The selection of the number of components of mixture distributions was carried out using both bootstrap-based and cross-validation-based information criterion. We illustrate our methods using data from a randomized clinical trial conducted in Qidong, People's Republic of China.

Linear Prediction of Extreme Values in Environmental Statistics

Corrado Crocetta, Università di Foggia; Nicola Loperfido, Università di Urbino; ◆ Cinzia Franceschini, Università di Foggia

Università di Foggia, Via IV Novembre, 1, Foggia, 71100 Italy

cinziafranceschini@msn.com

Key Words: *minimum, maximum, skew-normal distribution, spatial statistics, pollution*

In enviromental statistics, there is often the need to predict extreme values of given features (such as heat, rainfall, and pollution). When extreme values lead to emergencies, quick updating of predictions is necessary, too. We shall first assume that the underlying process is normal and then relax this assumption by considering the more general skew-normal process. The method is applied to real enviromental data.

Bayesian Inferences on Environmental Exceedances and Their Spatial Locations

◆ Peter F. Craigmile, The Ohio State University; Noel Cressie, The Ohio State University; Thomas J. Santner, The Ohio State University; Youlan Rao, The Ohio State University

The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210

pfc@stat.ohio-state.edu

Key Words: *Bayesian hierarchical models, geostatistics, integrated weighted quantile squared error loss, loss functions, Bayes predictor, spatial statistics*

A frequent problem in environmental science is the prediction of extrema and exceedances. It is well known that Bayesian and empirical-Bayesian predictors based on integrated squared error loss tend to "overshrink" predictions of extrema toward the mean. We propose a new loss function called the integrated weighted quantile squared error loss (IWQSEL) as the basis for prediction of exceedances and their spatial location. The loss function is based on an ordering of the underlying spatial process using a spatially averaged cumulative distribution function. We illustrate this methodology with a Bayesian analysis of surface-nitrogen concentrations in the Chesapeake Bay.

The Power Prior as a Tool for Improving the Benefit Transfer in Environmental Valuation

◆ Ulrike G. Lehr, University of Hohenheim

University of Hohenheim, Institut fuer VWL, 520F, Stuttgart, 70593 Germany

lehr@uni-hohenheim.de

Key Words: *contingent valuation method, Bayesian analysis, power prior, benefit transfer, environmental economics, reclamation*

The contingent valuation method (CVM) is a popular method in environmental economics to elicit people's willingness to pay for a positive change in environmental quality or the prevention of a negative change. However, the most incentive compatible elicitation method yields right- and left-censored survival-type data and requires large samples and expensive surveys to obtain valid results. Therefore, the benefit transfer approach has been suggested, which assigns results from past studies to a new policy site. Unfortunately, most of the suggested methods in the past failed. Recently, environmental economists started to look into using Bayesian methods as a tool for the benefit transfer. This paper uses the power prior for the first time on CVM data to incorporate historical data. Hence, it combines the results of previously done studies with the results of a small on site study. A full-scale CVM study on the reclamation of open pit coal mining areas in East Germany is used as reference data and several datasets are used to construct the power prior. It can be shown that the Bayesian benefit transfer succeeds in reducing survey costs by large and facilitates cost-benefit analysis.

21 Designing and Improving Surveys for Health Policy ▲

Section on Survey Research Methods, Social Statistics Section, Section on Health Policy Statistics

Sunday, August 8, 2:00 pm–3:50 pm

A New Design for the Canadian National Seat Belt Survey

◆ Aline Chouinard, Transport Canada

Transport Canada, 330 Sparks St., Ottawa, ON K1A 0N5 Canada

Chouina@tc.gc.ca

Key Words: *seat belt survey, observational survey, low traffic volume, spatial data*

The National Seat Belt Survey, an observational survey designed in 1978 and conducted almost annually until 2001. Despite the overall national seat belt ge rate of 89.9% ±0.6 in 2001, an average of 40% of all vehicle occupants killed and 21% of those seriously injured are still unbelted at the time of a collision. The National Seat Belt Survey was redesigned in 2001 to include rural communities into the survey. The design comprises several stratification levels and multiple stages of sampling. The stratification levels are the province, the region, and the rural/urban strata. The strata definition not only takes into account the population size of communities but also their population density and proximity (or lack thereof) to a city. An inventory of all the intersections contained within each stratum is obtained from the Canadian Highway Information System. The stages of sampling include the selection of the intersections, the day of the week, and the time of day. Sampling schemes that minimize the risk of sampling intersections with low traffic volumes will be presented. Finally, the estimation model for calculating the seat belt ge rate will be discussed.

Statistics for Global Health

◆ David J. Fitch, Universidad del Valle de Guatemala; Roberto Molina Cruz, Universidad del Valle de Guatemala; Patrick D. Baier, George Washington University

Universidad del Valle de Guatemala, Departmento de Matamatica, Apartado Postal 82, 01901, Guatemala

dfitch@uvg.edu.gt

Key Words: *LQAS, developing countries, decision-taking*

We take a second look at the use of Lot Quality Acceptance Sampling (LQAS) method for evaluating health projects success. As the method was designed for industrial quality control we go over the methodology to evaluate its applicability in assessing health programs in developing countries, especially because the great interest it has for important donor organizations like the United States Agency for International Development (USAID). We review its use, discuss how it is doing, and when it may not be efficiently providing information needed by program administrators. We also propose some other methods we believe may be more appropriate to decision-taking in health programs. We undertake a series of simulation studies to estimate the potential of LQAS and these other methods; e.g., our studies suggest there may be little gain in decision taking from sample sizes greater than 10 or 20. We observe the great potential for statistics for contributing to the health of the world's poor. We invite donor organizations like ID and the statistical community, especially the members of the ASA and the IASS, to work together on better statistical methods to be used in the developing countries.

The Effects of Correcting for Sample Selection Bias in Internet Panel Surveys Based on Random Digit Dialing Sampling When Estimating the Demand for Preventative Health Care

Trudy A. Cameron, University of Oregon; ◆ J.R. DeShazo, University of California, Los Angeles; J. Michael Dennis, Knowledge Networks, Inc.; Rick J. Lee, Knowledge Networks, Inc.

University of California, Los Angeles, 3250 Public Policy Bldg., Los Angeles, CA 90095-1656

deshazo@ucla.edu

Key Words: *sample selection bias, nonresponse adjustment, Knowledge Networks, internet*

When evaluating survey-based demand for preventative health care, this paper will (i) estimate the probabilities of self-selection and retention in the web-enabled research panel, and (ii) develop a statistical correction for differences observed in (i). The data sources are the telephone numbers sampled for the RDD recruitment for the Knowledge Networks research panel, Census long-form information; and sample disposition for cases sampled for a large-sample stated preference study. The authors model (i) presence in the final estimating sample for the stated preference survey, vs. (ii) absence from this sample as a consequence of nonrecruitment to the panel, attrition from the panel prior to targeting the survey, and nonresponse to the invitation to participate in the survey. We will identify systematic variation in "propensity to be present in the estimating sample." Preliminary specifications will generate a various fitted propensity measures (or inverse Mills ratio) to be used in an a theoretic fashion as additional regressors in the multinomial logit model that explains respondent demand for preventative health care.

The Impact of 2000 Census-based Population Controls on Health Estimates in the National Health Interview Survey

◆ Carrie Lynch, National Center for Health Statistics; Van L. Parsons, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 3118, Hyattsville, MD 20782

celynch@cdc.gov

Key Words: *multiple-race, population controls, weights, standard errors*

In 1997 the Office of Management and Budget (OMB) issued revised standards for the collection and presentation of federal data on race and ethnicity. The 2000 Census adopted these standards, and other federal surveys were to comply by January 2003. The National Health Interview Survey (NHIS), which is a major national household survey that monitors the health of the U.S. civilian noninstitutionalized population, has been allowing respondents to select more than one race since 1976. The NHIS also adopted the revised OMB standards with the fielding of the 2000 NHIS, but until 2003, the survey weights were produced using 1990 Census-based population controls and old race categories. This paper describes the procedure of bridging 2000 Census based population controls distributed by multiple-race groups to single-race groups for the development of NHIS weights. This paper also compares selected health estimates using 1990-based weights and 2000-based weights. The results presented in this paper demonstrate that, in practice, the number of significant differences of proportions is negligible. However, the switch to new population controls affects total estimates.

Soap Operas, Primetime Dramas, and Public Health: Analysis of National Consumer Panel Survey Data for Health Communication Planning

◆ William E. Pollard, Centers for Disease Control and Prevention; Vicki Beck, USC Annenberg Norman Lear Center

Centers for Disease Control and Prevention, 1600 Clifton Rd., NE, Mailstop D-42, Atlanta, GA 30345

bdp4@cdc.gov

Key Words: *health surveys, public health*

Audience analysis is an essential component in planning health communication for disease prevention and health promotion. The Centers for Disease Control and Prevention and the USC Annenberg School for Communication provide outreach and tech-

nical assistance to writers and producers of television dramas to provide accurate timely information about disease, injury, and disability. The analysis of national survey data for guiding this communication work will be discussed. The data are consumer mail panel surveys conducted annually from 1999 to 2003 with nationally representative samples of 3,000 to 4,000 respondents per year. Results show that various age groups, education and income levels, and minority groups most at risk for preventable disease are among those with the highest rates of viewing of soap operas and primetime dramas. Furthermore, many of these regular viewers report learning about health from these shows and report taking one or more actions as a result. Methodological issues in the analysis of consumer mail panel data are discussed and the implications for disseminating public health messages are described.

Measuring Self-report Symptoms of Depression in Organ Transplant Patients: Validation of a Multidimensional Model of CES-D Item Data

◆ Irene D. Feurer, Vanderbilt Transplant Center; Hongxia Liu, Vanderbilt School of Nursing and Transplant Center; Kenneth T. Thomas, Vanderbilt Transplant Center; Ronald M. Salomon, Vanderbilt University Medical Center; Theodore Speroff, Vanderbilt University Medical Center; C. Wright Pinson, Vanderbilt Transplant Center

Vanderbilt Transplant Center, 801 Oxford House, Nashville, TN 37232-4753

irene.feurer@vanderbilt.edu

Key Words: *CES-D, instrument scaling, latent variable structure, depression, organ transplantation*

This research evaluated the latent variable structure and implications for scoring a depression (DEP) screening instrument in organ transplant patients. Adult organ transplant patients completed the Center for Epidemiologic Studies Depression Scale (CES-D) and the 36-item Short Form (SF-36). Responses (N=1264) were randomly assigned to exploratory (EXP, n=651) and confirmatory (CON, n=613) samples. Statistical methods included EXP principal component analysis, and CON bifactor and discriminant analysis. An EXP three-component solution that accounted for 54% of the variance was interpreted as somatic symptoms, isolation, and positive affect. A CON bifactor model (60% variance) with a general factor (alpha=.84) and three group factors was superior to a two-group model (chi-square p<.001). A CON discriminant function based on the three factor scores (model p<.001) successfully classified 87% of cases (sensitivity=70%, specificity=93%, chi-square p<.001) in relation to the SF-36 mental component norm (50+/-10). Conclusion: A multidimensional CES-D scoring system is indicated for transplant patients. Funding: AHRQ R03-HS13036 and Roche Laboratories, Inc.

22 Nonparametrics ⌘

IMS

Sunday, August 8, 2:00 pm–3:50 pm

Asymptotic Distributions of Buckley-James Estimator with Discontinuous Error Distributions

◆ Qiqing Yu, SUNY, Binghamton; Fanhui Kong, SUNY, Binghamton

SUNY, Binghamton, Dept. of Mathematical Sciences, Binghamton, NY 13902

qyu@math.binghamton.edu

Key Words: *right-censorship, linear regression model, asymptotic normality, identifiability conditions*

We consider the linear regression model with right-censored data. The Buckley-James estimator (BJE) is the most appropriate extension of the least squares estimator from the complete data case to the right-censored data case. Lai and Ying (1991) establish asymptotic normality of the BJE under a set of regularity conditions, which includes smoothness conditions on the underlying distributions and includes an identifiability condition. We study asymptotic properties of the BJE under an identifiability condition that is weaker than the one in Lai and Ying, and under a set of regularity conditions that does not require continuity. We show that under certain regularity conditions, the BJE has an asymptotic truncated normal distribution.

Nonparametric Functional Mapping of QTL

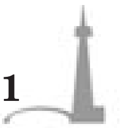
◆ Jie Yang, University of Florida; George Casella, University of Florida

University of Florida, Dept. of Statistics, 103 Griffin/Floyd Hall - PO Box 118545, Gainesville, FL 32611

jyang81@ufl.edu

Key Words: *functional mapping, nonparametric statistics, mixed model, genetic algorithm*

Functional mapping is a powerful tool for detecting major genes responsible for different phenotypic curves, and was first developed by Ma, Casella, and Wu (2002). The methodology uses a parametric functional form, usually derived from a biological law, to drive a maximum-likelihood-based test for a significant QTL (quantitative trait loci). However, in many situations there is no obvious functional form and, in such cases, this strategy will not be optimal. Here we propose to use nonparametric function estimation, typically implemented with B-splines, to estimate the underlying functional form of phenotypic trajectories, and then construct a nonparametric test to find evidence of existing quantitative trait loci. Using the representation of a nonparametric regression as a mixed model, we can easily derive a likelihood ratio test statistic. Using that statistic, we can then calculate the p value directly. Simulation studies show that our method is both powerful and quick, and we also provide an application to a real dataset.



Verification of the Variance Structure in Quasi-likelihood Functions

◆ Steven G. Bai, Pennsylvania State University

Pennsylvania State University, 330A Thomas Bldg., Statistics Dept., University Park, PA 16802

bai@stat.psu.edu

Key Words: *quasi-likelihood, squared integrable estimating equations, method of moment, minimization of asymptotic variance*

The main purpose of this paper is to develop a method for verifying the validity of the variance assumption in a quasi-likelihood. The theoretical statisticians of all persuasions agree in the importance of the role played by the likelihood function in statistical inference. To obtain such likelihood it is necessary to have a probabilistic mechanism for the response. However, the inferences can be hard to draw from experiments in which there is insufficient information to construct a likelihood function. When such situations occur, a quasi-likelihood can be defined based only on the first two moments of the distribution. To use the quasi-likelihood effectively, however, we need to verify our assumptions on these moments. This paper is focused on the verification of the second moment. We assumed that the variance function is in the family of \mathcal{V} = \{ V(\α): \ \α \in \mathbb{R} \}.

Estimation of a CDF and Its Derivative

◆ François Perron, Universitè de Montrèal

Universitè de Montrèal, Dèpartement de math, Universitè de Montrèal, PO Box 6128, Succ. centre-ville, Montrèal, PQ H3C 3J7 Canada

perronf@dms.umontreal.ca

Key Words: *splines, order statistics, nonparametric, CDF, estimation, Hoefdings bounds*

We consider the problem of estimating a CDF F by an estimator G based on a sample of size n. As n increases, we want G to be smooth on the regions where F is absolutely continuous and we want to obtain uniform convergence as well. Our approach uses splines with m nodes and the nodes depends on the sample. We show that our results hold if m tends to infinity as n tends to infinity. When F is absolutely continuous we show that the derivative of G tends to the derivative of F almost surely if m and n/m tend to infinity as n tends to infinity. We give some applications.

Modeling Random Fluctuations in Molecules

◆ Harshinder Singh, West Virginia University

West Virginia University, Health Effects Laboratory Division, NIOSH, Dept. of Statistics, Morgantown, WV 26506-6330

hsingh@stat.wvu.edu

Key Words: *dihedral angles, entropy, molecular dynamic simulations, von Mises Distribution*

Probabilistic modeling and the estimation of entropy of random fluctuations in the internal coordinates of molecules is important for studying their properties and functions. The entropy of a molecule depends mainly on random fluctuations in the dihedral angles of the molecule. Large molecules, such as peptides and proteins, have many dihedral angles and entropy estimation for such molecules is a challenging problem. The traditional approach assumes a multivariate Gaussian distribution for the internal coordinates of molecules. However, the assumption of normality is not valid in many situations. We discuss the circular probability modeling approach for modeling the dihedral angles in molecules. Applications of circular distributions are illustrated using data on the dihedral angles of some molecules. In general, the distributions of coordinates can have arbitrary shapes and we also discuss a nonparametric approach to estimation of entropy of molecules based on kth nearest neighbor distances between n sample points, where k (< n) is a fixed positive integer. Applications of the proposed estimators are illustrated by using them for estimation of entropy of some molecules.

Nonparametric Simultaneous Confidence Bands for Current Status Data

◆ Michael R. Kosorok, University of Wisconsin, Madison

University of Wisconsin, Madison, K6/428 Clinical Science Center, 600 Highland Ave., Madison, WI 53792-4675

kosorok@biostat.wisc.edu

Key Words: *current status data, extreme value distribution, confidence bands, nonparametric maximum likelihood, bootstrap, rate of convergence*

We study the problem of constructing simultaneous confidence bands for the cumulative distrubution function based on nonparametric maximum likelihood estimation (NPMLE) with current status data. Recent results by Hooghiemstra and Lopuhaa (1998) and Groeneboom (1989) shed light on the extremal limit behavior of the uniform norm of the NPMLE minus the true distribution. We verify that this norm, after suitable standardization, converges to the extreme value distribution. We study the uniform rate of convergence to this extreme value distribution, and show that the rate can be significantly improved through use of the bootstrap.

Tree-based Extrapolation Diagnostics for Machine Learning

◆ Giles Hooker, Stanford University

Stanford University, Statistics Dept., Stanford University, CA 94305

gilesh@stanford.edu

Key Words: *CART, diagnostics, extrapolation, tree, machine learning, nonparametric*

Machine-learning applications typically occur in very high-dimensional contexts and usually involve datasets of highly correlated predictors. This results in the hypercube bounding the data containing large unpopulated regions. Moreover, many learning procedures are highly flexible, meaning that their behavior is uncontrolled in these regions. This is a problem when the distribution of predictor variables shifts. It is also a concern for functional diagnostics which often require the function to be evaluated on a product measure. We present a diagnostic for extrapolation as a test statistic for a point originating from the data distribution against a uniform null hypothesis. This allows us to employ general classification methods to estimate this statistic. Further, we observe that CART can be given an exact distribution as an argument to provide a more stable estimate. This is the basis of our extrapolation-detection procedure. We explore some advantages of this approach and present examples of it working in practice.

23 Gross Flow and Calibration ▲

Section on Survey Research Methods

Sunday, August 8, 2:00 pm–3:50 pm

A Look at Several Gross Flows Estimation Methods with the Current Population Survey

◆ Stephen M. Miller, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 4985, Washington DC, 20212

miller_s@bls.gov

Key Words: *gross flows, labor force transitions*

Twenty years ago the Conference on Gross Flows in Labor Force Statistics, jointly sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics, was held to examine estimators of month-to-month change between labor force statuses. The need for such estimates was highlighted by an earlier report by the Levitan Commission. Since that conference several additional estimation methods have been proposed in both the statistics and economics literature. This paper presents an empirical examination of many of the estimators suggested over the last 20 years, applying them to recent data from the Current Population Survey. These estimators are also compared to a particular method under consideration by the Bureau of Labor Statistics.

A Proposed Estimation Procedure for CPS Gross Flows

◆ Edwin L. Robison, Bureau of Labor Statistics

Bureau of Labor Statistics, 1809 Belvedere Blvd., Silver Spring, MD 20902

robison.ed@bls.gov

Key Words: *gross flows, raking, weighting, Current Population Survey*

Gross flows estimates describe the month-to-month transitions from one labor force status to another. The method currently in use for CPS data applies a simple ratio adjustment forcing the weighted sum of matched cases (same person in sample two consecutive months) to duplicate overall male and female control totals for the current month. The current month margin of a gross flows table does not match the current month stock estimate. Subtables do not add up to known population controls for the current month, excepting male and female totals. An alternate gross flows weighting procedure for labor force estimates is proposed. Control totals are created for consecutive months by gender/race/labor force state by summing the full-sample CPS microdata weights for each month; adjustments account for persons flowing in-scope and out-of-scope between months. Matched cases are weighted, then raked and reweighed, to produce estimates that match controls for both months.

Sample Design and Weighting Features of the Safe Schools/Healthy Students Evaluation

◆ Shelton Jones, RTI International; Lei Li, RTI International; Joseph McMichael, RTI International; Harper Gordek, RTI International; Allison Burns, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

smj@rti.org

Key Words: *virtual classes, Common Core of Data, stratified random sampling*

Improving school safety and promoting healthy development among school-aged children requires cooperative efforts with schools, communities, families, law enforcement, and other partnering organizations. The Safe Schools/Healthy Students Initiative funded 97 sites composed of one or more local educational agencies nationwide to support local partnership development and safe school police and program implementation. This evaluation study conducts surveys of students, classroom teachers, principals, school administrators, district coordinators, and other administrators. A sampling frame of virtual classes was constructed based on the Common Core of Data (CCD) database and stratified random sampling was employed to obtain probability samples. Three rounds of surveys were conducted in 2000, 2001 and 2003 to accommodate longitudinal between year comparisons as well as cross-sectional estimation. This paper discusses characteristics of the multisurvey sample design and sample weighting procedures implemented in the evaluation study.

Trimming Extreme Weights in Household Surveys

◆ Benmei Liu, Westat; David L. Ferraro, Westat; J. Michael Brick, Westat; Erin Wilson, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

benmeiliu@westat.com

Key Words: *outliers, sampling weights, mean square error*

Outliers in household surveys usually arise when there are extreme values for the observed characteristics or when the sampling weights for a few of the cases are very large relative to the other cases. Outliers can seriously inflate the variance of the survey estimates, and may be highly influential for the estimates for small subgroups. The literature describes a few weight-trimming procedures that have been developed to handle outliers due to variable weights. However, most of these procedures were developed for use in surveys in which the variability in the weights is caused by weighting adjustments, or else the methods are very survey specific. In many household surveys, these methods may not apply because extreme weights are often a result of differential probabilities of selection rather than weight adjustments. This paper will explore various methods to trim extreme weights where such weights naturally occur due to differential probabilities of selection. We will evaluate the advantages and disadvantages of various trimming techniques when applied to household surveys of this type.

Effect of Collapsing Rows/Columns of Weighting Matrix on Weights

◆ Jay J. Kim, National Center for Health Statistics; Lawrence H. Cox, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

pzk3@cdc.gov

Key Words: *weighting matrix, cell-collapsing, undercoverage, bias, variance, Horvitz-Thompson estimator*

In calculating sample weights, we often have cells with undercoverage problems in some subpopulations. In this situation, the cells in the weighting matrix are collapsed to avoid such coverage problem. The cell-collapsing was discussed previously, but it has never been investigated systematically. We introduce a method by which we can collapse cells (rows or columns) with undercoverage problem and show how cell weight is transfered from one cell to another. We also show the impact of collapsing on the bias and variance of Horvitz-Thompson estimator. We will propose new collapsing strategies which correct for deficiencies of the current approaches.

Estimating Basic Weights for the U.S. Consumer Price Index

◆ Sylvia G. Leaver, Bureau of Labor Statistics; Darin T. Solk, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3655, Washington, DC 20212

Leaver_S@bls.gov

Key Words: *probability sampling, consumer expenditure*

We present findings from current research on quote-level weights used in constructing elementary indexes for the Commodities and Services component of the U.S. Consumer Price Index. These weights are complex in their construction and rely on expenditure estimates from two supporting surveys, the Consumer Expenditure Survey and the Telephone Point of Purchase Survey, both of which contribute to the variability of the quote-level weights. We also present experimental results, including estimates of indexes and standard errors, for a subset of item strata with which we use smoothed estimates of expenditures.

An Alternative to the Principal Person Method for Weighting in the American Community Survey

◆ Keith Albright, U.S. Census Bureau; Alfredo Navarro, U.S. Census Bureau; Mark E. Asiala, U.S. Census Bureau

U.S. Census Bureau, Decennial Statistical Studies Division, Washington, DC 20233

keith.a.albright@census.gov

Key Words: *American Community Survey, New York City Housing Vacancy Survey, weighting, post-stratification, principal person*

The American Community Survey (ACS) is a monthly survey conducted by the Census Bureau that collects demographic and socioeconomic data about households and persons and is intended to replace the decennial census long form. Testing of the ACS began in 1996, and is now in 36 counties. An issue of considerable concern about the ACS estimates is that the estimates of occupied housing units do not equal the estimate of householders and the estimates of married men do not equal the estimates of married women. This is because the ACS currently uses the principal person method for weighting. An alternative to the ACS weighting is the person weighting method used by the New York City Housing Vacancy Survey (NYCHVS). This method will produce estimates of householders that agree with the estimates of households and occupied housing units. Additionally, it will produce estimates of married women and men that agree. This paper examines the effect of the NYCHVS weighting on other ACS housing unit and population estimates. We will first briefly describe the current weighting method for the ACS and describe the NYCHVS method.

24 Statistical Applications in Economics and Finance ▲ ⌘

Business and Economics Statistics Section

Sunday, August 8, 2:00 pm–3:50 pm

On a Simple Econometric Approach for Utility-based Asset Pricing Model

◆ Jack C. Lee, National Chiao Tung University; Cheng-Few Lee, National Chiao Tung University; H.F. Ni, National Chiao Tung University

National Chiao Tung University, Institute of Statistics and Graduate Institute of Finance, Hsinchu, Taiwan

jclee@stat.nctu.edu.tw

Key Words: *Bayesian, Box-Cox transformation, real data, relative risk aversion*

The *Journal of Finance* has published an important paper, "A Simple Econometric Approach for Utility-Based Asset Pricing Model," by Brown and Gibbon (1985). The main purpose of this paper is to extend the research of Brown and Gibbon (1985) and Karson et al.(1995) in estimating the relative risk aversion (RRA) parameter β in utility-based asset pricing model. First, we review the distributions of RRA parameter estimate β(hat). Then, a new method to the distribution of β(hat) is derived, and a Bayesian approach for the inference of β is proposed. Finally, empirical results are presented by using market rate of return and riskless rate data during the period December 1925 through December 2001.

Imminence of Event and Scoring Models in Assessing Portfolio Risk

◆ Timothy H. Lee, CLConsulting

CLConsulting, 595 Croydon Lane, Alpharetta, GA 30022

tlee080551@aol.com

Key Words: *classification, score, performance window*

Statistical classification is a popular method for assessing likelihood of event (e.g., default in payment or response to solicitation) based on each individual's current profiles. A common way of using such an approach is to create scores, discriminants, or similar quantities as bases for classifying individuals into groups. It is sometimes expected that a good score would reflect the imminence of occurrence of event and can separate one group from the other with satisfactory accuracy as well. The following issues were investigated: (i) Do scores that are developed using same performance windows between profile time and occurrence of event outperform scores developed with varying distances? (ii) How can the imminence of event be incorporated with scores for portfolio decision-making? (iii) How to measure model performance incorporating both accuracy of separation and sensitivity to imminence of event. Finally, proposed a dual concordance rate as a way of measuring model performance and discussed the issues of using short-term time series data for classification, in practitioners' perspective. For the analysis, time series data is generated.

Financial Anomaly Detection: A Six Sigma Approach to Detecting Misleading Financials and Financial Decline

◆ Radu Neagu, General Electric Global Research Center; Deniz Senturk, General Electric Global Research Center; Christina LaComb, General Electric Global Research Center; Murat Doganaksoy, General Electric Global Research Center

General Electric Global Research Center, 1 Research Circle, Bldg. K1-4C41A, Niskayuna, NY 12309

neagu@research.ge.com

Key Words: *financial ratios, financial statements, peer grouping, z-scores, six sigma, financial anomaly*

Recently, several large companies have collapsed at the top of their game amidst SEC charges of fraudulent financials. Even companies not engaged in fraudulent activities but suffering financial decline are oftentimes undetectable from the handful of financial measures investors and creditors typically examine. With innovative intelligence extraction techniques, holistic and effective insight into a company's financial health can be achieved. We use six sigma tools and extend their applicability to address the problem of profiling a company's financial performance. This way inconsistencies and/or "warning signs" of misleading financial statements are raised and singled-out for separate analysis. The results of applying these techniques on a sample of 22 publicly held U.S. companies prove that our methods give early notice of misleading financial statements and financial behavior many months ahead of events such as SEC investigations or significant drops in company's stock price. We performed a simulation study to understand the dependence and sensitivity of our findings relative to the size of the sample of companies considered in the study and the main results are presented.

An Introduction to Statistical Applications in Consumer Credit Portfolio Management

◆ Randy Bartlett, Wells Fargo Financial Acceptance

Wells Fargo Financial Acceptance, 515 Plymouth Rd., K-8, Plymouth Meeting, PA 19462

randeroid@aol.com

Key Words: *risk management, marketing, consumer banking, portfolio managment, credit analysis, modeling*

How do consumer banks apply statistical techniques to manage their portfolios? How can they better apply those techniques? What are the misapplications, the missed opportunities, the misunderstandings; just how could we better fit this together? Where do Six Sigma, Total Quality Management, and Cross-Functional Teams belong? We will present one humble view of the big picture and poke fun at ourselves, et al.

Drift Function Modeling and Testing in Continuous Time Interest Rate Models

◆ Myung Suk Kim, Texas A&M University; Suojin Wang, Texas A&M University

Texas A&M University, 3143 TAMU, College Station, TX 77843-3143

myung@stat.tamu.edu

Key Words: *drift function, generalized likelihood test, consistent nonparametric model specification test, wild bootstrap method*

We test statistical models for the drift function using the weekly U.S. Treasury Bill yields data. Parametric linear and nonlinear regression models are applied to the estimation of the drift function derived by Stanton (1997). We investigate the correctness of the assumed drift models using the consistent nonparametric model specification test and the generalized likelihood ratio test. Both of the tests indicate that there is no strong statistical evidence against the assumed nonlinear drift model as well as the assumed linear drift model. While a weak linear trend is plausible, there is no clear statistical evidence of supporting any exclusive parametric linear or nonlinear drift model.

Effect of Exchange-rate Volatility on Trade Volume: Evidences from a Threshold Model

◆ Hui S. Chang, University of Tennessee, Knoxville; Yanhong Zhang, University of Tennessee

University of Tennessee, Knoxville, Dept. of Economics, Knoxville, TN 37996-0550

hchang@utk.edu

Key Words: *exchange rate, grid search, cointegration analysis*

Effects of exchange-rate volatility on tade volume have been studied extensively with diverse results. One drawback of previous studies lies in the assumption of a constant, or linear, effect of exchange-rate volatility on trade volume. In theory, the uncertainty associated with exchange-rate volatility will trigger trading firms' reaction only when volatility reaches certain level. Minor volatility should have little effect on trade volumes. A grid-search procedure is used to find the threshold which triggers the reaction of trading firms to exchange-rate volatility. Based on monthly data for bilateral trade between the U.S. and other six G-7 countries from 1989 to 2002, a GARCH model is used to estimate the volatility variable, and the cointegration analysis is adopted to study the threshold effect of volatility along with the effect of other explanatory variables on trade volume. Results show that most of the control variables have correct signs and are significant statistically. Results also confirm that threshold effect exists for five of the six countries studied.

25 Special Themes, Tools, and Topics to Incorporate in a Statistics Course

Section on Statistical Education

Sunday, August 8, 2:00 pm–3:50 pm

History of Science and Statistical Education: Examples from Fisherian and Pearsonian Schools

◆ Chong Ho Yu, Aries Technology/Cisco Systems

Aries Technology/Cisco Systems, 331 West Musket Place, Chandler, AZ 85248

asumain@yahoo.com.hk

Key Words: *Fisher, Pearson, philosophy, biology, Mendelism, evolution*

Many students share a popular misconception that statistics is a subject-free ethodology derived from invariant and timeless mathematical axioms. It is proposed that statistical education should include a component regarding the aspect of history/philosophy of science. This article will discuss how biological themes and philosophical presumptions drove Karl Pearson and R.A. Fisher to develop their statistical schools. Pearson is pre-occupied with between-group speciation and thus his statistical methods, such as the chi-squared test, are categorical in nature. On the other hand, variation within species plays a central role in Fisher's framework and therefore Fisher's approach, such as partitioning variance, is more quantitative than Pearson's in terms of the measurement scale. In addition, Fisher adopted a philosophy of embracing causal inferences and theoretical entities, such as infinite population and gene, while Pearson disregarded unobservable and insisted upon description of the data at hand. These differences lead to the subsequent divergence of two hypothesis-testing methods, developed by R.A. Fisher and Neyman/E.S. Pearson, son of Karl Pearson, respectively.

Writing-to-learn in Elementary Statistics

◆ Ananda A. Jayawardhana, Pittsburg State University; Cynthia Woodburn, Pittsburg State University

Pittsburg State University, 1701 South Broadway, Pittsburg, KS 66762

ananda@pittstate.edu

Key Words: *writing, statistics*

Pittsburg State University is a regional comprehensive university in southeast Kansas. As a part of the Writing Across the Curriculum (WAC) program, students are required to take two writing-to-learn classes between English Composition and Introduction to Research Writing. A variety of disciplines across campus offer writing-to-learn courses. Over the last several years, both authors have taught mathematics and statistics classes with a writing-to-learn component. We have tried journal writing, formal papers, impromptu writing, informal weekly writing, and formal biweekly writing with double submissions. We plan to explain the expectations of the university, our strategies, and the lessons we learned with special emphasis to elementary statistics. Even though most of our students are not statistics majors we encourage them to learn to think clearly, organize information, analyze data, and present the results.

Using Journal Articles in Intro Stat

◆ John C. Turner, U.S. Naval Academy

U.S. Naval Academy, 572C Holloway Rd., Annapolis, MD 21402

jct@usna.edu

The use of "real" data has been encouraged in introductory statistics courses. In many cases, it is far to easy to skip over the "real world" part and get right to the calculations. I report on my experience with using entire journal articles in introductory courses. There are a number of distinct advantages to this approach. An important advantage is that the students feel that they have learned something they can acutally apply to their field of interest.

Getting Students to Put Their Hearts into Statistical Topics

◆ Michael C. Mosier, Washburn University

Washburn University, Dept. of Mathematics and Statistics, 1700 SW College, Topeka, KS 66621

mike.mosier@washburn.edu

Key Words: *hands-on experience, classroom exercise, teaching sampling distributions, data collection, coverage probability*

In an introductory statistics course, students will generally groan loudly when required to go outside of class and collect some actual data. However, incorporating such a project into the course can have several benefits for the students. It forces them to think about issues involved in the production of quality data, they gain an understanding that "real" data is often messy (unlike most of the datasets they will encounter in the textbook), and finally, the students develop a feeling of ownership of the data. This enhances their interest in what the data has to reveal, and in the proper methods for analyzing it. We suggest a data collection exercise based on heart rate, which is relatively easy to implement, and then show how the resulting data has successfully been used in the classroom to introduce the concepts of the normal distribution, the sampling distribution of sample means, and the coverage probability of confidence intervals. Examples of analyses performed and presentations made using the students' own data will be given, as well as some examples of things that went wrong and sparked some interesting class discussions.

Introducing a Consulting Experience in a First-level Statistics Course

◆ Deborah Lurie, Saint Joseph's University

Saint Joseph's University, 5600 City Ave., Philadelphia, PA 19131

lurie@sju.edu

Key Words: *pedagogy, life sciences, consulting*

Although students in introductory statistics courses are encouraged to analyze real data, often they fail to get a clear understanding of how and why the data were collected and what role the statistician plays in the research process. To help clarify this process, a consulting experience was included in an applied statistics course for undergraduate biology majors. This experience involved a class visit by a physiologist to explain what blood pressure is, how it is measured, and what types of studies are done; a data analysis project, analyzing a published data set on blood pressure; and a written assignment that requires students to explain the data analysis to the physiologist. Being exposed to the entire research process highlights the team approach found in many research centers and the statistician's need for good communication and reasoning skills as well as computational skills. Students enjoyed the project and upon completion demonstrated a deeper understanding of statistical inference.

The Skeleton in Our Closet: Nonsampling Error

WITHDRAWN

◆ Philip Cline, Washington and Lee University

Washington and Lee University, Williams School of Management, Lexington, VA 24450

pcline@wlu.edu

Key Words: *nonsampling error, introductory statistics*

Instructors of introductory statistics tend to teach inference as though the only source of error in an estmate is sampling error. But do any of us know of a survey or even an experiment that has managed to drive nonsampling error to zero? The paper sets out a taxonomy of sources of nonsampling error and suggests a practical way to deal with it in introductory statistics classes.

Introducing Regression Concepts via Relationships that are Thought to be Deterministic

◆ Christopher J. Malone, Winona State University

Winona State University, Dept. of Mathematics & Statistics, PO Box 5838, Winona, MN 55987

cmalone@winona.edu

Key Words: *teaching, regression*

How do you introduce regression concepts? Regression is often introduced with a scatterplot showing the relationship between two variables. This discussion usually moves quickly onto the estimation of the slope and y-intercept which necessitates the need for hypothesis tests and confidence intervals. Regression is more intuitive than this and it should be taught more intuitively. For example, the concept of a residual does not require the estimation of a mean function if one expects a particular functional relationship to be present. Likewise, the concept of R-Squared may precede the estimation of regression parameters. Much of the discussion will be motivated through the Haystack dataset which was initially presented by Ezekiel (1941) and is used extensively in Cook and Weisberg (1999).

26 Small Area Estimation #1 ▲

Section on Government Statistics, Section on Survey Research Methods

Sunday, August 8, 2:00 pm–3:50 pm

Small-area Estimation with Autocorrelated Observations and Stochastic Benchmark Constraints

◆ Richard B. Tiller, Bureau of Labor Statistics; Danny Pfeffermann, Hebrew University and University of Southampton

Bureau of Labor Statistics, Postal Square Bldg., Room 4985, 2 Massachusetts Ave. NE, Washington, DC 20212-0001

Tiller_R@bls.gov

In order to reduce the variances of state labor force estimates derived from the U.S. Current Population Survey (CPS), the Bureau of Labor Statistics (BLS) uses state-space models that are fitted to each of the direct CPS series, independently between the states. The models combine a model for the true population values with a model for the sampling errors. At the end of each calendar year, the model-dependent state estimates are benchmarked to the corresponding CPS annual average. This approach has the disadvantages of no real time benchmarking and instability in the benchmarking process because even the annual CPS averages are subject to relatively large sampling errors due to the high correlations between the monthly estimators. This paper investigates a new approach to benchmarking that constrains the separate monthly model-dependent state estimates in groups of states, (or nationwide) to sum to the corresponding aggregate CPS estimates in real time. The use of this approach requires joint modeling of the direct estimators in several states and adding the sampling errors and the benchmark constraints to the observation equations.

Nonparametric Small-area Estimation Using Penalized Splines

◆ Jean Opsomer, Iowa State University

Iowa State University, Ames, IA 50011

jopsomer@iastate.edu

Key Words: *mixed model, prediction, nonparametric regression*

Penalized spline regression provides a convenient framework for constructing nonparametric small-area estimators. At the population level, the relationship between the variable of interest and continuous auxiliary variables is modeled as a smooth but otherwise unspecified function, while the small-area effects are incorporated in the model through a random effects specification. We present the resulting nonparametric small area estimator and discuss its statistical properties. The methodology is applied to a survey of lakes in the northeastern U.S.

M-quantile Models for Small-area Estimation

◆ Raymond Chambers, University of Southampton; Nikolaos Tzavidis, University of Southampton

University of Southampton, S3RI, Highfield, Southampton, SO17 1BJ UK

rc6@soton.ac.uk

Key Words: *mixed-effects models, weighted least squares, robust inference, quantile regression, influence functions*

Traditionally, model-based small-area estimation relies on mixed-effects (multilevel) models. Here, we investigate the use of M-quantile models for this purpose. Unlike the mixed-effects approach, which models the expected value of the conditional distribution of the response given the covariates, the M-quantile approach models any set of M-quantile (percentile-like) values of this conditional distribution. By doing so, we avoid normality assumptions on the error term as well as imposing modeling restrictions analogous to random intercepts or random slopes. Instead, between-area variability is captured via variation in area-specific M-quantile "scores." The M-quantile approach is illustrated and contrasted with the mixed effects approach using real-life datasets. Results from Monte Carlo simulation studies indicate that M-quantile models may provide an alternative, and in some cases preferable, solution for small-area estimation problems.

A Method of Reconciling Discrepant National and Subnational Employment and Unemployment Estimates

◆ Swamy A.V.B. Paravastu, Bureau of Labor Statistics; Jatinder S. Mehta, Temple University; i-Lok Chang, American University

Bureau of Labor Statistics, Room 3885, Postal Square Building, 2 Massachusetts Ave., NE, Washington, DC 20212

Paravastu_s@bls.gov

Key Words: *small-area estimation, geographical variation, nonlinear regressions, cross-sectional heteroscedasticity, iteratively rescaled generalized least squares, errors-in-variables*

The principal goal of this paper is to improve the estimation methods adopted in the Local Area Unemployment Statistics (LAUS) program within the Bureau of Labor Statistics (BLS). To jointly model all the available data on employment and unemployment for small areas, the estimators applied to these data are considered. Even though some of these estimators are unbiased in probability sampling, the estimates of employment or unemployment provided by these estimators are different, containing different magnitudes of nonresponse and measurement-error biases and nonsampling and sampling errors. This paper develops a method of estimating these biases and errors. It takes a pair of estimates of employment or unemployment for each of several geographical areas and finds a good model of their conditional variations across areas both at a point in time and through time. This model improves one of the pair of estimates and corrects the other estimate for nonresponse and measurement-error biases and for sampling and nonsampling errors. The improved estimate is equal to the corrected estimate. The method's practical behavior is demonstrated on a real dataset.

A Study of Mass Imputation in Small-area Estimation

◆ Nancy Robbins, U.S. Census Bureau; Richard Moore, U.S. Census Bureau

U.S. Census Bureau, Suitland, MD 20746

nancy.l.robbins@census.gov

Key Words: *mass imputation, small-area estimation*

The Survey of Business Owners (SBO), a survey conducted every five years, is the most comprehensive survey source of basic economic statistics on businesses owned by people of Black, Hispanic, Asian-Pacific Islanders, or American Indian-Native Alaskan ancestry and women. It publishes information on the aggregate number, receipts, payroll, and employment of minority-owned sole proprietorships, partnerships, and corporations. While SBO is designed to provide reliable estimates of race/ethnicity at the state by two-digit standard industrial classification, requests are often received for estimates at more detailed geographic and/or industry levels. Direct survey estimators fail for these finer levels of detail because the sample is not representative at these levels. We have proposed a method of mass imputation whereby data is imputed both for nonresponding and nonsampled cases to create a complete universe. The results are encouraging when the estimates from this method are compared to the direct estimates in small areas.

Exploring the Feasibility of Using Small-area Estimation to Estimate Health Behaviors in Remote Areas in Taiwan

◆ Hsing-Yi Chang, National Health Research Institutes

National Health Research Institutes, 2F, #109, Min-Chuan East Rd., Section 6, Taipei, AL 114 Taiwan

hsingyi@nhri.org.tw

Key Words: *small-area estimation, survey, interview, health*

With the increase demand of local information, Taiwan is considering using the method of small-area estimation in a Health Interview Survey. Local governments need data in small areas, especially in remote areas in Taiwan, to design suitable programs or policies. The combined synthetic-regression method will be used. Attention will be given to small area with area-specific effects, including age structure and ethnic effects. Data for this study came from the Health Behavior-KAP survey, which was conducted in 2002. The survey sampled 30,000 subjects older than 15 years. Each county was a stratum. Therefore, 23 strata were used. The dataset had the advantage in providing national estimates as well as local estimates. We will identify key variables in predicting features of interests in health behavior. Then the proportion of the local symptom variables to national value will be estimated and used to predict local values. Once an ideal model is established, the local data will be used to validate the estimated value. We hope to establish a model for estimating health related issues, such as disease prevalence and the risk factors in remote area in Taiwan.

Model-based Approaches for a Land-cover Map Accuracy Assessment of Northeastern Iowa

◆ Ben Skalland, Iowa State University; Tapabrata Maiti, Iowa State University; Sarah Nusser, Iowa State University

Iowa State University, Center for Survey Statistics and Methodology, Department of Statistics, Ames, IA 50011

skalland@iastate.edu

Key Words: *geospatial data, map accuracy, generalized linear model, survey weight*

A complex sample of pixels from a remotely sensed land-cover map of Northeastern Iowa was taken in order to estimate the accuracy of the map. Based on field visits to the sample locations, a binary response variable was created indicating whether or not a pixel was correctly classified by the mapper. This variable can be used to create a sample-based estimate of the overall accuracy of the map. We use this same data to fit a weighted, logistic regression model for pixel-level accuracy, and suggest ways that this model can be useful to a user of a remotely sensed land-cover map.

27 Statistics of the Brain

Biometrics Section, ENAR

Sunday, August 8, 2:00 pm–3:50 pm

A Complex Data Method to Compute fMRI Activation

◆ Daniel B. Rowe, Medical College of Wisconsin; Brent R. Logan, Medical College of Wisconsin

Medical College of Wisconsin, Dept. of Biophysics, 8701 Watertown Plank Rd., Milwaukee, WI 53226

dbrowe@mcw.edu

Key Words: *fMR, MRI, neuroimaging, brain activation, functional imaging*

In functional magnetic resonance imaging, Fourier “image reconstruction” results in complex valued proton spin densities that make up our voxel time course observations. The complex part of the proton spin density is a result of phase errors due to magnetic field unhomogeneities. Nearly all fMRI studies obtain a statistical measure of functional “activation” based on magnitude image time courses. However, the image information is contained in both the real and imaginary parts or in the magnitude and phase. A more accurate model should use the correct distributional specification and all the information contained in the data. A model is presented that uses the original complex form of the data and not just the magnitude. While the two are equivalent for high signal-to-noise ratios, the additional data used to estimate the complex model parameters results in improved poser for low signal to noise ratios.

Flexible Modeling of Correlated Binary Data for Estimation of Neuron Firing Rates and Synchrony between Neurons

◆ Christel Faes, Limburgs Universitair Centrum; Helena M. Geys, Limburgs Universitair Centrum; Marc Aerts, Limburgs Universitair Centrum; Geert Molenberghs, Limburgs Universitair Centrum; Carmen Cadarso-Suarez, University of Santiago de Compostela

Limburgs Universitair Centrum, Center for Statistics, Universitaire Campus, Diepenbeek, 3590 Belgium

christel.faes@luc.ac.be

Key Words: *synchrony, pseudo-likelihood, correlated binary data, association measure*

A fundamental methodology in neurophysiology is eletrophysiology which records the electrical signals produced by individual neurons within the brain of awake-behaving animal. The main goals are to estimate the temporal evolution and effect of covariates on the neuron firing rates, as well as on the synchrony between neurons. Synchrony refers to the observation that in many cases, action potentials emitted from different neurons are emitted at the same time, or very close in time. While multivariate methods of the analysis of continuous outcomes are well understood, multivariate methods for correlated binary data are less developed. A joint model must allow different time- and covariate-depending firing rates for each neuron, and must account for the association between them. The association between neurons might depend on covariates as well. To describe how “synchronous” two spike trains are, a variety of association measures can be used. Focus is on the specification of a flexible marginal model for multivariate correlated binary data together with a pseudo-likelihood estimation approach, to adequately and directly describe the measures of interest.

An Expectation-maximization Algorithm for Analyzing Multicenter Repeated Functional Magnetic Resonance Imaging Data

◆ Kelly H. Zou, Harvard Medical School; Steven D. Pieper, Brigham and Women’s Hospital; Meng Wang, Brigham and Women’s Hospital; Douglas N. Greve, Massachusetts General Hospital; Simon K. Warfield, Brigham and Women’s Hospital; William M. Wells, III, Brigham and Women’s Hospital; Ron Kikinis, Brigham and Women’s Hospital; First Birn, Harvard University

Harvard Medical School, Dept. of Health Care Policy, 180 Longwood Ave., Boston, MA 02115

zou@bwh.harvard.edu

Key Words: *image processing, expectation-maximization algorithm, sensitivity, specificity, functional magnetic resonance imaging, multicenter study*

Functional magnetic resonance imaging (fMRI) has significantly contributed to understanding both normal and diseased human brains. A large amount of variability often exists in the magnitude, spatial distribution, and statistical significance of the resulting fMRI maps due to differences in equipment and other site-specific differences. Therefore, understanding the effect of these differences in a multicenter functional imaging trial, an efficient pooling and comparison mechanism is desirable, particularly due to the costly imaging, demanding tasks, and analytical burden. We have applied and extended a recently developed expectation-maximization (EM) algorithm, namely Simultaneous Truth and Performance Level Estimation (STAPLE), for evaluating multicenter repeated fMRI data derived from the same set of study subjects across all centers. Over repeated runs per visit at each study center, at each given activation threshold, we used STAPLE to calculate a three-dimensional best estimate brain activation map. Based on such estimated voxel-wise “gold standard,” sensitivity, specificity, and predictive values were then estimated using voxel counts.

Extracting Features of T2 Distributions from Magnetic Resonance Images of Human Brain Using Gamma Variate-fitting

◆ Angshuman Saha, General Electric Global Research Center; Sudeshna Adak, General Electric Global Research Center; Tandon Reeti, General Electric Global Research Center; John Schenck, General Electric Global Research Center; Earl Zimmerman, Albany Medical Center

General Electric Global Research Center, GE India Technology Center, EPIP, Phase 2, Hoodi Village, Whitefiled Rd., Bangalore, 560066 India

Angshuman.Saha@ge.com

Key Words: *neuroimaging, Alzheimer's disease, histogram shape features, parametric smoothing, classification*

Excessive iron deposition and increased atrophy in the brain are known to be hallmarks of Alzheimer's disease (AD). Such anomalies leave specific signatures in T2 images from a 3T magnetic resonance scan: excess iron reduces T2 while atrophy filled with cerebrospinal fluid (CSF) increases T2. Thus, compared to controls, T2 distributions for Ads show thicker lower tails (due to iron) and thicker upper tails (due to CSF). These differences can be used to diagnose and monitor AD. In the neuroimaging literature, smoothing T2 distributions with gamma variate curves have been suggested for visual appeal. We extract subject specific parameter estimates from a gamma variate fit of the T2 distribution, which are then used to discriminate Ads from controls. We find that fitting a single gamma to the entire histogram is inadequate. We have developed a method for fitting a three component Exponential-Gamma-Exponential model to describe the lower tail (iron), mid section and upper tail (CSF) that provides much better fit and clinically meaningful parameters. We also show that, some of these T2 distribution shape parameters outperform the more traditional volumetric features.

Resampling for Brain Volumetric Analysis

◆ Wei Zhu, SUNY, Stony Brook; Yeming Ma, National Institutes of Health; Nora D. Volkow, National Institutes of Health; Manlong Rao, SUNY, Stony Brook; Gene-Jack Wang, Brookhaven National Laboratory

SUNY, Stony Brook, Dept. of Applied Math & Statistics, Stony Brook, NY 11794-3600

zhu@ams.sunysb.edu

Key Words: *bootstrap, jackknife, resampling, PET, SPM, imaging*

Brain activation/deactivation volume is as important an indicator of brain functional activities as (de)activation intensity. The widely used Statistical Parametric Mapping software (SPM) reports the (de)activation volume from one condition to the other. However, no statistical test is available in SPM to compare the (de)activation volume between two or more stimuli. There are two approaches to accomplish the statistical comparison of (de)activation volumes. The first approach is highly intuitive. For each subject, one would claim a certain brain region or voxel being activated from baseline to stimulus if it exceeds a certain threshold, and deactivated if it is less than another threshold. The difficulty with this approach lies in the determination of the thresholds, which is usually heuristic and has no statistical basis. We propose to adopt the statistical Resampling method including both the bootstrap method and the jackknife method as a natural solution to the brain volumetric analysis. The major difficulty with resampling is the computational intensity. We have overcome this obstacle by developing the SPM Batch-Mode software in-house.

Classic Poisson Mixture Modeling with Application to Neuroscience

◆ Zhuoxin Sun, University of Pittsburgh; Ori Rosen, University of Pittsburgh; Allan R. Sampson, University of Pittsburgh

University of Pittsburgh, 27th Fl. CL, Dept. of Statistics, Pittsburgh, PA 15260

zhsst3@pitt.edu

Key Words: *mixture models, EM algorithm, mixtures of Poissons*

In some neurological post-mortem tissue studies, each observation can be viewed as coming from one of a pre-specified number of populations where each population corresponds to a possible type of neurons. We employ standard mixture models with pre-specified number of components to model such data. The EM algorithm is implemented to estimate the parameters in the model. Also we provide a discussion of the statistical and computational issues concerning estimation of Poisson mixtures. An application is given to compare schizophrenic and control subjects with regard to the parvalbumin mRNA expression level of the neurons in the prefrontal cortex. Our approach is compared to a more standard approach to analyze such data.

28 Imputation ▲

Section on Survey Research Methods

Sunday, August 8, 2:00 pm–3:50 pm

Imputation Strategy for a Health and Nutrition Survey

◆ Jeffrey M. Gossett, UAMS Pediatrics; Pippa M. Simpson, Arkansas Children's Hospital; Chan-hee Jo, UAMS Pediatrics; Rajiv Goel, UAMS Pediatrics; James G. Parker, UAMS Pediatrics; Margaret Bogle, U.S. Department of Agriculture

UAMS Pediatrics, 1120 Marshall St., Slot 512-43, Little Rock, AR 72202

GossettJeffreyM@uams.edu

Key Words: *imputation, survey*

The FOODS 2000 was a cross-sectional telephone survey conducted in the spring of 2000 in the Lower Mississippi Delta regions of Arkansas, Louisiana, and Mississippi to assess health and nutrition. Missing data is a common problem in survey data. The purpose of the survey was to investigate the relationships between health and nutrition factors adjusting for control variables such as income, race, weight, and nutritional assistance program participation. It is particularly troublesome to have missing control variables. If we use all available cases, the sample size varies considerably. We investigate single and multiple imputation strategies. The FOODS 2000 is a complex weighted sample. We discuss strategies for incorporating predetermined weights in the imputation. This work was funded under the Lower Mississippi Delta Nutrition Intervention Research Initiative, USDA ARS grant # 6251-53000-003-00D.

Practical Suggestions on Rounding in Multiple Imputation

◆ Recai M. Yucel, Institute for Health Policy

Institute for Health Policy, 50 Staniford St., Suite 901, Boston, MA 02115

ryucel@partners.org

Key Words: *rounding, categorical data, missing data, multiple imputation*

In the last decade, substantial progress has been made in the area of missing data. Many statistical methods and their implementations in software products (e.g., Splus 6 "missing" library, SAS PROC MI, SOLAS) have become available for practitioners in a numerous research areas. The key idea underlying most of these methods is to "replace" missing values by random draws from the conditional distribution of the missing data given the observed data. For convenience some methods (e.g., norm module of missing library in Splus 6, SAS PROC MI) impose a multivariate normal distribution on the variables that are incompletely observed. When these variables are not of a "normal" nature but rather categorical, practitioners are often advised to round the imputed value to the nearest integer (or category) that is within the defined region. We provide some practical suggestions for rounding using commonly available software while avoiding potential biases and more efficient results in terms of marginal distribution consistency. We also consider relevant complications in the structure of the variables of interest such as ordinal or nominal variables.

Comparing Estimates and Variances for a Dataset with Missing Values and Multiple Hot-deck Imputations

◆ Nuria Diaz-Tena, Mathematica Policy Research, Inc.; Frank Potter, Mathematica Policy Research, Inc.

ndiaz-tena@mathematica-mpr.com

Key Words: *item nonresponse, imputations, missing values, multiple imputation, hot-deck imputation*

The National Survey of SSI Children and Families (NSCF) collected data about families with children with disabilities who have applied for Supplemental Security Income (SSI) since 1992. The NSCF supports descriptive analysis of the characteristics of current beneficiaries and their families, and studies the effects of welfare reform. The national sample consisted of 8,535 eligible completed interviews, and some interviews had item nonresponse. We imputed data for missing values for some variables. We present an example of how to impute the missing values (imputing the values of the households' cars and associated debts). The household may have one or two cars for the parents, besides a car for the SSI applicant. These data have been imputed by hot-deck imputation procedures taking into account the income, the kind of employment, and all the household reported data. We imputed multiple times using the hot-deck procedure to check if we can detect a between variance among the different imputations. We compared the difference in the estimates between the multiple imputations, as well as the variances of the estimates with missing values, and the multiple complete datasets.

Imputation and Unbiased Estimation: Use of the Centered Predictive Mean Neighborhoods Method

Avinash Singh, RTI International; ◆ Eric Grau, RTI International; Ralph Folsom, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

egrau@rti.org

Key Words: *predictive mean matching, nearest neighbor imputation, multivariate imputation, predictive mean neighborhoods, unbiased estimation*

Methods for determining the predictive distribution for multivariate imputation range between two extremes, both of which are commonly employed in practice: a completely parametric model-based approach, and a completely nonparametric approach such as the nearest neighbor hot-deck (NNHD). A semiparametric middle ground between these two extremes is to fit a series of univariate models and construct a neighborhood based on the vector of predictive means. This is what is done under the predictive mean neighborhoods (PMN) method, a generalization of Rubin's predictive mean matching method. Because the distribution of donors in the PMN neighborhood may not be centered at the recipient's predictive mean, estimators of population means and totals could be biased. To overcome this problem, we propose a modification to PMN which uses sampling weight calibration techniques such as the GEM (generalized exponential model) method of Folsom and Singh to center the empirical distribution from the neighborhood. Empirical results on bias and MSE, based on a simulation study using data from the 2002 National Survey on Drug Use and Health, are presented to compare the centered PMN with other methods.

Potential Methodologies for Count Imputation for the Decennial Census

◆ Richard A. Griffin, U.S. Census Bureau

U.S. Census Bureau, 104, Arrowhead Cir., Lansdale, PA 19446

richard.a.griffin@census.gov

Key Words: *EM algorithm, multinomial distribution, log linear model, spatial models*

Count imputation was used for Census 2000 for housing unit records lacking a status designation of occupied, vacant, or nonexistent, as well as for known occupied units with unknown population count. A "hot-deck" imputation methodology was used to determine donors to be used for donees requiring imputation. Count imputation as implemented for Census 2000 was a deterministic method in that given the census data the imputed values are fixed. Alternative stochastic imputation methods which randomly select imputed values from a distribution could also be used. Since the results of count imputation effect many important uses of the Census such as allocation of congressional seats and revenue distribution, the Census Bureau is conducting research on imputation alternatives to the Census 2000 methodology. This paper presents results from simulations of alternative stochastic imputation methodologies using log linear models on Census 2000 data. These methodologies assume a multinomial distribution and take advantage of the monotone missing data

pattern to produce explicit maximum likelihood estimates by the factored likelihood method for some log linear models.

29 Clinical Trials ⊞

Biopharmaceutical Section

Sunday, August 8, 2:00 pm–3:50 pm

Impact on Error Rates When Futility Criteria is Based on Secondary Endpoint

◆ Akiko Okamoto, Johnson & Johnson Pharmaceutical R&D, LLC; Yiwen Ma, Johnson & Johnson Pharmaceutical R&D, LLC; Surya Mohanty, Johnson & Johnson Pharmaceutical R&D, LLC

Johnson & Johnson Pharmaceutical R&D, LLC, 1125 Trenton-Harbourton Rd., Titusville, NJ 08560

aokamot@prdus.jnj.com

Key Words: *futility, interim analysis, error rate*

Decision to curtail a clinical trial may need to be based on an endpoint other than the primary endpoint. These decision rules will have an impact on the error rate. These rates will be some function of the dependence between the primary endpoint and the decision-making endpoint. The same concept could be extended to a multi-arm trial in curtailing some of the ineffective arms. The operating characteristics for such situations under various relationships between primary endpoint and decision-making endpoint will be discussed.

The Application of Enhanced Parallel Gatekeeping Strategies

◆ Xun Chen, Merck & Co., Inc.; Edmund Luo, Merck & Co., Inc.; Thomas Capizzi, Merck & Co., Inc.

Merck & Co., Inc., RY34-A316, Merck & Co., Rahway, NJ 07065

Xun_chen@merck.com

Key Words: *enhanced procedures, gatekeeping strategy, interrelationships, matched tests*

The parallel "gatekeeping strategy" proposed by Dmitrienko et al. (2003) provides a flexible framework for the pursuit of strong control on studywise Type I error rate. This paper further explores the application of the weighted Simes parallel gatekeeping procedure recommended by Dmitrienko et al. and proposes some modifications to better incorporate the interrelationships of different hypotheses in actual clinical trials and to achieve better power performance. We first propose a simple method to quantitatively control the impact of secondary tests on the testing of primary hypotheses when applying the weighted Simes parallel gatekeeping procedure. We then introduce an idea of matching the gatekeeping relationship between individual primary and secondary tests whenever applicable to more appropriately address the logical relationships between primary and secondary tests. Our simulation study demonstrates that when applicable, the enhanced gatekeeping procedures generally results in more powerful tests than the ordinary parallel gatekeeping procedure in Dmitrienko, et al.

Fallback and Gatekeeping Strategies for Primary and Secondary Endpoints

◆ Brian L. Wiens, Amgen, Inc.; Alexei Dmitrienko, Eli Lilly and Company; Peter H. Westfall, Texas Tech University

Amgen, Inc., m/s 24-3-C, 1 Amgen Center Dr., Thousand Oaks, CA 91320-1799

bwiens@amgen.com

Key Words: *closed test, multiple comparisons, power*

Clinical trials that pre-specify multiple endpoints require strategies to control the Type I error rate. One tactic is to classify endpoints as primary or secondary, but this alone is insufficient. A number of methods have been proposed, including fixed-sequence testing, Bonferroni adjustments, and gatekeeper approaches. A fallback procedure, which uses a fixed-sequence testing approach while always allowing for testing of all hypotheses, will be discussed. Parallel gatekeeper strategies can be used to control the Type I error rate for multiple primary endpoints while allowing consideration of secondary endpoints if at least one primary endpoint is significant. We compare and contrast these two methods. Representing each strategy as a closed testing procedure allows for evaluation and comparison, and for consideration of improvements in power without inflating the Type I error rate. Combining the two strategies is also explored.

Application of Statistical Sampling Principles to Clinical Trial Management

Guy Cohen, Pfizer Inc.; ◆ Cynthia Siu, Roche Pharmaceuticals; Isma Benattia, Wyeth Pharmaceuticals; Darryl Penenberg, Pfizer Pharmaceuticals; Vito Calamia, Pfizer Pharmaceuticals; Joan Rex, Wyeth Pharmaceuticals; Christine Clemente, Pfizer Pharmaceuticals

Roche Pharmaceuticals, 340 Kingland St., Nutley, NJ 07110

Cynthia.Siu@Roche.com

Key Words: *clinical trials, study management, Deming sampling methods*

As clinical trials have become larger, more lengthy, and more complex, due to the nature of the diseases and conditions under study, it is necessary to use statistical sampling principles (Deming's enumerative study framework) in order to ensure efficient clinical trial management and contain spiraling R&D costs and assure quality. Sampling plans and sampling methods can be developed to randomly pre-select investigation sites for auditing and close monitoring (monitoring reports). Similar statistical sampling methods can be utilized to assure study integrity, quality control for data collection, database cleaning, and tables in final study report. This paper will describe a sampling scheme and show how it will avoid pitfalls of trying to achieve 100% quality control while preserving the responsibility required of all new drug sponsors: to ensure that the study meets or exceeds all GCP and all regulatory requirements.

A Case Study of a Group-sequential Response Adaptive Clinical Trial

◆ Roy Tamura, Eli Lilly and Company; Janelle Erickson, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, DC 6152, Indianapolis, IN 46285

tamura_roy_n@lilly.com

Key Words: *adaptive design, clinical trial, play-the-winner rule*

In most clinical trials, the probability of treatment assignment is determined a priori and is fixed for the duration of the trial. A case study of a one-time group-sequential response adaptive clinical trial will be presented. The focus of the talk will be on the rationale for the adaptive trial, the logistics in implementation, and the choice of an allocation rule. Recent work by Rosenberger, et al., (2001) and Hu and Rosenberger (2003) illustrate the trade-off among allocation rules between power and expected number of treatment failures.

On Statistical Properties of QT Correction Methods

◆ Yibin Wang, Novartis Pharmaceuticals; Guohua Pan, Johnson & Johnson Pharmaceutical R&D, LLC; Alfred Balch, Novartis Pharmaceuticals

Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07936

yibin.wang@pharma.novartis.com

Key Words: *QT prolongation, correction for HR, statistical property, QT-RR correlation*

There is an increasing regulatory emphasis on assessing drug-induced QT interval prolongation. Since QT interval is correlated with heart rate (HR), assessment of QT prolongation should be made at a standard HR, resulting in the need to correct QT interval (QTc) for HR. Numerous methods using fixed correction factors have been proposed; however, none can fully remove the QT-HR relationship when there is substantial variability among subjects or subject populations. This study investigates the statistical properties of QT correction methods that use individual-based (QTcI), population-based (QTcP), or fixed (QTcF) correction factors. It is found that, under both the linear and log-linear models for the QT-HR relationship, QTcP and QTcF are biased with VAR(QTcF) < VAR(QTcP) < VAR(QTcI). Furthermore, QTcI is unbiased under the linear model, but biased under the log-linear model. The Type I error may be inflated in an analysis using QTcP or QTcF (or QTc under the log-linear model) as the response. Therefore, HR should be included in such an analysis as a covariate to adjust for the remaining correlation of QTc with HR. This approach is equivalent to one-step analysis that corrects for HR by using the uncorrected QT interval as the response in a model while including HR as a covariate.

The Growing Role of the Biostatistician in Safety Assessment of Clinical Trials for Regulatory Approval

◆ Michael Nessly, Merck & Co., Inc.

Merck & Co., Inc., Clinical Biostatistics, BL3-2, PO Box 4, West Point, PA 19486

nesslym@merck.com

Key Words: *clinical trials, safety analyses, graphical analysis, reference ranges, laboratory toxicity, antiretroviral therapy*

There is a strong perception that issues surrounding safety of approved drugs, biologicals and devices are increasing in scope and scale. Safety data will be under increasingly greater scrutiny in the future. Despite the large quantity of safety data acquired during clinical drug testing, safety data are rarely collected nor analyzed to their fullest potential For simple trial designs, the standard safety analysis approach of tables of adverse experience counts and listings is generally adequate. However, due to complexities of long-term longitudinal clinical trials, biostatisticians are more frequently being engaged in the analysis and review of safety data in regulatory submissions. Simplistic and sometimes automated approaches of tabulation of safety data as crude frequencies are not the most relevant means to characterize the safety profile of a new treatment. There are cases where exploration of all of the data are revealing with respect to potential underlying mechanisms. The use of more involved methodology in context of recent safety initiatives involving antiretroviral therapy will be presented. Requirements for submission safety analysis datasets will be discussed.

30 Introductory Overview Lecture on Proteomics: A New Field for Statistical Involvement

ASA, ENAR, WNAR, SSC, IMS

Sunday, August 8, 4:00 pm–5:50 pm

Proteomics: A New Field for Statistical Involvement

◆ Francoise Seillier-Moiseiwitsch, University of Maryland, Baltimore County

University of Maryland, Baltimore County, Bioinformatics Research Center, Baltimore, MD 21250

seillier@math.umbc.edu

We will describe: the technology involved in two-dimensional gel electrophoresis and mass spectrometry; the steps required to acquire the raw data; features of the raw data; themes common to these two areas of proteomics; normalization; and preprocessing techniques. We will give an overview of the statistical and computational techniques used to identify differentially expressed proteins. Pitfalls will be highlighted with our analyses of published datasets.

Statistical Perspectives of Two-dimensional Gel Analysis

◆ Anindya Roy, University of Maryland, Baltimore County

University of Maryland, Baltimore County, Dept. of Mathematics and Statistics, 1000 Hilltop Cir., Baltimore, MD 21250

anindya@math.umbc.edu

We will describe the technology involved in two-dimensional gel electrophoresis and mass spectrometry; the steps required to acquire the raw data; features of the raw data; themes common to these two areas of proteomics; normalization; and preprocessing techniques. We will give an overview of the statistical and computational techniques used to identify differentially expressed proteins. Pitfalls will be highlighted with our analyses of published datasets.

31 Women's Health: What About It? ▲ ⊞

Section on Health Policy Statistics, Cmte on Women in Statistics, Social Statistics Section, Section on Statistics in Epidemiology

Sunday, August 8, 4:00 pm–5:50 pm

Assessing Tumors and Patient Prognosis in a Context of Tumor Heterogeneity: An Example with Noninvasive Breast Cancer

◆ Judy-Anne Chapman, Henrietta Banting Breast Centre

Henrietta Banting Breast Centre, University of Toronto and Dept. of Statistics and Actuarial Science, University of Waterloo, 11 Dayman Ct., Kitchener, ON N2M 3A1 Canada

jachapma@utstat.toronto.edu

Key Words: *genomics, breast cancer, sample size, image analysis, heterogeneity*

The assessment of tumors is increasingly performed with very small amounts of tissue while there has been a proliferation of the number of biomarkers being assessed. The Kananaskis working group on quantitative methods in tumor heterogeneity established a research framework for assessing genetic markers of tumor progression, in the context of intratumor heterogeneity. The tenets of this framework will be outlined. A case study will be described for the heterogeneous context of noninvasive breast cancer, breast ductal carcinoma in situ (DCIS). Computer image analysis of nuclei within 10 replicates (five ducts in each of two fields) indicated significant heterogeneity within ducts, between ducts, and between fields. This heterogeneity impacted prognostic investigations. There are sample size implications in the domain of number of cells and area of tumor utilized for investigations by image analysis and microdissection.

Do Women with HIV Receive Appropriate Clinical Care in the U.S.?

◆ Jayanti Mukherjee, Bristol-Myers Squibb Company; Yong Yuan, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, Pharmaceutical Research Institute, 5 Research Pkwy., Wallingford, CT 06492

jayanti.mukherjee@bms.com

Key Words: *women, HIV, clinical care, HAART*

The rate of HIV infection among women in the U,S. is increasing, as reported by the CDC. In addition, when compared to men, women have poorer access to care and medications, resulting in higher viral loads and lower CD4 counts. However, very few studies have tried to investigate differences in the rate of Highly Active Anti-Retroviral Therapy (HAART) utilization between males and females. We investigate demographic and social characteristics of women with HIV in the U.S. and accessibility to current standard of clinical care and compare these to similar data for men with HIV. A longitudinal cohort of HIV positive patients (HIVInsight?), who initiated HAART for the first time from 1999-2003, will be included in the analysis. Demographic and socioeconomic characteristics—indicators of health status, medical access, and clinical care—will be compared between men and women and will be used as covariates in multivariate models to investigate the association between gender and time to the first HAART use. Expect to find quality of care gaps for women as compared to men with respect to HIV infection.

Using Time and Space for Managing Care: Nurses' Experiences from Canada and the United States

◆ Pat Armstrong, York University

York University, Vari Hall, 4700 Keele St., Dept. of Sociology, Toronto, ON M3J 1P3 Canada

patarmst@yorku.ca

Key Words: *health care reform, nursing, time and space management*

Major health care reforms are under way in Canada and the United States. In both countries, the reforms are presented as ways of improving care while reducing costs. The emphasis in reforms on both sides of the border has been on managing the labor of the mainly women who provide care and on managing access for those seeking care, the majority of whom are also women. Drawing on interviews with nurses in British Columbia and California, Ontario and New York State, this presentation explores the ways time and space are used to manage the largest occupational category in health care and to manage access to care. The strategies are often visible and based on old means of controlling the workforce as well as on assumptions about women's work. But other, often less visible, strategies are also emerging, reflecting relations in the new economy of care. Whatever the strategies, assumptions about gender and gendered practices play a central role.

Whom Do We Reach? Breast Screening for Women after 70

◆ Leslie A. Gaudette, Health Canada; Robert A. Spasoff, University of Ottawa

Health Canada, Centre for Chronic Disease Prevention and Control, 120 Colonnade Rd., Room 265B, Ottawa, ON K1A 1B4 Canada

leslie_gaudette@hc-sc.gc.ca

Key Words: *screening mammography, evaluation, breast cancer, elderly, modeling, population health*

Established evaluation criteria were used to assess population health impacts of extending population-based breast cancer screening to Canadian women aged 70-79. Experimental evidence and empirical data were reviewed; the MISCAN microsimulation model, adapted to the Canadian population, estimated benefits and harms. Depending upon screening participation level and sojourn time, continuing to screen women aged 70-79 resulted in an estimated 534 to 569 false positive mammograms per 10,000 screens, 22 to 42 additional biopsies, 10 to 25 extra cancers, and 12.2 to 13.7 prevented cancer deaths, and about 30 fewer women diagnosed with stage T2+ cancer. Quality adjustment of life-years gained reduced the benefit by up to 31% to 48% depending upon discount factor and sojourn times. Between 733 and 821 screens are needed to avert one breast cancer death; the gain in life expectancy is about four days per screen. An estimated 459 breast cancer deaths can be prevented per year in Canada if the national target level of 70% screening participation is reached in women aged 70-79. Overall, favorable population health impacts were found for extending breast cancer screening to women aged 70-79.

32 SAMSI Session on Scalability in Statistics and Science ▲ ▦

Statistical and Applied Mathematical Sciences Institute, Section on Statistics and the Environment

Sunday, August 8, 4:00 pm–5:50 pm

Scalability of Models in Data Mining

◆ David Banks, Duke University

Duke University, ISDS Box 90251, Durham, NC 27708

banks@stat.duke.edu

Key Words: *Curse of Dimensionality, large p, small n, clustering, classification, data mining, selection*

The SAMSI data mining year examined a range of problems in the area, from use of unlabeled sample in classification to issues in overcompleteness. A major theme concerned the analysis of high-dimension, low sample-size datasets. This talk reviews various strategies for handling these problems, which are closely related to the Curse of Dimensionality and also relevant to data quality. Our work explores the effect of combining smart algorithms, aggressive feature selection, and combinatorial search. The ideas are illustrated through several examples, and we draw conclusions and give advice for those who must analyze such data.

Multiscale Analysis of High Performance Materials

◆ Ralph C. Smith, North Carolina State University; Alan E. Gelfand, Duke University

North Carolina State University, Dept. of Mathematics, Box 8205, Raleigh, NC 27695

rsmith@eos.ncsu.edu

Key Words: *multiscale, materials, models, controls, hysteresis, SAMSI*

We discuss the SAMSI Program on Multiscale Model Development and Control Design for advanced materials. The focus of the program is the synergistic investigation of deterministic and stochastic issues related to model development, numerical approximation, and control design for piezoceramic, magnetic, shape memory alloy, and ionic polymer compounds to achieve unique design objectives. The modeling component focuses on the development of fundamental energy relations, stochastic homogenization techniques, and quantification of highly variable spatial and temporal scales to provide a framework for characterizing and designing advanced materials. Numerical techniques are developed with the dual goals of achieving the requisite accuracy while providing the efficiency necessary for real-time implementation. The control component focuses on the use of deterministic and stochastic analysis to quantify model and process uncertainties in a manner which facilitates robust control designs. Examples will be drawn from problems arising in deformable mirror design, artificial muscle development, tendon design to minimize earthquake damage, and atomic force microscopy.

Multiscale Statistical Analysis and Modeling of Burstiness in Internet Traffic

◆ J. Stephen Marron, University of North Carolina

University of North Carolina, Statistics and Operations Research Dept., Chapel Hill, NC 27599-3290

marron@email.unc.edu

Key Words: *multiscale, network, internet*

This talk is an overview of research done during the SAMSI program on Network Modelling for the Internet. This work involved many interesting collaborations between statisticians, probabilists and network researchers. Some of these, including the development of new multiscale statistical techniques and new multiscale models motivated by deep data analysis, will be presented.

33 Markov Chain Monte Carlo Algorithms ▦

Section on Bayesian Statistical Science, IMS, Section on Statistical Computing

Sunday, August 8, 4:00 pm–5:50 pm

Simulating Diffusions

◆ Gareth Roberts, Lancaster University

Lancaster University, Dept. of Mathematics and Statistics, Lancaster, LA1 4YF UK

g.o.roberts@lancaster.ac.uk

The talk will show how diffusions (satisfying some regularity conditions) can be simulated exactly without the need for any time discretization. The method is essentially based upon rejection sampling from candidate (and tractable) diffusion candidates. The accepted sample path is produced as a skeleton which can be "filled in" using simple Brownian bridge probabilities to produce values at different times.

A Mixture Representation of the Stationary Distribution

◆ Jim Hobert, University of Florida; Christian Robert, Universite Paris Dauphine

University of Florida, Dept. of Statistics, Griffin-Floyd Hall, Gainesville, FL 32611

jhobert@stat.ufl.edu

When a Markov chain satisfies a minorization condition, its stationary distribution can be represented as an infinite mixture. The distributions in the mixture are associated with the hitting times on an accessible atom introduced via the splitting construction of Athreya and Ney (1978) and Nummelin (1978). This mixture representation is closely related to perfect sampling and has potential applications in Markov chain Monte Carlo.

Perfect Sampling with Non-Markovian Update Functions

◆ Mark L. Huber, Duke University

Duke University, Box 90320, Durham, NC 27713

mhuber@math.duke.edu

Key Words: *perfect sampling, coupling from the past, non-Markovian coupling*

Protocols for perfect sampling such as coupling from the past (CFTP) normally uses couplings described via an update function that is Markovian and unchanging over time. However, CFTP still works with an update function that is both non-Markovian and varies with time. Moreover, employing such update functions can dramatically improve CFTP, making the algorithm both faster and easier to implement. We will illustrate this with several examples on discrete and continuous state spaces.

34 Adaptive Sampling Applications in Homeland Security ▲ ⌘

Section on Statistics in Defense and National Security, Section on Statistical Computing, Social Statistics Section, Cmte on Statisticians in Defense and National Security, Section on Survey Research Methods, Section on Government Statistics

Sunday, August 8, 4:00 pm–5:50 pm

Adaptive Sampling in Network and Spatial Settings

◆ Steven K. Thompson, Pennsylvania State University

Pennsylvania State University, Dept. of Statistics, 326 Thomas Bldg., University Park, PA 16802-2111

skt@stat.psu.edu

Difficult sampling situations sometimes require an adaptive design for effective investigation of a population of interest. Examples include studies of hidden human populations such as injection drug users and others at high risk for HIV/AIDS or hepatitis C, surveys of unevenly distributed ecological populations and natural resources, and surveillance of epidemic outbreak patterns. Rapid assessment and response to bioterrorism incidents may also require an adaptive approach. In spatial settings, adaptive sampling designs include designs that add additional units to the sample in regions of high encountered values. In network settings, adaptive designs include procedures that follow links from interesting units to find new units to add to the sample. Some new, flexible types of adaptive sampling methods will be described.

35 Robust Analysis for Large Datasets ▲

Section on Physical and Engineering Sciences, Section on Statistical Graphics, Section on Quality and Productivity

Sunday, August 8, 4:00 pm–5:50 pm

Robust Mixture Modeling

◆ Geoffrey J. McLachlan, University of Queensland

University of Queensland, Dept. of Mathematics, St. Lucia, Brisbane, 4072 Australia

gjm@maths.uq.edu.au

Key Words: *finite mixture models, EM algorithm, multiresolution kd-trees, t-distributions, mixtures of factor analyzers*

Finite mixture models are being increasingly used to model the distributions of a wide variety of random phenomena and to cluster datasets. We shall focus on the use of normal mixture models to cluster datasets of continuous multivariate data. We shall consider a robust approach to clustering by modeling the data by a mixture of t-distributions. With this t-mixture model-based approach, the normal distribution for each component in the mixture model is embedded in a wider class of elliptically symmetric distributions with an additional parameter called the degrees of freedom. The advantage of the t-mixture model is that, although the number of outliers needed for breakdown is almost the same as with the normal mixture model, the outliers have to be much larger. We also consider the use of the t-distribution for the robust clustering of high-dimensional data via mixtures of factor analyzers. Finally, we consider the robust fitting of normal mixtures using multiresolution kd-trees.

Clustering and Classification Based on the L1 Data Depth

◆ Rebecka J. Jornsten, Rutgers University

Rutgers University, 501 Hill Center, Busch Campus, Dept. of Statistics, Piscataway, NJ 08854

rebecka@stat.rutgers.edu

Key Words: *clustering, classification, data depth, validation, microarray, gene expression*

Clustering and classification are important tasks for the analysis of microarray gene expression data. Classification of tissue samples can be a valuable diagnostic tool for diseases such as cancer. Clustering samples or experiments may lead to the discovery of subclasses of diseases. Clustering can also help identify groups of genes that respond similarly to a set of experimental conditions. In addition to these two tasks it is useful to have validation tools for clustering and classification. Here we focus on the identification of outliers—units that may have been misallocated, or mislabeled, or are not representative of the classes or clusters. We present two new methods: Ddclust and Ddclass, for clustering and classification. These robust nonparametric methods are based on the intuitively simple concept of data depth. We apply the methods to several gene expression and simulated datasets. We also discuss a convenient visualization and validation tool—the Relative Data Depth (ReD) plot.

Robust Regression for Microarray Data Analysis

◆ Arnold J. Stromberg, University of Kentucky; Johanna Hardin, Pomona College; Hanga Galfalvy, New York State Psychiatric Institute; Steven C. Grambow, Duke University Medical Center

University of Kentucky, Dept. of Statistics, 817 Patterson Office Tower, Lexington, KY 40506-0027

astro@ms.uky.edu

Key Words: *microarray, robust, regression*

Experimental designs for microarray studies are becoming more sophisticated as the technology becomes more commonplace. Biological researchers are using least squares regression models to analyze data for each gene. Since outliers are common in microarray data, we investigate the use of MM linear regression estimators which are able to downweight outliers while remaining efficient for normal errors.

36 Software Development in Survey Organizations ▲ ▚

Section on Survey Research Methods, Section on Government Statistics, SSC, Section on Statistical Computing, Social Statistics Section

Sunday, August 8, 4:00 pm–5:50 pm

Software Process Improvement Efforts in the Economic Directorate

◆ Howard Hogan, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Suitland, MD 20746

howard.r.hogan@census.gov

Key Words: *CMM, TSP, PSP*

This presentation will examine the economic current surveys plan to improve the way statistical software is developed to support surveys. It examines the existing culture and the need to bridge the gap between business expertise and software process expertise required to manage organizational and technological change. It is the goal of the Economic Directorate to evolve toward a methodology that integrates software engineering and management excellence with survey research and computing. This presentation discusses the general approach, including the management steps necessary to make the transition toward software process improvement. It looks at areas of progress and challenges ahead that include the limited management and technical resources available. To improve the way software is developed, we need to be cognizant of the factors that affect change and implement strategies to ensure successful implementation. Included will be a discussion of our experience with the Capability Maturity Model (CMM), the Team Software Process (TSP), and the Personal Software Process (PSP).

Agile Component-based Software Engineering over the Survey Life Cycle

◆ Michael S. Tilkin, NORC, University of Chicago

NORC, University of Chicago, 1350 Connecticut Ave., NW, Suite 500, Washington, DC 20036

tilkin-mike@norc.uchicago.edu

Modern surveys require an array of distinct software components that must work together to produce increasingly sophisticated solutions. This paper examines the survey life cycle and discusses the processes, organizational structures, and technologies used by one survey organization to conduct large-scale social science surveys. The approach utilizes a component-based architectural framework and agile software engineering practices to produce software solutions for distinct projects with varying needs and constraints.

The Processing Environment behind a Statistical Program

◆ Claude Poirier, Statistics Canada

Statistics Canada, Generalized System Methods Section, Tunney's Pasture, RHCoats Bldg 11th Floor, Ottawa, ON K1A 0T6 Canada

claude.poirier@statcan.ca

In developing a statistical program, a national institute must also develop the related technical environment, ideally including common tools. This paper examines the strategy of developing such an environment in considering the scope of the common tools, the user requirements, the hardware and software components, the budget issue, the staffing aspects, and the management pressures. In this context, past lessons must be considered in identifying the set of robust, common, and standardized methods which would most likely satisfy a wide range of internal applications. Furthermore, the environment should be reliable, simple, and flexible enough to be expanded through the years. Given that the end-uses of the tools determine the success of such a project, the promotion of the resulting products is addressed. Statistics Canada's positive as well as negative experience offers the basis for this presentation.

37 Gene Expression QTL Analysis

▲ ⌘

WNAR, ENAR

Sunday, August 8, 4:00 pm–5:50 pm

A Comparison of Transcript-centric and Marker-centric Approaches for Mapping Expression Trait Loci

◆ Christina Kendziorski, University of Wisconsin, Madison; Meng Chen, University of Wisconsin, Madison

University of Wisconsin, Madison, 1300 University Ave., 6729 Medical Sciences Center, Dept. of Biostatistics & Medical Informatics, Madison, WI 53726

kendzior@biostat.wisc.edu

Key Words: *experimental design, power calculations, microarrays, quantitative trait loci (QTL) mapping, expression trait loci (ETL) mapping*

The development of statistical methods for mapping quantitative traits has received considerable attention. Effective methods now exist to account for different types of crosses or family structures, different kinds of phenotypes, the presence of multiple genes affecting the trait and their genetic interactions, and the multiple testing issues that arise from tests at many markers. A number of groups have recently applied these QTL methods to the problem of mapping mRNA abundance measurements by considering each individual transcript as a quantitiative trait. However, most QTL mapping methods were developed to address the case where a small number of traits (oftentimes, just one) are being mapped. In expression trait loci mapping, thousands of traits are considered simultaneously and the repeated application of individual tests is not the most efficient strategy. I will present an empirical Bayes modeling approach to enable ETL mapping. The inefficiency of the single trait method and the utility of the proposed method are demonstrated using microarray and genotype data from an F2 mouse cross in a study of diabetes.

Statistical Methods for Constructing Genetic Effect Network between QTL, Gene Expressions, and Phenotypes

◆ Zhao-Bang Zeng, North Carolina State University

North Carolina State University, Dept. of Statistics, Bioinformatics Research Center, Raleigh, NC 29695-7566

zeng@stat.ncsu.edu

Key Words: *gene expression, QTL-mapping, genetic effect network, factor analysis*

Microarray gene expression technology has recently been used in QTL (quantitative trait loci) mapping studies to map QTL that regulate the expression of genes. In this kind of study, mRNA abundance of many genes is measured from tissue samples in a number of segregating individuals by using microarray. Phenotypic values of a few quantitative traits and genome-wide molecular markers are also measured in each segregating individual. With these data, we can analyze the genome-wide association between gene expression profiles and molecular markers to map gene expression QTL (eQTL) as well as the association between quantitative traits and molecular markers (QTL-mapping). We can also build a genome-wide genetic effect network between genomic region variation, expression levels of a set of genes and some quantitative trait phenotypes. This talk will discuss some statistical methods and issues toward to building this comprehensive genetic effect network. A dataset from forest trees will be used to illustrate the methods.

Genetic Networks Derived from Gene Expression Data in Segregating Mouse Populations

◆ Eric E. Schadt, Rosetta Inpharmatics

Rosetta Inpharmatics, 12040 115th Ave. NE, Kirkland, WA 98034

eric_schadt@merck.com

The reconstruction of genetic networks in mammalian systems is one of the primary goals in biological research, especially as such reconstructions relate to elucidating not only common, polygenic human diseases, but living systems in general. I present a statistical procedure for inferring causal relationships between gene expression traits and more classic clinical traits, including complex disease traits. This procedure has been generalized to the gene network reconstruction problem, where naturally occurring genetic variations in segregating mouse populations are used as a source of perturbations to elucidate tissue-specific gene networks. Differences in the extent of genetic control between genders and among four different tissues are highlighted. I also demonstrate that the networks derived from expression data in segregating mouse populations using the novel network reconstruction algorithm are able to capture causal associations between genes that result in increased predictive power, compared to more classically reconstructed networks derived from the same data. This approach to causal inference in large segregating mouse populations over multiple tissues elucidates fundamental aspects of transcriptional control and allows for the objective identification of key drivers of common human diseases.

38 Statistics in the Courtroom

Section on Bayesian Statistical Science, Section on Statistical Consulting

Sunday, August 8, 4:00 pm–5:50 pm

Problems Arising in Statistical Expert Testimony

◆ Joseph L. Gastwirth, George Washington University

George Washington University, Dept. of Statistics, Washington, DC 20052

jlgast@gwu.edu

Key Words: *expert witness, court, law, ethics, consulting, testimony*

The presentation of statistical findings in a courtroom differs from that in the usual scientific talk or article as all witnesses answer questions put to them by the lawyers. Thus, particular findings that you might deem relevant or even essential may not be presented to the court. Also, the statistical expert usually obtains the basic data from the lawyer. Lawyers may not inform you that some types of information exists or may not request some types of information you feel is needed during the discovery process. This talk will describe some of the problems the speaker has encountered when working in the legal setting. Some suggestions for improving the quality of statistical and other scientific evidence provided to courts and regulatory panels will be presented.

Responsibilities of the Statistician as Expert Witness

◆ Joseph B. Kadane, Carnegie Mellon University

Carnegie Mellon University, Dept. of Statistics, Pittsburgh, PA 15213-3890

kadane@stat.cmu.edu

Key Words: *Bayesian, expert witness, responsibilities, ethics, court, consulting*

This Bayesian speaker will address the question of how to handle the variety of ways of doing statistics in the context of the responsibilities of a statistical expert witness. An expert witness is typically hired and paid by one of the sides in a legal dispute, but the fundamental responsibilities of the expert are to the court. The expert is sworn “to tell the truth, the whole truth, and nothing but the truth,” not “only those truths that help my client.” Experts may not be paid on a contingency basis (i.e., with compensation dependent on the outcome of the case). This talk addresses how the resulting tensions may be balanced.

A Lawyer’s View of the Statistical Expert

◆ Steven L. Willborn, University of Nebraska College of Law

University of Nebraska College of Law, 103 Ross McCollum Hall, PO Box 830902, Lincoln, NE 68583-0902

willborn@unl.edu

Key Words: *law, expert witness, court, testimony, reports, consulting*

The author will discuss the role of statistical experts in legal proceedings, the nature of the relationship between the experts and the lawyers, and differences between statistics and law that may cause conflict or confusion. He will explain how the law views the role of experts in legal proceedings including discussion of the qualifications necessary to be an expert, the nature of the questions that may be put to experts, and the types of answers courts and juries expect. He will describe the normal course of legal proceedings including engagement, drafting of a report, deposition, and courtroom testimony. In particular, he will discuss legal strategies and philosophical differences between the law and statistics.

39 TEAMS— Planning the Statistics Education of Future Teachers ⌘

Section on Statistical Education, Section on Quality and Productivity, Advisory Cmte on Teacher Enhancement

Sunday, August 8, 4:00 pm–5:50 pm

TEAMS—Planning the Statistics Education of Future Teachers

◆ W. Robert Stephenson, Iowa State University; ◆ Denise Mewborn, University of Georgia; ◆ Murray H. Siegel, Sam Houston State University; ◆ David Thiel, Southern Nevada Professional Development Program

Southern Nevada Professional Development Program, 515 W. Cheyenne Ave., Suite D, North Las Vegas, NV 89030

dthiel@interact.ccsd.net

Key Words: *K-12 mathematics, teacher preparation*

Statistics is a hot topic in K-12 mathematics! Whether referred to as statistical literacy, quantitative literacy, or data analysis, the topic is one of the key components of the mathematics curriculum. Building a nucleus of teachers who can effectively teach the data analysis called for at the K-12 level heavily depends on cooperation among mathematical scientists, including statisticians, and teacher educators. This need is outlined in the “The Mathematical Education of Teachers (MET) Report,” released in 2001 by CBMS. The ASA has responded to the MET report by supporting an inaugural conference of teacher educators, which was held in October 2003. The Teacher Education: Assessment, Methods and Strategies conference (TEAMS) brought together teams of statisticians, mathematicians, mathematics educators, educators from related disciplines, and experienced teachers. This panel, which is composed of two math educators, a statistician, and an experienced teacher, will present the goals of the inaugural TEAMS conference, particularly in regards to effective cooperation among stakeholders.

40 The Advanced Technology Program: Innovation and Outcomes of a Public/Private Partnership

Section on Government Statistics

Sunday, August 8, 4:00 pm–5:50 pm

The Role of Small Firms in the U.S. R&D and Innovation System: Evidence from the Advanced Technology Program

◆ Gary Anderson, National Institute of Standards and Technology

National Institute of Standards and Technology, Advanced Technology Program, 100 Bureau Dr., Gaithersburg, MD 20899-4710

gary.anderson@nist.gov

Key Words: *Advanced Technology Program, innovation, R&D surveys*

The Advanced Technology Program (ATP) is a federal science and technology program whose mission is to accelerate the development of long-term, risky technologies for broad national benefit through partnerships with the private sector. Small companies have played an active and vital role in these awards. Of the 709 awards given by ATP between 1990 and 2003, 65% have been awarded directly to small companies or to joint venture projects lead by small companies. ATP tracks the technical achievements, dissemination of technical information and commercialization of products and services during the projects lifetime and for six years after completion. The purpose of this paper is twofold. First, using data collected from ATP participants and evaluation data from other ATP sources, this paper examines the relationship between company size and R&D and innovation. This is accomplished by examining a wide array of performance criteria used to assess the program. Second, this paper will use the ATP data on the R&D and innovation by small companies to draw lessons for other statistical agencies engaged in the collection of R&D and innovation data.

Federal R&D Funding—Outcomes of Award Competition in the Advanced Technology Program

◆ Stephen Campbell, National Institute of Standards and Technology; Andrew Wang, National Institute of Standards and Technology

National Institute of Standards and Technology, 100 Bureau Dr., Stop 4710, Gaithersburg, MD 20899

stephen.campbell@nist.gov

Key Words: *federal R&D program, R&D award competition, R&D project funding*

The Advanced Technology Program (ATP) supports innovation in the United States through competitively awarded funding to companies pursuing early-stage, high-risk R&D. Each year, companies propose R&D projects to ATP, and project awards are made on the basis of technical and economic merit through a competitive review process. A special survey was conducted of over 400 companies participating in project proposals to ATP in the year 2000 award competition. This unique survey provides information on both awardees and nonawardees, and allows comparisons on outcomes for the two groups. The survey data are combined with existing archival data to conduct analysis. We analyze characteristics of proposed projects and applicant companies, factors that explain award selection, and outcomes for companies. We present findings on how projects proposed to ATP compare to other company R&D projects on key dimensions such as risk and time horizon; how awarded projects differ from nonawarded proposals; and how the award decision subsequently impacts the company.

Industry R&D Collaboration—Project Performance and Outcomes in the Advanced Technology Program

◆ Andrew Wang, National Institute of Standards and Technology; Stephen Campbell, National Institute of Standards and Technology

National Institute of Standards and Technology, 100 Bureau Dr., Stop 4710, Gaithersburg, MD 20899

andrew.wang@nist.gov

Key Words: *R&D collaboration, R&D project outcomes, federal R&D program*

The Advanced Technology Program (ATP) is a public-private partnership to support technology innovation in the United States through competitive funding awards to companies pursuing high-risk R&D. A key mission of the ATP is to promote collaborative R&D in U.S. industry. Since 1991, ATP has funded nearly 200 research joint ventures involving over 800 companies, universities, and other organizations. A special survey of ATP joint venture participants was conducted to assess the characteristics and outcomes of these ATP-supported R&D collaborations. The survey data are combined with existing archival data to conduct analysis. We analyze characteristics of these joint venture projects, the experience of project participants in the R&D collaboration, and research and commercial outcomes. In particular, we present qualitative perceptual measures of project success reported by participants, quantitative measures of research outcomes such as patent applications, and quantitative measures of commercial outcomes such as product revenues and cost savings.

41 Creative Approaches to Combat the Brutal Environment of Telephone Surveys ▲

Section on Survey Research Methods, Social Statistics Section

Sunday, August 8, 4:00 pm–5:50 pm

Two Advance-letter Experiments to Raise Response Rates

◆ Charles D. Shuttles, Nielsen Media Research; Paul J. Lavrakas, Nielsen Media Research; Jennie Lai, Nielsen Media Research

Nielsen Media Research, 501 Brooker Creek Blvd., Oldsmar, FL 34677

chuck.shuttles@nielsenmedia.com

This paper will present the results of two experimental studies on the use of sending advance letters to mailable RDD households prior to contacting them via telephone. The first test was conducted in August 2003. It tested several different conditions (n = 1,000 in each condition) comparing various amount of cash sent along with an advance letter vs. an advance post card. The advance letter by itself did not raise telephone response rates. However, $1 raised them two percentage points compared to $3 and $5 raising them approx. 10 percentage points. The second experiment will be conducted in February 2004 and will test for the main effects of sending three-levels of cash in an advance letter and sending a promotional flyer, and the interaction of those two factors; each randomly assigned condition will have 10,000 RDD numbers assigned to it. The second experiment will investigate the impact of the treatments on response rates to the RDD telephone survey and tin a subsequent mail survey stage that follows up the RDD survey stage.

Altering the Hold Period for Refusal Conversion Cases in an RDD Survey

◆ Sherman Edwards, Westat; Dave Martin, Westat; Charles DiSogra, University of California, Los Angeles; David Grant, University of California, Los Angeles

Westat, 1650 Research Blvd., Rockville, MD 20850

ShermEdwards@westat.com

Key Words: *RDD survey, call scheduling*

At the 2003 AAPOR conference, Edwards et al. explored the scheduling of calls for refusal conversion in an RDD survey by comparing the day of week and time of day when the refusal occurred and when the first callback after refusal was made. A collateral finding in that research was that the refusal conversion rate increased steadily as the hold period increased up to about three weeks after the initial refusal. However, the length of the hold period was not varied experimentally, so the results may have been due to some factor related to the scheduling operation. The 2003 California Health Interview Survey introduced an experiment to assess the effects of different hold periods. Initial refusals designated for conversion were randomly assigned to hold periods of one, two, or three weeks. To date, more than 30,000 screener refusals have been recontacted, and more than one third have resulted in completed screeners. The analysis will compare the conversion rates, as well as the response rates for the extended interview for those completing the screener, across the three hold-period groups. It will also compare the characteristics.

Optimizing Call Scheduling in an RDD Survey

◆ Whitney E. Murphy, NORC, University of Chicago; Colm O'Muircheartaigh, NORC, University of Chicago; Cynthia Howes, NORC, University of Chicago

NORC, University of Chicago, 55 E. Monroe, Suite 4800, Chicago, IL 60603

murphy-whitney@norcmail.uchicago.edu

Key Words: *call attempts, calling rules, call history*

The Reach 2010 Risk Factor Survey requires a minimum of seven call attempts to be made over a two-week period in order to complete a household screener. For the first two years of the study, calling rules and a detailed schedule were developed to meet these requirements. As the sample for each community was released in batches and sample sizes for later batches were calculated using results from earlier batches, the data collection period for a particular community could sometimes be several months long. Because REACH 2010 collects yearly cross-sectional samples and is interested in charting progress from year to year, it was in the project's best interest to maintain comparability of time periods and compress the data collection period for any one community as much as possible. After analyzing call history data from the first two years of data collection, we decided that reducing the number of days over which the seven calls were made would allow for a shorter overall data collection period for each community without compromising data quality. In the third year of data collection, the calling rules have been modified.

Predicting Wave Nonresponse from Prior Wave Data Quality

◆ Brian J. Meekins, Bureau of Labor Statistics; Roberta L. Sangster, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 1950, Washington, DC 20212

meekins_b@bls.gov

Key Words: *panel attrition, data quality, nonresponse*

Prior research on nonresponse and data quality in panel surveys is largely focused on the effect of reducing unit nonresponse on survey estimates. This paper models panel attrition or subsequent wave nonresponse, from both noncontact and refusal, using aspects of data quality from previous waves of a panel survey. Data from the Telephone Point-of-Purchase Survey (TPOPS) for the years 2001 to 2003 are used in this analysis. Respondents for the TPOPS are initially recruited using an RDD design and are subject to three more waves of interviewing. A total of 35,477 respondents are used. Panel attrition is a problem with about 76% of the initially completed respondents participating in wave 2, decreasing to 68 percent in waves 3 and 4. Measures of data quality include: expenditure rounding, outlet address completeness, number of outlets/unique outlets, duration of interview, and item nonresponse, especially to income and race. Factor analysis is used to create a single factor of data quality. Covariates also include the difficulty of contacting the respondent and the respondent's reluctance to complete the interview on the prior waves.

Modeling the Likelihood of Interviews and Refusals: Using Call History Data to Improve Efficiency of Effort in a National RDD Survey

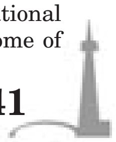
◆ Roberta L. Sangster, Bureau of Labor Statistics; Brian J. Meekins, Bureau of Labor Statistics

Bureau of Labor Statistics, Office of Survey Methods Research, 2 Mass. Ave. NE, Rm 1950, Washington, DC 20212

Sangster_R@bls.gov

Key Words: *nonresponse, call history, RDD survey, call attempts, calling rules*

Proportional hazard models for competing risks are used to develop improved calling strategies. The premise is that completed interview and refusals share similar risks for noncontact. However, the underlying structure may be different. For example, a sampling unit with five answering machine outcomes early in the survey period may be more likely to refuse to participate, while a sampling unit with no answering machine outcomes may be more likely to participate. The number of times an outcome occurs or the proximity between the occurrences may affect the likelihood of certain final outcomes. By modeling the likelihood for refusals and interviews, we hope to find ways to develop calling rules that improves efficiency and optimizes the chance of completion. This study uses call history records for two years of a large national RDD survey (2001-2002). The call records include the outcome of



every call attempted for 125,932 cases (671,482 call attempts). Each survey was conducted over an eight-week period with a maximum number of 30 attempts for each implementation. Ultimately, the survey researchers plan on testing the calling strategies developed in this study.

42 Statistical Thinking and Understanding in the Pre-clinical Pharmaceutical Industry ▲ ⌘

Biopharmaceutical Section

Sunday, August 8, 4:00 pm–5:50 pm

What Good is Statistical Thinking?

◆ Bill Pikounis, Merck & Co., Inc.

Merck & Co., Inc., 126 East Lincoln Ave., Rahway, NJ 07065-0900

v_bill_pikounis@merck.com

In today's business climate, the discovery, basic, and pre-clinical development phases of new chemical entities in pharmaceutical research would seem to be in need of good systems and processes. Biological and measurement variability, for example, need to be recognized and accounted for, so that research is more effective (better drugs) and efficient (faster to the clinic). Certainly as statisticians we feel that we have a rich toolkit to help in these needs. But how much impact can we make given the complexity of research and the culture of scientific laboratories? What is feasible and what is not? We will draw from some of our experiences across the spectrum of discovery, basic research, and pre-clinical development to address these issues.

Statistical Thinking in the "Omics" Era

◆ Alex Varbanov, Procter & Gamble

Procter & Gamble, HCRC, 8700 Mason Montgomery Rd., Mason, OH 45040

varbanov.ar@pg.com

Key Words: *statistical issues, experimental design, genomics, data analysis*

The "omics" era has a significant influence on the preclinical pharmaceutical industry. Modern technologies from areas like genomics, proteomics, and metabonomics are widely used for performing experiments to identify and validate targets for a variety of therapeutic areas. Statistical thinking in the design, analysis, and interpretation of such studies is key to extracting important and correct information from "omics" data. Several statistical issues that are most influential on experimental results and conclusions are presented. This includes experimental design, use of appropriate statistical model, accounting for data noise and multiple comparisons in data analysis, and inferring biochemical networks. These statistical issues are relevant to all "omics" areas.

Generating and Using Trustworthy Data from Plate-Based Assays

◆ James Kenyon, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, 311 Pennington-Rocky Hill Rd., Pennington, NJ 08534

james.kenyon@bms.com

Key Words: *statistical thinking, plate assay, drug discovery, trustworthy data*

Generating and using data from plate-based assays is critical to current drug discovery practices in the pharmaceutical industry. It is used in high-throughput single-dose screens, estimating dose response curves, generating the response for in-silico prediction models, and more. The results of these experiments are used to evaluate toxicity or other adverse effects; desired beneficial activities; methods of transport; and delivery properties, to name a few key areas. Defining and understanding what is meant by "trustworthy" data is essential to all these activities. This includes identifying and evaluating the key characteristics critical to generating or identifying such data. Statistical thinking can be a natural and effective approach to obtaining the keys to success.

Statistical Considerations in Biomarker Method Development and Evaluation

◆ Viswanath Devanarayan, Eli Lilly and Company

Eli Lilly and Company, Lilly Research Laboratories, Lilly Corporate Center, DC 2233, Indianapolis, IN 46285

devan@lilly.com

Key Words: *model selection, assay validation, nonlinear regression, calibration*

An overview of important statistical considerations for some commonly used biomarker methods/assays in pre-clinical and clinical research will be provided. This includes model selection, variability assessment, optimization, performance evaluation and validation, method comparison, parallelism, etc. Careful consideration of the inherent statistical issues can greatly enhance the quality of results and conclusions generated from the biomarker methods. Consequently, there is a growing appreciation in the scientific community for statistical collaboration in these and related topics. Along with some background and overview, the value of applying good statistical thinking and practices in some of these topics will be illustrated with examples.

43 Statistical Genetics: QTL-mapping ▲ ⊞

Biometrics Section, ENAR

Sunday, August 8, 4:00 pm–5:50 pm

A Rank-based Nonparametric Algorithm for QTL-mapping with the Use of Genome-wide Scan Data

◆ Hui Wang, Columbia University; Tian Zheng, Columbia University; Shaw-Hwa Lo, Columbia University

Columbia University, 544 West 113th St. # 2F, New York, NY 10025

hw331@columbia.edu

Key Words: *complex traits, Quantitative Trait Locus, transmission disequilibrium, association, genotype, epistasis*

During the past two decades, a lot of new technologies have been developed in the molecular biology, which enable people to focus their studies based on the whole-genome scans. For example, the identification of large number of marker loci provide much information in the mapping of complex traits, which helps people precisely localize the genes that are disease susceptible. However, since the complex traits involve multiple genes with possible interactions, which is more complicated than simple Mendelian disease, statistical data analysis must be carefully designed to dig out information powerfully and efficiently. We developed a novel method named Quantitative Backward Genotype Transmission Association (QBGTA) algorithm, which used genotype information in the quantitative-trait data to detect those responsible markers and the interaction effects. It is a nonparametric method using the rank information, so it is quite robust to the trait distributions, and fairly powerful, especially for detecting the multiple genes with epistasis.

Regression-based Multivariate Quantitative Trait Locus Mapping

◆ Steven G. Buyske, Rutgers University

Rutgers University, Statistics Dept., Hill Center, 110 Frelinghuysen Rd., Piscataway, NJ 08854

buyske@stat.rutgers.edu

Key Words: *linkage, statistical genetics, genetics*

In cases of highly correlated traits, a multivariate approach to genetic mapping of a quantitative trait locus (QTL) can be superior to univariate approaches. We have developed a regression-based multivariate QTL method that has greater power than the multivariate New Haseman-Elston approach and the multivariate variance components approach. The proposed method has been implemented as an extension to Merlin and is computationally fast.

Fine Mapping of QTL in Complex Pedigrees

◆ Natascha Vukasinovic, Monsanto Animal AG Genomics

Monsanto Animal AG Genomics, 700 Chesterfield Pkwy. W, Mail Stop BB2D, Chesterfield, MO 63017

nvukasin@charter.net

Key Words: *linkage analysis, linkage disequilibrium, quantitative trait loci, pedigree*

To date, two general approaches have been used to map quantitative trait loci (QTL) in human and animal populations: linkage analysis (LA) and linkage disequilibrium (LD) approach. The LA approach utilizes information on recombination events between genetic markers in a pedigree. With the LA approach, the QTL position is estimated with a large error, caused by a small number of recombinations in the genotyped pedigree that commonly comprises only two to three generations. The LD approach utilizes information on historical recombinations, thus providing more precise estimates of the QTL position. However, the LD approach can produce false positive results if LD results from causes other than the mutation at the QTL. Recently, methods that combine advantages of both LA and LD approaches have been proposed. We apply the LA/LD approach in complex multigenerational pedigrees and investigate its efficiency and power for QTL mapping under different scenarios regarding pedigree structure, degree of linkage disequilibrium, and availability of marker information.

QTL-mapping of Gene Expression Levels of a Eucalyptus Backcross

◆ Jessica M. Maia, North Carolina State University

North Carolina State University, 1907 Trexler Ct., Raleigh, NC 27606

jess_maia@yahoo.com

Key Words: *QTL-mapping, quantitative trait, gene expression*

Quantitative trait loci (QTL)-mapping finds genomic regions associated with a quantitative trait such as wood density in trees or hypertension in humans. Gene expression levels within an organism are highly correlated and can also be treated as quantitative traits. My research consists of summarizing thousands of gene expressions levels into fewer components and then finding putative gene regulatory regions associated with them. The data used in this research consists of 2,610 gene expression levels and 18 phenotypic traits measured in 88 trees from a backcross of two eucalyptus strains: eucalyptus grandis and eucalyptus globulus.

Gene Mapping in Recombinant Inbred Lines

◆ Karl W. Broman, Johns Hopkins University

Johns Hopkins University, Bloomberg School of Public Health, Dept. of Biostatistics, 615 North Wolfe St., Baltimore, MD 21205

kbroman@jhsph.edu

Key Words: *statistical genetics, QTL-mapping, linkage analysis, mixed effects models*

Recombinant inbred lines (RILs) are powerful tools for the genetic analysis of complex traits in model organisms, such as the mouse.

RILs are developed by crossing two inbred strains followed by repeated sibling matings, in multiple independent lineages, to produce a new set of inbred strains whose genomes consist of a mosaic of the two parental strains' genomes. RILs are uniquely suited for the coordinated analysis of multiple invasive phenotypes, and for the study of gene-by-environment interactions, as one may phenotype multiple genetically identical individuals. The standard statistical methods for gene mapping in RILs are relatively rudimentary. We explore the use of mixed effects models for the genetic analysis of RILs.

44 Disclosure Limitation and Confidentiality Protection with Microdata Research ▲ ⌘

Section on Government Statistics, Section on Health Policy Statistics, Social Statistics Section

Sunday, August 8, 4:00 pm–5:50 pm

Sharing Confidential Data Among Multiple Agencies Using Multiply Imputed, Synthetic Data

◆ Christine N. Kohnen, Duke University; Jerome Reiter, Duke University

Duke University, Box 90251, Durham, NC 27708-0251

cnk@stat.duke.edu

Key Words: *disclosure limitation, combining rules, multiple imputation, synthetic data*

As the amount of data collected by agencies increases, they may find it beneficial to share their information to create one large database. This new combined database would then provide all involved agencies with a better view of the units recorded; however, such an endeavor would require the cooperation of all involved agencies to disseminate possibly sensitive information. Such a request may not be granted as it could reveal respondents' actual data. One approach to safely sharing their data is for agencies to provide synthetic data rather than real data. This talk will discuss the use of multiple imputation to share data among statistical agencies. We will present inferential methods for combining multiple datasets and will illustrate the approach with simulations.

Full Rank Minimal Sufficient Statistics for Disclosure Limitation and Variance Estimation: A Practical Way to Release Count Information

◆ Yves Thibaudeau, U.S. Census Bureau; William E. Winkler, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop Code 9100, Washington, DC 20233-9100

yves.thibaudeau@census.gov

Key Words: *loglinear model, minimal sufficient statistic, contingency table, disclosure limitation, sampling variance*

Thibaudeau derives a full-rank minimal sufficient statistic (FRMSS) for contingency tables, under the assumption that a hierarchical log-linear model governs the sampling process. The dimensionality of this statistic is equal to the number of degrees of freedom of the model. The objective of the paper is to show that: 1. Selected components of the FRMSS carry much of the information requested by the user, so it is reasonable to release only these components, thereby also limiting information disclosure. 2. The representation of a contingency table by selected FRMSS components lends itself well to the computation of sampling variances for estimators of interest. We conduct simulations to illustrate these results.

Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data

◆ Martha H. Stinson, U.S. Census Bureau; John M. Abowd, Cornell University and U.S. Census Bureau

U.S. Census Bureau, 2A Groves Ave., Alexandria, VA 22305

martha.stinson@census.gov

Key Words: *measurement error, SIPP, administrative data*

We quantify sources of variation in annual job earnings data collected by the Survey of Income and Program Participation (SIPP) to determine how much of the variation is the result of measurement error. Jobs reported in the SIPP are linked to jobs reported in a new administrative database, the Detailed Earnings Records (DER) drawn from the Social Security Administration's Master Earnings File, a universe file of all earnings reported on W-2 tax forms. As a result of the match, each job potentially has two earnings observations per year: survey and administrative. Unlike previous validation studies, both of these earnings measures are viewed as noisy measures of some underlying true amount of annual earnings. While the existence of survey error resulting from respondent mistakes or misinterpretation is widely accepted, the idea that administrative data is also error-prone is new. Possible sources of employer reporting error, employee under-reporting of compensation such as tips, and general differences between how earnings may be reported on tax forms and in surveys, necessitates the discarding of the assumption that administrative data is a "true" measure of the quantity collected.

45 From Genotypes to Expression Arrays: Complex Analysis of Genetic Data

Section on Bayesian Statistical Science

Sunday, August 8, 4:00 pm–5:50 pm

Stochastic Search Gene Suggestion: A Bayesian Hierarchical Model for Gene-mapping

◆ Michael D. Swartz, Rice University, Texas A&M University & University of Texas M.D. Anderson Cancer Center; Marek Kimmel, Rice University; Peter Mueller, University of Texas M.D. Anderson Cancer Center; Christopher I. Amos, University of Texas M.D. Anderson Cancer Center

Texas A&M University & U.T. M.D. Anderson Cancer Center, Dept. of Statistics, 6100 Main MS 138, Houston, TX 77251-1892

mswartz@stat.rice.edu

Key Words: *Bayesian model selection, gene mapping, Markov chain Monte Carlo, Bayesian hierarchical model*

Mapping the genes for a complex disease, such as Rheumatoid Arthritis (RA), involves finding multiple genetic loci that may contribute to the onset of the disease. Pairwise testing of the loci leads to the problem of multiple testing. To avoid multiple tests, we can look at haplotypes; but this results in a contingency table with sparse counts. Using case-parent triad data, we extend the Bayesian conditional logistic regression model developed by Thomas, et al., by defining prior distributions on the allele main effects that model the genetic dependencies present in the HLA region of Chromosome 6. We also added a hierarchical level for model selection that accounts for both locus and allele selection. Thus we cast the problem of identifying genetic loci relevant to the disease into a problem of Bayesian model selection. We evaluate the performance of the procedure with some simulated examples, and then apply our procedure to identifying genetic effects influencing susceptibility to RA. This research is supported by a Genetic Epidemiology Fellowship supported by NCI grant R25-CA57730, Robert Chamberlain, PhD, Principal Investigator.

Sources of Measurement Error in Haplotype-tagging Studies of Complex Disease

◆ Peter Kraft, Harvard School of Public Health

pkraft@hsph.harvard.edu

Key Words: *haplotypes, tag SNPs, effect estimation*

Population-based case-control studies measuring associations between haplotypes of single nucleotide polymorphisms (SNPs) and disease are increasingly popular, in part because haplotypes of a few "tagging" SNPs may serve as surrogates for variation in relatively large sections of the genome. Due to current technological limitations, haplotypes must be inferred from unphased genotypic data. Thus there are two sources of measurement error in haplotype-tagging studies: technical error from imputing haplotypes from unphased multilocus genotypes and structural error arising from the less-than-perfect correlation between the tag-SNP haplotypes and the causal variant. I review the simple and flexible "expectation substitution" approach to inferring and analyzing haplotypes from unphased genotypes. I show via simulation that for relatively common variants with moderate relative risks the technical-error bias in estimates of haplotype effects is negligible when using the expectation-substitution method. However, structural error can lead to a dramatic loss of power, especially if two or more haplotypes carry the causal variant at moderate frequencies.

SNPs, Haplotypes, and Model Selection in Candidate Gene Regions

◆ David Conti, University of Southern California; William Gauderman, University of Southern California

University of Southern California, 1540 Alcazar St. CHP-220, Los Angeles, CA 90033

dconti@usc.edu

Key Words: *SNPs, haplotypes, association analysis, Bayes model averaging*

Modern molecular techniques make discovery of numerous single nucleotide polymorphims (SNPs) in candidate gene regions feasible. Conventional analysis of multilocus data ranges from either independent tests with each variant or the use of haplotypes in association analysis. The first technique ignores the dependencies between SNPs, while the second, though it may increase power, often introduces uncertainty by estimating haplotypes from population data. Additionally, as the number of loci expands, ambiguity in haplotype estimation increases and resolution of the specific causal variant may become problematic. We present a genotype-level analysis to jointly model the SNPs and we introduce a modified SNP¥SNP interaction term to capture the underlying haplotype phase information. By reparameterizing the information from multiple SNPs into linear combinations of SNP and phase terms, we frame the analysis of multilocus data into a model selection paradigm. Within this paradigm, we propose a Bayes model-averaging procedure, which highlights key SNPs and phase terms while incorporating uncertainty in model selection. Prior distributions are modified with genetic information.

A Bayesian Method for Class Discovery and Gene Selection

◆ Mahlet G. Tadesse, University of Pennsylvania; Naijun Sha, University of Texas, El Paso; Marina Vannucci, Texas A&M University

University of Pennsylvania, Dept. of Biostatistics, 423 Guardian Dr., Blockley Hall, Philadelphia, PA 19104-6021

mtadesse@stat.tamu.edu

Key Words: *Bayesian variable selection, clustering, DNA microarray data analysis, mixture models, Markov chain Monte Carlo*

A common goal in DNA microarray data analysis is the discovery of new classes of disease and the identification of relevant genes. We propose a Bayesian method for simultaneously uncovering the cluster structure of the observations and identifying genes that best discriminate the different groups. We formulate the clustering problem in terms of a multivariate normal mixture model with an unknown number of components and use the reversible jump MCMC technique. We handle the problem of selecting a few predictors among the prohibitively large number of variable subsets through the introduction of a binary inclusion/exclusion latent vector and stochastic search methods. We illustrate the methodology with a microarray data from an endometrial cancer study.

Statistical Methods for Haplotype Estimation and Association Studies

◆ Zhaoxia Yu, Rice University

Rice University, 6100 Main St., MS138, Houston, TX 77005

yu@rice.edu

Key Words: *haplotype, block, association, linkage, disequilibrium*

Recently it has been shown that the human genome has a haplotype block structure so that it can be divided into discrete blocks, such that each block contains adjacent loci and only a limited number of haplotypes. This is of great interest due to the potential usefulness in association mapping. Yet, the definition of blocks is not well defined from a statistical point of view and most of the existing definitions are only based on pairwise linkage disequilibrium estimates. I will introduce a new definition which uses not only linkage disequilibrium but also includes other information, such as physical distance. I will also talk about the efficiency it brings together with other factors which affect the power to detect association.

46 NISS/SAMSI on Data-mining

Section on Statistical Computing, Section on Government Statistics, Section on Statistical Graphics, Section on Physical and Engineering Sciences

Sunday, August 8, 4:00 pm–5:50 pm

Bayesian Nonparametric Approach for Analyzing Mass Spectrometry Data

◆ Leanna L. House, Duke University; Merlise Clyde, Duke University

Duke University, 2126 Sprunt Ave., Durham, NC 27705

house@stat.duke.edu

Key Words: *Bayesian, overcomplete, nonparametric, gamma-convolution model, proteomics, MALDI-TOF MS*

Motivated by the science of expression proteomics, we develop a Bayesian nonparametric approach for predicting disease status from high-dimensional proteomic profiles assessed by MALDI-TOF Mass Spectrometry (MS). We develop a nonparametric model for each individual's spectrum through a process convolution, using marked gamma processes. The gamma-convolution model has several desirable features: ensures non-negativity of the modeled protein abundance, allows for shifting of peak locations in the observed spectra, and captures dependencies. Locations of jumps in the gamma process can be used to identify peaks, and associated marks used to identify which peaks are associated with differential protein expression, and predictive of disease status. The gamma-convolution model can be represented as an overcomplete kernel regression model with a Poisson number of components. Using this representation, we capitalize on a reversible jump Markov chain Monte Carlo algorithm to sample from the posterior distribution. Using Bayesian model averaging, we make probabilistic statements concerning the predictions of patient disease status.

Bayesian Sample Sizes for Design and Inference

◆ Bertrand S. Clarke, University of British Columbia/SAMSI

University of British Columbia/SAMSI, 1808 White Pine Dr., Durham, NC 27705

bertrand@stat.duke.edu

Key Words: *sample size, asymptotics, posterior*

Bayesian statistical inference is often based on the posterior density. We use asymptotic expressions for the expected value, under a fixed parameter and under the marginal for the data, of certain functionals of a posterior. The generality of our approach permits us to do sample size calculations and to evaluate the effective size of a sample in terms of independent data. Our approach is not tied to specific parametric families and does not rely on computationally intensive simulation methods. This is joint work with several collaborators.

Comparison of Two Multiple-tree Algorithms on High-throughput Screening Data from Drug Discovery: RandomForest and Partitionator

◆ Katja S. Remlinger, North Carolina State University; Jacqueline M. Hughes-Oliver, North Carolina State University

North Carolina State University, 2304-C Bedford Ave., Dept. of Statistics, Raleigh, NC 27607

ksremlin@stat.ncsu.edu

Key Words: *classification trees, data-mining, dimension reduction, variable importance, accumulation curves*

In drug discovery, large chemical libraries are screened to identify active compounds. Screening the entire library is not very cost- or time-efficient. Methods are needed that can predict the biological activity of a compound based on that compound's chemical structure. Data-mining techniques are good candidates for this task. They perform well on large datasets, and they are very flexible. This paper compares two multiple tree algorithms, RandomForest and Partitionator, on a dataset from drug discovery that was used in the KDD Cup 2001. We first give a brief description of both algorithms, point out their differences, and then compare their performance on the KDD dataset. Both algorithms achieve weighted accuracies on predicting the test set activities that are in the top 5% of all competitor results. Furthermore, we propose three different approaches to define the screening order of test set compounds that are suitable for multiple-tree algorithms.

Characterization of Overcompleteness in Function Approximation

◆ Ernest P. Fokoue, SAMSI

SAMSI, 19 T W. Alexander Dr., PO Box 14006, Research Triangle Park, NC 27709

epf@samsi.info

Key Words: *overcompleteness, function approximation, sparsity, basis set*

In recent years, the use of overcomplete sets has become increasing popular in signal processing, machine learning, statistics and many other fields interested in learning functional dependencies from the data. Overcompleteness has been particularly noted to offer a host of advantages when it comes to sparse function representation. I will present some ideas and results on how to characterize the level of overcompleteness under predictive optimality criterion.

A Bayesian Model for Relating Browsing Behavior to Site Structure on the World Wide Web

◆ Murali Haran, National Institute of Statistical Sciences

National Institute of Statistical Sciences, 19 T W. Alexander Dr., Research Triangle Park, NC 27709

mharan@niss.org

Key Words: *hierarchical Bayes, World Wide Web, clickstream data*

A critical problem for users of the World Wide Web is that many sites are difficult to navigate, hard to use, and have confusing structure. This can be largely attributed to a mismatch between the designer's understanding of the structure for the web site, and actual user behavior. One approach to understanding user behavior is to exploit the rich instrumentation in the online world. Web servers create voluminous log files that record every hit to every page on the site. We propose Bayesian statistical models that use this data to relate visitor transition patterns to site structure. These models help relate site activity to site structure, can correlate site modifications with activity changes, and make stochastic predictions. We illustrate our approach with an example from a commerce web site.

47 Research at EPA-Supported Statistics Centers ▲ ⊞

Section on Statistics and the Environment

Sunday, August 8, 4:00 pm–5:50 pm

Model Selection for Geostatistical Models

◆ Andrew A. Merton, Colorado State University; Jennifer A. Hoeting, Colorado State University

Colorado State University, Fort Collins, CO

merton@stat.colostate.edu

Key Words: *AIC, kriging, autocorrelation function*

We consider the problem of model selection for geospatial data. The importance of accounting for spatial correlation has been discussed in other contexts, but the effect of spatial correlation on the choice of covariates in the model has not been fully explored. We consider kriging for geostatistical models to predict a response at unobserved locations, which involves the fitting of explanatory variables and an autocorrelation function. Spatial correlation is typically ignored in the selection of explanatory variables and this can influence model selection results. To address this problem, we consider the Akaike Information Criterion (AIC) as applied to a geostatistical model. We offer a heuristic derivation of the AIC in this context and provide simulation results that show that using AIC for a geostatistical model is superior to the traditional approach of ignoring spatial correlation in the selection of explanatory variables. An example further demonstrates these ideas.

On the Recovery of Stratospheric Ozone

◆ Serge Guillas, University of Chicago; Michael L. Stein, University of Chicago; Donald Wuebbles, University of Illinois, Urbana-Champaign

University of Chicago, CISES, 5734 S. Ellis Ave., Chicago, IL 60637

guillas@uchicago.edu

Key Words: *ozone, trend analysis*

Since the implementation of the international controls on ozone-depleting chemicals, an important focus in studies of stratospheric ozone has been on the detection of a turnaround in the downward trend. We model the trend using the University of Illinois at Urbana-Champaign 2-D (UIUC 2-D) chemical-transport model of the global atmosphere. We carry out a study in the spectral domain on a cohesive dataset from the SBUV(/2) satellite system at northern midlatitudes. We prove that the model is better at capturing the long-range correlation of the data than assuming a piecewise linear trend. In a second time, we compare several statistical trend models, based on a regression either on a piece-wise linear trend, or on the Effective Equivalent Stratospheric Chlorine (EESC) with or without a constant halocarbon emissions run of the UIUC 2-D model. The fit is better with the EESC approach, and the estimate of the EESC regression coefficient is more precise with the help of the UIUC 2-D model. The smallest number of years necessary to detect a recovery is obtained when the EESC is considered in the analysis.

Estimating the Uncertainty of Space-time Predictions of Ozone in Chicago

◆ Dana Draghicescu, University of Chicago

University of Chicago, 5734 South Ellis Ave., Chicago, IL 60637

draghice@uchicago.edu

Key Words: *space-time covariance functions, space-time kriging, parametric bootstrap, ozone*

Producing reliable space-time maps of air pollutants is central for environmental science. We propose a semiparametric model for space-time prediction of ground-level ozone at the ZIP-code level in metropolitan Chicago. The nonparametric space-time trend and the parametric space-time covariance function of the residual process are estimated from monitoring data. Local space-time interpolation is then used to predict ozone from the same data. Therefore estimation of the prediction errors becomes very challenging. We use simulations and resampling techniques to assess the variability of the space-time kriging predictions. Computational problems generated by the high dimension of the data are also discussed.

Hierarchical Bayesian Models for Seasonal Radio Telemetry Habitat Data

◆ Megan C. Dailey, Colorado State University; Alix I. Gitelman, Oregon State University; Fred L. Ramsey, Oregon State University

Colorado State University, Dept. of Statistics, Fort Collins, CO 80523-1877

dailey@stat.colostate.edu

Key Words: *radio telemetry, habitat selection, Bayesian, persistence*

Radio telemetry data used for habitat selection studies typically consists of a sequence of habitat types for each individual indicating habitat use over time. Existing models for estimating habitat selection probabilities have incorporated covariates in an independent multinomial selections (IMS) model and an extension of the IMS to include a persistence parameter. These models assume that all parameters are fixed through time. However, this may not be a realistic assumption in radio telemetry studies that run through multiple seasons. We extend the IMS and persistence models using a hierarchical Bayesian approach that allows for the selection probabilities, the persistence parameter, or both, to change with season. These extensions are particularly important when movement patterns are expected to be different between seasons, or when availability of a habitat changes throughout the study period due to weather or migration. The models are motivated by radio telemetry data for fish in which seasonal differences are expected and evident in the data.

Bayesian Inferences of Disease Epidemics: Smallpox Outbreaks and Public Policy

◆ Bret D. Elderd, University of Chicago; Vanja Dukic, University of Chicago; Greg Dwyer, University of Chicago

University of Chicago, Center for Integrating Statistical and Environmental Science, 5734 S. Ellis Ave., Chicago, IL 60637

belderd@uchicago.edu

Key Words: *disease epidemics, MCMC, smallpox, SEIR model*

Public-policy debate about the release of smallpox into the general population has centered on whether the protection of the public can be better achieved by mass vaccination or trace vaccination. Mass vaccination inoculates the entire population, whereas trace vaccination inoculates those who have been in contact with infected individuals. Previous studies have used relatively complex deterministic models. These models forecast disease epidemics and subsequently public policy based on a single point estimate of the disease reproductive rate (i.e., the number of newly infected individuals arising from a single infected individual). We present a Bayesian analysis of past smallpox epidemics where the number of individuals who have died in each epidemic is known. For this analysis, we used a Markov chain Monte Carlo simulation coupled with a simple set of differential equations. This analysis provided an estimate of the distribution of the disease reproductive rate rather than a single point estimate. By understanding the variation surrounding the disease's spread, a more informed set of decisions can be reached with regards to the public policy of smallpox inoculation.

48 Failure

Section on Physical and Engineering Sciences

Sunday, August 8, 4:00 pm–5:50 pm

An Updated Bibliography of Accelerated Test Plans

◆ Wayne B. Nelson, Wayne Nelson Statistical Consulting

Wayne Nelson Statistical Consulting, 739 Huntingdon Dr., Schenectady, NY 12309

WNconsult@aol.com

Key Words: *accelerated testing, test plans, bibliography, extrapolation*

This paper provides a current bibliography of accelerated test plans, which will aid practitioners in selecting plans and will stimulate researchers to develop needed plans. Performance of existing plans has been evaluated using analytic theory or simulation. Such plans deal with factors including: life distributions (exponential, Weibull, lognormal, etc.); different life-stress or degradation-stress relationships; types of data censoring (right and interval); types of stress loading (constant, step, ramp, cyclical, stochastic, field use); specimen size and geometry; optimization criteria (minimum variance or determinant, cost, etc.); constraints on the test region or test order of specimens; allocation of specimens to test stress levels. Many plans need to be developed, especially for degradation models. The author welcomes additions to the bibliography and on request will email a current copy as a searchable Word file (*Wnconsult@aol.com*).

Nonparametric Failure Distribution Analysis for Accelerated Degradation Test

◆ Shuen-Lin Jeng, Tunghai University

Tunghai University, No. 181, Sec. 3, Taichung-Kan Rd., Dept. of Statistics, Taichung, 407-04 Taiwan

sljeng@mail.thu.edu.tw

Key Words: *accelerated degradation test, nonparametric failure distribution, pseudo-failure time, random parameter degradation model*

A typical procedure of a life test is to treat the survive observations as censored data. This way, there is an enormous information loss on estimating the failure distribution when the test units have high reliability. Accelerated degradation data could provide information of failure distribution for units that have not failed in the normal ge condition when the degradation measurements are closely related to the failure mechanism. We will first review the existence procedures on finding the failure distribution from degradation data. When the degradation model is nonlinear and several parameters of the model are random, the failure distribution is hard to be calculated and is usually evaluated with numerical methods. An easier but approximately correct way is to use "pseudo-failure times" which is generated from the degradation model. Following the idea of using pseudo-failure times, we suggest a nonparametric procedure to obtain the failure distribution and give comparisons with previous results.

Reliability Characteristics of Failure-time Distribution for Performance Degradation

◆ Suk Joo Bae, University of Tennessee; Paul H. Kvam, Georgia Institute of Technology; Way Kuo, University of Tennessee

University of Tennessee, 406, Ferris Hall, Knoxville, TN 37996

sbae@utk.edu

Key Words: *degradation, failure rate, reliability, bathtub curve*

Degradation modeling has a rich history in manufacturing, especially in the electronics industry. In general, degradation measurements have great potential to improve lifetime data analysis. We investigate the reliability characteristic of lifetime distribution generated from a general degradation model, where some of the model coefficients are random variables. We also seek to characterize the degradation models that lead to particular families of lifetime distribution.

Bayesian Accelerated Degradation Models

◆ Arzu Onar, University of Miami; Fridtjof Thomas, Swedish National Rd. and Transport Research Institute

University of Miami, 406 KE Faculty Wing, Management Science Department, Coral Gables, FL 33129

aonar@miami.edu

Key Words: *mixed effects models, MCMC, power-law model*

This presentation will discuss Bayesian accelerated degradation models, which have been applied to a dataset collected during a designed experiment at the Florida Department of Transportation under accelerated conditions. Acceleration was achieved both via increased temperatures and via the use of a Heavy Vehicle Simulator. The experiment was designed to compare three binder types as well as to investigate the effect of increased temperatures. The analysis presented here utilizes mixed effects models, where the parameter estimates as well as the required distributions are obtained via MCMC. Competing models, which include the frequently encountered power-law model as well as other more complex models suggested by the data, are compared and contrasted both from a statistical and practical perspective. The implications in terms of inference and extrapolation as well as in terms of future designs of such experiments are also discussed.

Enhanced Monte Carlo Estimation of the Probability of Failure in Aircraft Turbine Engines

◆ Peter W. Hovey, University of Dayton

University of Dayton, 300 College Park, Dayton, OH 45431-2316

peter.hovey@notes.udayton.edu

Key Words: *Monte Carlo, Gumbel, aircraft safety*

Aircraft turbine engines must be designed to achieve an extremely high reliability. Current design strategies are focused on achieving a specific probability of failure for the engine. Traditional Monte Carlo techniques require excessive computing time because of the complexity of the finite element calculations that determine when a failure occurs. A new method for analyzing Monte Carlo results based on extreme value theory is discussed that significantly decreases the number of simulations that are required, thus increasing computation speed.

A Bivariate Treatment of Truncated Warranty Data

◆ Jeffrey A. Robinson, General Motors R&D Center; Stefanka Chukova, Victoria University

General Motors R&D Center, Mail Code 480-106-256, Vehicle Development Research Lab, Warren, MI 48090-9055

jeffrey.a.robinson@gm.com

Key Words: *warranty, bivariate, survivor*

This article deals with a type of truncation that occurs with typical automotive warranties. Warranty coverage and the resulting claims data are limited by mileage as well as age. Age is known for all sold vehicles all the time, but mileage is only observed for a vehicle with a claim and only at the time of the claim, i.e., the data are zero-truncated with respect to mileage. An unknown number of vehicles leave coverage due to the mileage constraint, and the number of vehicles eligible to generate a claim at any age-mileage combination is subject to uncertainty. We focus on bivariate treatments of this problem. We review briefly nonparametric procedures for estimating the bivariate cumulative hazard function or related bivariate survivor function. In this context these functions describe the distribution of the age or mileage to the first warranty claim, say, of a given type. We compare the bivariate analyses with univariate results for both age and mileage and discuss the practical similarities and differences between them. Real data examples illustrate the ideas.

The Distribution of the Present Value of a Nonrenewing Pro Rata Warranty

◆ John I. McCool, Pennsylvania State University, Great Valley

Pennsylvania State University, Great Valley, 30 E. Swedesford Rd., Malvern, PA 19444

mpt@psu.edu

Key Words: *Weibull, warranty, present value, life distribution, pro rata*

The value to the purchaser of a limited warranty such as given by tire manufacturers on certain premium grades, is regarded as a random variable. The customer is entitled to a credit or rebate, computed on a pro rata basis if the product fails to achieve a specified lifetime in ge units (e.g., miles). The present value of such a warranty depends on the consumer's ge rate U, the discount rate [α], the warranty period W in ge units and the distribution of life in ge units (e.g., miles) and may determine whether it is worthwhile for the consumer to purchase such a warranty if it is not included in the purchase price or to incur the maintenance expenses required to maintain the validity of the warranty. Under the assumption that the time to failure follows a two parameter Weibull distribution, the distribution of present value is found numerically for selected cases and the relative economics of warranty value are examined as a function of ge rate, discount rate, the warranty period and the life distribution parameters.

49 Topics in Quality Control ◬

Section on Quality and Productivity

Sunday, August 8, 4:00 pm–5:50 pm

Diagnosing Shifts the Second and Higher Moments in Multivariate Changepoint Settings

◆ Joe H. Sullivan, Mississippi State University; Zachary G. Stoumbos, Rutgers University; Robert L. Mason, Southwest Research Institute; John C. Young, McNeese State University

Mississippi State University, Box 9582, MSU, MS 39759

jsullivan@cobilan.msstate.edu

Key Words: *covariance matrix, multivariate gamma distribution, statistical process control*

In multivariate changepoint and statistical process control situations, primary emphasis is on detection of a shift in the distribution of random observations. Also important is the diagnosis of what parameter or parameters have changed in the usual parametric setting. We parameterize the multivariate gamma distribution by the mean vector, standard deviations, and correlations, in contrast with the usual parameterization by shape and scale, for easier interpretation. We diagnose shifts in any combination of elements of the correlations, standard deviations, and means.

Methods for Estimating the Time of a Shift in the Mean for Multivariate Data

◆ Steven E. Rigdon, Southern Illinois University, Edwardsville; Nicole J. Munden, Southern Illinois University, Edwardsville

Southern Illinois University, Edwardsville, Dept. of Mathematics & Statistics, Edwardsville, IL 62026

srigdon@siue.edu

Key Words: *multivariate quality control, changepoint, EWMA*

Suppose that a multivariate control chart, such as the Hotelling T-squared chart or the multivariate EWMA chart, signals that the mean vector has changed. We would like to give estimates for the time when the change occurred as well as the amount of the shift. We derive the maximum likelihood estimate for the time of the shift. Another set of estimates is obtained by considering the multivariate EWMA control chart. We compare these various estimates of the time of the shift through simulation.

Multivariate Process Control Under Regulatory Variables

◆ Amitava Mitra, Auburn University

Auburn University, College of Business, Office of the Dean, Suite 516, Auburn, AL 36849-5240

mitra@business.auburn.edu

Key Words: *multivariate process control, influential observations*

In most consumer products, multiple characteristics may have to be monitored to meet customer requirements. Further, since the product is created by a process, it is desirable to identify the operational level of some of the process regulatory variables. The impact of these regulatory variables, as well as that due to unknown factors, influence the product characteristics. We identify the impact of these two sources of variability and propose a scheme to analyze out-of-control conditions.

On a Statistic to Assess the Randomness of Stability

◆ Kevin Anderson, Intel Corporation; Russ Sype, Intel Corporation

Intel Corporation, 1600 Rio Rancho Blvd., M/S RR5-454, Rio Rancho, NM 87124-1092

kevin.c.anderson@intel.com

Key Words: *nonparametric, SPC, stability, random*

Process control is a critical enabler of modern semiconductor manufacturing. It is not uncommon for a single semiconductor fabrication facility to maintain over 75,000 control charts on product, process, equipment, incoming material, and environmental characteristics. Manufacturers have committed resources to Advanced Process Control (APC) on their equipment to automate this control labor. APC implementation is justified only when a process's data are serially correlated. Some important questions semiconductor engineers and statisticians ask: Is this process stable? Would this process profit from APC? Is the APC system functioning properly? The answers to these questions hinge upon the degree of randomness in the attendant data. Procedures for testing randomness have been developed, and are appealing due to their familiarity and ease of computation. Unfortunately, these tests are also sensitive to non-normality. This presentation will describe the development of a rank-based procedure for testing for i.i.d. data, and will detail its use in the assessment of stability and randomness, providing case studies of its application in semiconductor manufacturing.

Inferences on the Parameters and System Reliability for a Failure-Truncated Power Law Process: A Bayesian Approach Using a Changepoint

◆ Mary Richardson, Grand Valley State University; Asit Basu, University of Missouri, Columbia

Grand Valley State University, Dept. of Statistics, 2309 MAK, Allendale, MI 49401

richamar@gvsu.edu

Key Words: *power-law process, changepoint*

The reliability of a repairable system that is either improving or deteriorating depends on the system's chronological age. If such a system undergoes "minimal repair" at the occurrence of each failure so that the rate of system failures is not disturbed by the repair, then a nonhomogeneous Poisson process (NHPP) may be used to model the "age-dependent" reliability of the system. The power-law process (PLP) is a model within the class of NHPP models and is a commonly used model for describing the failure times of a repairable system. We introduce a new model that is an extension of the PLP model: the power law process changepoint model. This model is capable of describing the failure times of particular types of repairable systems that experience a single change in their rates of occurrence of failures. Bayesian inference procedures for this model are developed.

A Statistical Test for Compatibility of Two Software Usage Environments

◆ Daniel R. Jeske, University of California, Riverside

University of California, Riverside, Dept. of Statistics, Riverside, CA 92521

daniel.jeske@ucr.edu

Key Words: *software reliability, score test, operational profile*

An estimate of the user-perceived failure rate of a software system is typically obtained by extrapolating an analysis of failure data collected during the test and debug phase. An important underlying assumption that validates the extrapolation is that the software was tested in a manner that emulates how the users will interact with it. Testing of this type is referred to as operational profile testing. Operational profile testing is difficult to do in practice, and consequently the extrapolated failure rates frequently have limited usefulness. A partial score test is developed for testing the hypothesis that the test and user environments are identical. Application of the test will provide an early indication in the user phase that the that failure rates extrapolated from the test phase are not valid. It is shown how the failure rates can be recalibrated if the hypothesis is rejected. The size and power of the hypothesis test are discussed. The underlying software reliability model is assumed to be the nonhomogeneous Poisson process proposed by Goel and Okumoto. Extensions to other nonhomogeneous Poisson processes are discussed.

Bayesian Modeling of Accelerated Life Tests with Random Effects

◆ Ramon V. Leon, University of Tennessee; Avery Ashby, University of Tennessee; Jayanth Thyagarajan, University of Tennessee

University of Tennessee, 331 Stokely Management Center, Knoxville, TN 37996-0532

rleon@utk.edu

Key Words: *MCMC, WinBUGS, credibility interval, prediction interval, quantile*

We show how to use Bayesian modeling to analyze data from an accelerated life test where the test units come from different groups (such as batches) and the group effect is random and significant. Our approach can handle multiple random effects and several accelerating factors. However, we present our approach on the basis on an important application concerning pressure vessels wrapped in Kevlar 49 fibers where the fibers of each vessel comes from a single spool and the spool effect is random. We show how Bayesian modeling using Markov chain Monte Carlo methods can be used to easily answer questions of interest in accelerated life tests with random effects that are not easily answered with more traditional methods. For example, we can predict the lifetime of a pressure vessel wound with a Kevlar 49 fiber either from a spool used in the accelerated life test or from another random spool from the population of spools. We comment on the implications that this analysis has on the estimates of reliability (and safety) for the space shuttle, which has a system of 22 such pressure vessels.

50 Nonlinear and Generalized Linear Models

ENAR, Biometrics Section, General Methodology

Sunday, August 8, 4:00 pm–5:50 pm

Practical Considerations for Constructing Confidence Intervals for Nonlinear Functions

◆ Paul E. Duty, University of Missouri; Nancy Flournoy, University of Missouri

University of Missouri, 146 Middlebush, Dept. of Statistics, Columbia, MO 65211

paulduty@aol.com

Key Words: *Delta method, bootstrap, geometry, life expectancy, logistic model, quantal estimation*

Confidence intervals for nonlinear functions often suffer from undercoverage. We compare the Delta method, the transformation method, the bootstrap method, and a strategy we call the "composition method" for some specific problems. Using the geometry of the problem, we present criteria for selecting and evaluating which method of constructing confidence intervals is better in a given situation.

A Partial Coefficient of Determination for Generalized Linear Models

◆ Sundar Natarajan, NYU School of Medicine and the VA New York Harbor Healthcare System; Stuart R. Lipsitz, Medical University of South Carolina

NYU School of Medicine and the VA New York Harbor Healthcare System, 423 East 23rd St., Room 11101-S, New York, NY 10010

sundar.natarajan@med.nyu.edu

Key Words: *generalized linear models, partial correlation*

In a regression setting, the partial coefficient of determination is used as a measure of "standardized" partial association between an outcome and a covariate given all other covariates. In ordinary least squares linear regression with y as the response, the estimated partial coefficient of determination between y and x_K is the difference in the coefficient of determination for a regression model with covariates [x_1,...,x_{K-1},x_K] and a regression model with covariates [x_1,...,x_{K-1}]. For generalized linear models (GLM), no definition of partial coefficients of determination can be found in the literature. Zheng and Agresti (2000) propose a coefficient of determination for GLM, which is the squared correlation between the observed and fitted outcomes. Analogous to linear regression, we propose a partial coefficient of determination between y and x_K, which is the difference in the coefficients of determination for GLM with and without x_K. The bootstrap will be used to obtain confidence intervals for this new partial coefficient. To illustrate the method, we use a study evaluating racial differences in control of risk factors in diabetes.

A Goodness-of-fit Test for Logistic Regression Models with Continuous Predictors

◆ Xianjin Xie, University of Iowa; Jane Pendergast, University of Iowa; William R. Clarke, University of Iowa

University of Iowa, 1100 Arthur St., G6, Dept. of Biostatistics, Iowa City, IA 52240

xianjin-xie@uiowa.edu

Key Words: *goodness of fit, logistic regression*

When continuous predictors are present, classical Pearson and deviance goodness-of-fit tests to assess logistic model fit break down. The Hosmer-Lemeshow goodness-of-fit statistic is often used in these situations. Their procedure groups observations into G bins according to the percentiles of the estimated probabilities. It uses a Pearson chi-square statistic with G-2 degrees of freedom to compare the observed frequency of events to that expected using the model's average predicted value in each group. While simple to perform with satisfactory properties, it provides no further information on the source of any detectable lack of fit. Tsiatis (1980) proposed an alternative statistic which partitions the covariate space and uses a score statistic to test for regional effects. We propose a new method for goodness-of-fit testing which uses a very general partitioning strategy in the covariate space and is based on either a Pearson statistic or a score statistic. Properties of the proposed statistics are discussed and simulation studies comparing it to the existing tests are presented, demonstrating its usefulness in practice.

Comparing Discrimination Rules for Two Populations

◆ Esteban Walker, Cleveland Clinic Foundation; Yaomin Xu, Cleveland Clinic Foundation

Cleveland Clinic Foundation, 9500 Euclid Ave., Dept. of Vascular Surgery / S40, Cleveland, OH 44195

walkere1@ccf.org

Key Words: *classification, logistic regression, regression trees*

A diagnostic tool is a rule designed to identify the presence of a disease. These tools are often developed by comparing samples of healthy and diseased individuals. When the comparison is based on continuous variables, then the objective is to find thresholds that best distinguish between diseased and nondiseased individuals. These thresholds are crucial because they determine the performance of the diagnostic tool (e.g., sensitivity and specificity). We examine common methods to calculate such thresholds. The methods are compared based on properties like bias, variation, and robustness. Recommendations for the use of these procedures are provided.

An Index of Discrimination for Logistic Regression Models with Ordinal Responses

◆ Michael P. McDermott, University of Rochester; Christopher A. Beck, University of Rochester Medical Center

University of Rochester, Dept. of Biostatistics and Comp Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

mikem@bst.rochester.edu

Key Words: *model discrimination, proportional odds model, ROC surface*

A commonly used index of discrimination for a logistic regression model with a binary outcome is the so-called c-index, which is an estimate of concordance between predicted and observed responses. This index is equivalent to the area under the receiver operating characteristic (ROC) curve generated by varying the cutpoint for the predicted probability that is used to classify observations. An index of discrimination is proposed for a logistic regression model with an k-level ordinal outcome. The index is based on the idea of generalizing the concept of an ROC curve to that of an ROC surface when observations can be classified into three or more categories. The surface is generated by varying k-1 cutpoints for the linear predictor that is used to classify observations, with a larger value of the linear predictor being associated with an increased probability of being classified in a higher category. The volume under the ROC surface serves as an index of discrimination. Issues concerning estimation and interpretation of the volume under the ROC surface are discussed.

Modeling Ordinal Categorical Data

◆ John L. Fresen, Medical University of Southern Africa

Medical University of Southern Africa, Dept. of Mathematics and Statistics, PO Box 107, MEDUNSA, 0204, Pretoria, 0204 Republic of South Africa

jfresen@medunsa.ac.za

Key Words: *modeling, ordinal, categorical, data*

Two approaches are commonly used for analyzing ordinal categorical data. The first method is to use a loglinear model for the expected cell frequencies. If the underlying data, before categorization, follow a normal distribution then the linear predictor that is commonly used seems inappropriate. This situation is analyzed by considering the case where the variables follow a joint normal distribution and the case where one of the variables is a response variable and the rest are predictors. The second method, for the case of an ordinal categorical response variable, is to model the cumulative logits of the cell probabilities in terms of a linear predictor. Again, the linear predictor that is commonly used seems inappropriate if the underlying data follow the normal distribution. Different predictors for both these methods are discussed. The estimation is achieved by maximizing an approximate likelihood. Similar arguments hold for other distributions in which the variables have a correlation or regression structure. Examples are presented.

Augmented Marginal Log-linear Models for Capture-recapture Studies

◆ Elizabeth L. Turner, McGill University; Alain C. Vandal, McGill University/SMBD Jewish General Hospital

McGill University, Dept. of Mathematics and Statistics, 805 Sherbrooke St. West, Montreal, PQ H3A 2K6 Canada

eturner@math.mcgill.ca

Key Words: *capture-recapture, source dependence, conditional independence, augmented marginal log-linear models, epidemiology*

In capture-recapture modeling it is necessary to account for possible dependence between sources. We propose a novel technique, termed augmented marginal log-linear modeling (AMLLM), to account for both heterogeneity-induced and pure source dependence in capture-recapture studies when individual covariate data are collected. Central to this technique, under the assumption that sources are conditionally independent given the covariates, is a measure we call the coefficient of source dependence (CSD), defined for every set of sources. The CSDs are formed by two components: the distribution of the covariates in each source and the population distribution of the covariates. The first is estimated empirically and the second estimated along with the unknown population size by inclusion of the CSDs in the AMLLM. This technique permits, within a marginal log-linear model, the inclusion of covariate data and the estimation of the population covariate distribution and population size. It also avoids the problems of model selection at source level and random zeros. We illustrate the use of CSDs in the design of epidemiological capture-recapture studies.

51 Distribution Theory and Stochastic Processes

IMS

Sunday, August 8, 4:00 pm–5:50 pm

Searching for Searchers

◆ Colin L. Mallows, Avaya Labs Research; Jean Meloche, Avaya Labs Research

Avaya Labs Research, 233 Mt. Airy Rd., Basking Ridge, NJ 07920

colinm@research.avayalabs.com

Key Words: *search, functional equations*

We describe a search problem that has arisen in the context of network monitoring. A known, very large, region may contain one or more “agents.” Starting with just one agent, we search until another agent is found; this new agent can assist in the remaining search, and so on recursively.

Mutual Information for the Multinomial Distribution

◆ Walfredo R. Javier, Southern University, Baton Rouge; Arjun K. Gupta, Bowling Green State University

Southern University, Baton Rouge, Mathematics Dept., T.T. Allain, Baton Rouge, LA 70813-9757

waljavi9@aol.com

Key Words: *asymptotic distribution, Kullback-Leibler number, mutual information, multinomial distribution, normal quadratic form*

Mutual information, T(X), among the components of a q-variate random vector X is defined to be the expected value of the natural logarithm of the ratio of the joint density function and the product of the marginal densities of the component variables. This paper presents T(X) for the multinomial distribution. The sampling distribution of the maximum likelihood estimator of T(X) is derived as the distribution of a quadratic form from the approximating (q-1)-variate normal distribution; its characteristic function, mean and approximate variance are then computed. It is proposed to use this sampling distribution in inference problems involving the parameters of the multinomial distribution.

Statistical Tests for Detecting Dependence Relations among Random Variables Generating Random Sets

◆ Franz Streit, Université de Genéve

Université de Genéve, C.P. 240, Section de Mathématiques, CH-1211 Genéve, Switzerland

streit@math.unige.ch

Key Words: *parametric models for random sets, tests for model selection, score tests, choice of copula*

Stochastic models for observed realizations of random sets of the same type are investigated in order to judge whether dependence relations between measures taken from the same realization of the random set need to be incorporated into the stochastic model; score tests turn out to be very useful in this context. Mathematical models for random sets generated by individual radial deformation of a single deterministic prototype-set are considered. Stochastic dependence among the multiplicative deformation factors is obtained in representing the probability law of these factors by a general multivariate distribution. For a given fixed set of marginal distributions the dependence structure of the multivariate distribution is represented by its copula; different degrees of the possible strength of the interdependence among the deformation factors may be realized by a different choice of the copula.

Conditional Saddlepoint Approximations for Noncontinuous and Nonlattice Distributions

◆ John E. Kolassa, Rutgers University; John Robinson, University of Sydney

Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08855

kolassa@stat.rutgers.edu

Key Words: *conditional inference, saddlepoint approximation*

This manuscript presents an approximation to the distribution function of a smooth transformation of a random vector, conditional on the event that values of other smooth transformations of the same random vector lie in a small rectangle. This approximation is used to justify the application of standard saddlepoint conditional tail area approximations in circumstances more general than continuous and lattice cases currently justified in the literature. Applications to conditional inference are discussed.

Recurrence of Markov Chains with an Application to Admissibility

◆ Galin Jones, University of Minnesota

University of Minnesota, School of Statistics 313 Ford Hall, 224 Church St. SE, Minneapolis, MN 55455

galin@stat.umn.edu

Key Words: *admissibility, Markov chain, recurrence, irreducibility, transience*

Consider a discrete time, homogeneous Markov chain that lives on the non-negative real line. This talk will focus on conditions that guarantee the stability of such chains. On discrete state spaces the concepts of irreducibility and recurrence are well understood from a classical perspective. However, on more general spaces these concepts must be modified. I will introduce and connect some notions of recurrence using irreducibility. I will then present new criteria that can be used to establish recurrence. Finally, I will use this recurrence criterion to establish the admissibility of an estimator of the normal mean when the prior is Lebesgue measure on the non-negative real line.

Statistics in Music Analysis

◆ Steven J. Sepanski, Saginaw Valley State University

Saginaw Valley State University, 7400 Bay Rd., Dept. of Mathematics, University Center, MI 48710

sepanski@svsu.edu

Key Words: *music analysis, Markov chains, correlation*

We will look at statistical tools to analyze pieces of music, particularly improvised jazz instrumental solos. We will consider questions such as do certain players have certain phrases that define their style? Are there certain phrases that are almost always played over certain chord changes? We will look at dependency properties, such as Markovian, and if Markovian, of what order?

Nonlinear Tikhonov Regularization for Statistical Inverse Problems

◆ Nicolai Bissantz, University of Goettingen

University of Goettingen, Institute for Mathematical Stochastics, Maschmuehlenweg 8-10, Goettingen, 37073 Germany

bissantz@math.uni-goettingen.de

Key Words: *statistical inverse problems, nonlinear Tikhonov regularization, nonparametric regression, local polynomial estimators, cross-validation*

We consider nonlinear statistical inverse problems described by operator equations F(a)=u. Here a is an element of a Hilbert space which we want to estimate, and u is an L^2-function. The given data consist of measurements of u at n points, perturbed by random noise. We construct an estimator \hat{a}_n for a by a combination of a local polynomial estimator and a nonlinear Tikhonov regularization and establish consistency in the sense that the mean integrated square error (MISE) tends to 0 as n\to\infty under reasonable assumptions. Moreover, if a satisfies a source condition, we show for \hat a_n a convergence rate result for the MISE, as well as almost surely. Further, it is shown that a cross-validated parameter selection yields a fully data-driven consistent method for the reconstruction. Finally, the feasibility of our algorithm is investigated in a numerical study for a groundwater filtration problem and an inverse obstacle scattering problem.

52 Calibration ▲

Section on Survey Research Methods

Sunday, August 8, 4:00 pm–5:50 pm

Minimum MSE Regression Estimation—May be Upper Bound but not Lower Bound of Variance

◆ Mingue Park, University of Nebraska

University of Nebraska, Dept. of Statistics, 103 Miller Hall, Lincoln, NE 68588

mpark2@unl.edu

Key Words: *regression model, regression estimator, MSE*

Estimation for the population mean under the general regression model and general unequal sampling design is considered. Regression weights that are non-negative and generate a linear estimator with minimum model MSE is constructed. The variance estimator for the regression estimator based on the model is also suggested.

Calibration Methods for Foreign and Domestic Investment in Canada

◆ Leon Jang, Statistics Canada; Marie-Claude Duval, Statistics Canada

Statistics Canada, R.H. Coats Building, 11th Floor, Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

leon.jang@statcan.ca

Key Words: *calibration, country of control*

The Capital Expenditures Survey performed by Statistics Canada provides information on capital spending for the economy in Canada at large as well as by industrial classification and province. Estimates on capital spending for several countries of control to evaluate the capital in Canada owned by foreign countries has been requested by data users. Estimates by country of control were produced using domain estimation techniques. As country of control was not considered in the sample design, calibration methods were investigated to improve the estimation. This talk will provide some background on the survey and discuss the improvements made by calibrating to total capital spending and total revenues. Comparisons between calibrated and noncalibrated estimates at various industry and provincial levels will be presented. In addition, future enhancements to the survey to account for the country of control will be presented.

Golden and Silver Jubilee Year-2003 of the Linear Regression Estimators

◆ Sarjinder Singh, St. Cloud State University

St. Cloud State University, Department of Statistics, St. Cloud, MN 56301-4498

sarjinder@yahoo.com

Key Words: *GREG, calibration, estimation of total and variance, unequal probability sampling, startified sampling, two-phase sampling*

The linear regression estimator of population mean owed to Hansen, Hurwitz, and Madow (1953), and that of variance by Das and Tripathi (1978) are shown to be unique. We modify the methodology of Deville and Särndal (1992), and others. For example, note that Hidiroglou and Särndal (1995, 1998), Tracy and Singh (2000), Arnab and Singh (2003), Wu and Luan (2003), Singh (A.C.) and Wu (2003), Ash (2003), Patel and Chaudhari (2003), and Estevao and Sarndal (2003) papers can be improved using Singh (2003), "Advanced Sampling Theory with Applications," and by accounting for a very important point. Some changes in the well-known statistical packages such as GES, SUDDAN, CALMAR, etc., are recommended.

Survey Calibration to CPS Household Statistics

◆ Varma Nadimpalli, Westat; David R. Judkins, Westat; Adam Chu, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

nadimpv@westat.com

Key Words: *raking, post-stratification*

It is common practice on many household surveys to post-stratify the sampling weights to demographic statistics from the Census Bureau. On some surveys, it is also useful to post-stratify to household statistics from the Current Population Survey, such as the number of households containing children in a specific age range. These are less stable than demographic statistics on age, sex, race, and ethnicity. In addition, if the statistics are not published by the Bureau, then estimation can be delayed by problems in the release of pubic-use data files by the Bureau. One focus of this paper is on the smoothing of CPS household statistics for use in post-stratification and on forward extrapolation, as needed. The other focus is on how to reflect the variability of the CPS statistics in the corresponding estimate of sampling errors.

Calibrated Imputation for the Drug Abuse Warning Network

◆ Andrea R. Piesse, Westat; James L. Green, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

andreapiesse@westat.com

This paper presents an imputation approach used for the Drug Abuse Warning Network (DAWN). DAWN uses a stratified, single-stage sample of hospitals to produce estimates for targeted DAWN areas and the nation. In some DAWN areas, data from one or more units are critical to the validity of the estimates for that area. These reporters are generally critical because they contain unusually large volumes of DAWN events or are otherwise atypical of other units in their area. Occasionally, data from a critical reporter becomes temporarily unavailable (i.e., periodic unit nonresponse). We have developed an improved imputation approach, outside of the standard survey imputation procedure, to deal with missing data for critical reporters. The approach uses time series models to obtain totals for important statistics in the missing data period, then uses linear programming to calibrate real reported data from an adjacent period to these modeled totals. The end result is a complete dataset.

Calibration Estimation for Quantiles

◆ Torsten N.J. Harms, Université de Montréal; Pierre Duchesne, Université de Montréal

Université de Montréal, 625 Rue Outremont, Apt. 1A, Montreal, H2V 3M8 Canada

torsten@dms.umontreal.ca

Key Words: *survey, sampling, calibration, GREG, quantile*

In survey sampling, calibration—or GREG—estimators are of widespread use. However, their application is usually limited to the estimation of means or totals. A generalization of the calibration approach will be presented, where the measure of interest as well as for the known covariates of the population can be any arbitrary measure. In most cases there exists no analytic solution of the resulting calibration problem. However, in the case of quantiles, it is possible to find an analytic expression of the calibrated weights. These will be presented along with the resulting variance estimator and confidence intervals for this new quantile-estimator. A Monte Carlo study will evaluate this new estimator under different sampling designs and for different populations. Emphasis will be laid on the performance relative to other known estimators for quantiles that make use of auxiliary information such as difference-, ratio-, or model-based approaches.

Calibration-weighting and Its Effect on Measures of Variation

◆ Richard A. Moore, Jr., U.S. Census Bureau

U.S. Census Bureau, Washington, DC 20233

Richard.A.Moore.Jr@census.gov

Key Words: *calibration-weighting, random group replication, general variance function*

We are taking a random sample of n units from a universe with N units. Suppose a frame variable of interest (e.g., receipts) has a highly skewed distribution. Each selected unit is assigned a weight of N/n. Since the distribution is highly skewed, it is very likely that the estimate for the aggregate value of the weighted sum of this variable over the selected cases differs significantly from the actual aggregate sum over all units in the universe. Suppose we adjust the weights of the selected units, so that (1) the adjusted weighted sum of the frame variable over the selected units equals the sum of the variable for all cases in the universe, and (2) the sum of the adjusted weights equals the number of cases in the universe. This will then require some modifications to the method used for calculating the variance. Using the employer universe of the 2002 Survey of Business Owners, this paper compares several alternative variance estimation methods and focuses on the accuracy of the variances under a calibration-weighting scheme.

53 Mixture Models and Clustering ⊞

General Methodology

Sunday, August 8, 4:00 pm–5:50 pm

A Unified Approach to Studying Two-level Structural Equation Models and Linear Mixed Effects Models

◆ Jiajuan Liang, University of New Haven; Peter M. Bentler, University of California, Los Angeles

University of New Haven, 300 Orange Ave., School of Business, West Haven, CT 06516

jliang@newhaven.edu

Key Words: *covariance structure, EM algorithm, linear mixed effects model, mean structure, two-level structural equation model*

Two-level structural equation models and linear mixed effects models are two different types of statistical models. Existing publications usually study these two types of models separately. We will develop a unified approach to studying the two types of models simultaneously. The idea is based on the similar structure of the two types of models. A two-level structural equation model consists of level 1 effect and level 2 effect. A linear mixed effects model consists of a mixed effect and a random effect. So there are some common points in the structure of these two types of statistical models. In our unified approach, we will consider the mixed effect as the mean structure and the random effect as the covariance structure. Based on this restructuring of linear mixed effects models, we can obtain the same formulation of these two different types of models. An EM algorithm will be developed to fit the same formulation for both types of models. Numerical examples will be given to illustrate the effectiveness of our EM approach applied to linear mixed effects models.

Identifiability of Finite Mixtures

◆ Hajo Holzmann, University of Goettingen

University of Goettingen, Institut für Mathematische Stochastik, Maschmühlenweg 8-10, Göttingen, 37073 Germany

holzmann@math.uni-goettingen.de

Key Words: *finite mixture, identifiability, location-scale family, circular distributions, characteristic function*

Identifiability is a task of general interest in the theory of mixture models, and also has applications in related fields such as hidden Markov models. We present a general result about identifiability of finite mixtures of a family of distributions based on certain assumptions on the tail behavior of the corresponding characteristic functions. This is applied to location-scale families on the real line and to circular distributions. Particular cases include circular wrapped distributions of location-scale families, stable distributions, and the d-dimensional-wrapped normal distribution.

A Mixture Model Using Poisson Kernel Components

◆ Ke Yang, Pennsylvania State University; Bruce G. Lindsay, Pennsylvania State University

Pennsylvania State University, 422A Thomas Bldg., Dept. of Statistics, University Park, PA 16802

kyang@stat.psu.edu

Key Words: *Poisson kernel, EM algorithm, mixture model, nonparametric MLE, directional data*

The Poisson kernel originated from a famous Dirichlet problem. As a probability density function, it can be used to describe the distribution of directional data. The simplest Poisson kernel distribution is the univariate Poisson kernel distribution on a circle, which is well known as wrapped Cauchy distribution. Maximum likelihood estimation for the wrapped Cauchy distribution was studied by Kent and Tyler (1988). We extend their result to higher dimensions and construct an EM algorithm for the MLE. Then we consider the mixture model with the Poisson kernel used as the component density. We discuss the identifiability of the mixture model and its nonparametric MLE. This mixture model can be applied to directional data or normalized data, such as normalized microarray data.

Applying MASAL to Estimate the Distribution of Response Lag in Transcriptional Regulatory Networks

◆ Fenghai Duan, Yale University

Yale University, 470 Prospect St., Apt. 45, New Haven, CT 06511

f.duan@yale.edu

Key Words: *response lag, transcription factors, regulatory networks, MASAL*

Learning the distribution of response lag between transcription factors (TFs) and the genes they regulate (“target genes”) in cells could assist us not only to better understand the underlying cellular mechanism but also to build up more meaningful and more powerful genetic causal networks. Based on Spellman et al.’s yeast cell cycle expression data (1998) and Guzilm et al.’s transcriptional regulatory network data (2002), we could study the distribution of response lag between TFs and their target genes by calculating the mutual information (MI) at different time intervals followed by MASAL fitting (multivariate adaptive splines for longitudinal data). There are some interesting results, including the optimum response lag, the difference between positive-effect and negative-effect transcriptional regulatory networks with respect to response lag, and so on.

Simple Heterogeneity Variance Estimation for Meta-analysis

◆ Jeffrey N. Jonkman, Mississippi State University; Kurex Sidik, Wyeth Research

Mississippi State University, Dept. of Mathematics and Statistics, PO Box MA, Mississippi State, MS 39762

jonkman@math.msstate.edu

Key Words: *across-study variance, weighted estimation, confidence intervals*

We propose a simple method of estimating the heterogeneity variance in a random effects model for meta-analysis. The estimator is simple and easy to calculate, and has reduced bias compared to the most commonly used estimator in random effects meta-analysis, particularly when the heterogeneity variance is moderate to large. Unlike some existing estimators, it always yields a non-negative estimate of the heterogeneity variance. We find that random effects inference about an overall effect based on this estimator is more reliable than inference using the common estimator, in terms of coverage probability for an interval estimate.

Testing Homogeneity in Discrete Mixtures via L-Two Distances

◆ Richard J. Charnigo, University of Kentucky; Jiayang Sun, Case Western Reserve University

University of Kentucky, Dept. of Statistics and Division of Biostatistics, 851 Patterson Office Tower, Lexington, KY 40506-0027

richc@ms.uky.edu

Key Words: *D-test, discrete mixture, homogeneity, L-two distance, mixture distribution*

The recently developed D-test for homogeneity in continuous mixture distributions is competitive with likelihood ratio tests and easier to employ. We develop a D-test for homogeneity in discrete mixtures. We define a D-test statistic, study its convergence rates under the null and alternative hypotheses, and characterize its asymptotic null distribution under maximum likelihood and penalized maximum likelihood estimation frameworks. As in the continuous case, a generalized or weighted D-test may be performed; while the interpretation is different in the discrete setting, the generalization still allows the D-test to be adapted to the specific parametric family from which the mixture components come. We characterize the asymptotic null distribution of the generalized D-test statistic in the discrete setting and establish that similar results hold in the continuous case. The small-sample applicability of the asymptotic theory as well as the competitiveness of the D-test and its generalization are investigated for Binomial and Poisson mixtures. We also examine an epidemiological dataset that is not adequately described by a homogeneous Poisson model.

Effect of Using Principal Coordinates and Principal Components on Retrieval of Clusters

◆ William Warde, Oklahoma State University

Oklahoma State University, Statistics Dept., 301 MSCS, Stillwater, OK 74078

billw@okstate.edu

Key Words: *agglomerative clustering, principal coordinates, principal components*

Principal coordinate analysis is a more powerful technique than principal components analysis to ensure identification of groups of objects if some conditions are satisfied. The results of using principal coordinates prior to cluster analysis were investigated. Three different methods of standardization were examined and compared with no standardization using both principal coordinates and principal components. The retrieval ability of the known agglomerative clustering algorithms was improved by using principal coordinates. The results of applying principal coordinates based on the correlation coefficient instead of squared Euclidean distance prior to application of the clustering algorithms were less sensitive to changes in noise.

54 Robustness ⌘

General Methodology

Sunday, August 8, 4:00 pm–5:50 pm

Robust Regression with Neural Networks Using Iterative Learning Trimmed Elemental Estimators

◆ Matthew Hall, University of Kansas Medical Center; Matthew Mayo, University of Kansas Medical Center

University of Kansas Medical Center, 3901 Rainbow Blvd. MS1008, Kansas City, KS 66160

mhall4@kumc.edu

Key Words: *neural networks, robust regression, elementals*

It is well known that regression with neural networks (NNs) can be sensitive to the presence of outliers in the data. Since is it common to use the sum of squared errors as a metric when seeking to optimize the weights of the network, neural network linear regression shares many of the robustness problems that ordinary least squares (OLS) does. We will address the robustness problems for neural networks through the use of elemental subsets. Furthermore, we will reframe the work of Mayo and Gray (1997, 2001) in a neural network environment and compare the performance of an iterative learning version of Mayo and Gray's trimmed elemental estimators (TEEs) with that of OLS. We found that by using neural networks with iterative learning TEEs, the robustness of the parameter estimates is greatly improved over traditional methods.

A New Robust Regression Estimator with Self-tuning Weights

YouSung Park, Korea University; ◆ DongHee Lee, Korea University

Korea University, Dept. of Statistics, Anam-Dong, Sungbuk-Gu, Seoul, 136-701 Korea

ld0351@korea.ac.kr

Key Words: *WSTE, self-tuning, efficiency, influence function, high breakdown point, bad leverage outliers*

We introduce a new robust regression estimator called as weighted self-tuning robust regression estimator (WSTE). WSTE has data-adapted tuning constants which adjust by themselves and resists outliers by downweighting outlying observations systematically. We derive influence function of WSTE and show its asymptotic normality. Using five empirical datasets frequently used to assess robustness, we show that only WSTE perfectly detects known outliers in these data, compared to five competitive robust estimators. High-breakdown estimators such as the least median of squares, the least trimmed squares and the S estimators are impractical to compute exactly in large samples. However, since computation of the WSTE depends only on the number of independent variables but not on the sample size, WSTE is exactly computed and unique. Extensive simulation studies are performed using datasets with no outliers to evaluate efficiency, with large percentage of outliers to measure degree of breakdown point, and with bad leverage outliers to test bounded influence. We observe from these simulation studies that WSTE is superior to six competitive regression estimators.

Robust Tests for the Equality of Variances for Clustered Data

◆ Ivan Iachine, University of Southern Denmark; Hans C. Petersen, University of Southern Denmark

University of Southern Denmark, Campusvej 55, 5230 Odense M, Dept. of Statistics, Odense, DK-5230 Denmark

iachine@stat.sdu.dk

Key Words: *equality of variances, non-normality, robust tests, Levene's test, clustered data, twins*

Tests for the equality of variances are often needed in applications. In genetic studies the assumption of equal variances of continuous traits, measured in identical and fraternal twins, is crucial for heritability analysis. To test the equality of variances of traits, which are non-normally distributed, Levene (1960) suggested a method which was surprisingly robust under non-normality, and the procedure was further improved by Brown and Forsythe (1974). These tests assumed independence of observations. However, twin data are clustered—observations within a twin pair may be dependent due to shared genes and environmental factors. Uncritical application of the tests of Brown and Forsythe to clustered data may result in much higher than nominal Type I error probabilities. To deal with clustering we developed an extended version of Levene's test, where we replaced the ANOVA step with a regression analysis with standard errors computed using a clustered version of the Huber-White sandwich estimator, followed by a pseudo-t-test. We studied the properties of our procedure using simulated non-normal clustered data and obtained Type I error rates close to nominal and reasonable power.

Robust Instrumental Variable Estimator for Measurement Error Models

◆ Gabriela Cohen-Freue, University of Maryland, College Park; Ruben Zamar, University of British Columbia

University of Maryland, College Park

gcf@math.umd.edu

Key Words: *instrumental variables estimator, robust diagnostic, measurement error models, S-estimators, robust estimator*

This paper examines robust estimation in linear models with measurement errors. We derive robust instrumental variables (RIV) estimators using high breakdown point S-estimators of multivariate location and scatter matrix. The resulting estimator has bounded influence function and is in the class of weighted instrumental variables estimators. Moreover, this estimator is computationally inexpensive and a natural extension of the ordinary instrumental variables estimator. The asymptotic variances of the parameter estimates are also obtained and estimated. In addition, we construct a diagnostic tool to flag outliers and leverage points for the errors in variables models when instrumental variables are available. An illustration of the performance of our estimator is given by a numerical example.

Comparisons of Bias Curves of Some Robust Estimators of Scale

◆ Daniela Szatmari-Voicu, University of Calgary; John Collins, University of Calgary

University of Calgary, Calgary, AB Canada

danasv@shaw.ca

Key Words: *asymmetric contamination, robust estimation, scale parameter, minimum-GES (gross error sensitivity), minimax asymptotic bias comparisons*

We study the comparative asymptotic performance of three classes of scale estimators under [ε]-contamination of the distribution Fo from which the data arises. The classes studied are: M-estimators, L-estimators and the [γ]%-Shorth (the length of the shortest (1-γ)% of the data). Asymptotic bias curves are compared for different estimators which have been standardized to have the same gross error sensitivity (GES). The main results are: (1) Minimum-GES M-estimators are found when Fo is not necessarily strongly unimodal. These optimal estimators have piecewise constant influence functions which do not always correspond to the median absolute deviation (MAD). (2) When Fo is strongly unimodal, symmetrized interquantile ranges have better performance than their nonsymmetrized counterparts, in the sense that their bias curves are uniformly lower. However, for symmetric but nonunimodal Fo, symmetrization can sometimes result in uniformly higher bias curves. (3) The bias curve for the 50%-Shorth is uniformly lower than that of the MAD when Fo is symmetric and unimodal.

Robust Detection of Multiple Outliers in Grouped Multivariate Data

◆ Nedret Billor, Auburn University; Chrys Caroni, National Technical University of Athens

Auburn University, 235 Allison Lab, Auburn, AL 36849-5307

billone@auburn.edu

Key Words: *multivariate data, outliers, robust methods, cluster analysis*

Many methods have been developed for detecting multiple outliers in a single multivariate sample, but very little on the case where there may be groups on the data. We propose a method of simultaneously determining groups (as in cluster analysis) and detecting outliers, which are points that are distant from every group. Our method is an adaptation of the BACON algorithm proposed by Billor, Hadi, and Velleman for the robust detection of multiple outliers in a single group of multivariate data. There are two versions of our method, depending on whether or not the groups can be assumed to have equal covariance matrices. The effectiveness of the method is shown by a simulation study for different sample sizes and dimensions for two and three groups, with and without planted outliers in the data. When the number of groups is not known in advance, the algorithm could be used as a robust method of cluster analysis, by running it for various numbers of groups and choosing the best solution.

Robust Fitting of Multinomial Models Based on Assessment of Model

◆ Jiawei Liu, Pennsylvania State University; Bruce G. Lindsay, Pennsylvania State University

Pennsylvania State University, 325 Thomas Bldg., Dept. of Statistics, University Park, PA 16802

jiawei@stat.psu.edu

Key Words: *multinomial, goodness of fit, statistical tube, model selection, bootstrap, semiparametric*

We introduce a semiparametric tubular extension model in the multinomial setting. It consists of all multinomial distributions lying in a distance-based neighborhood of the parametric model of interest. Fitting such a tubular model allows one to use a parametric model while treating it as an approximation to the true distribution. In this paper, the Kullback-Leibler distance is used. Based on this idea one can define an index of fit of the model, to be the distance between the true multinomial distribution and the parametric model. The paper develops a likelihood ratio test procedure for testing the magnitude of the index. A semiparametric bootstrap method is implemented to better approximate the distribution of the LRT statistic. The approximation permits more accurate construction of a lower confidence limit for the model-fitting index.

55 Methods for Case-control Studies and Logistic Regression

Section on Statistics in Epidemiology, Biometrics Section

Sunday, August 8, 4:00 pm–5:50 pm

Impact of the Coexistence of a Composite Score and Its Components in a Multiple Logistic Regression Model on Predicting Clinical Outcomes

◆ Yen-Hong Kuo, Jersey Shore University Medical Center

Jersey Shore University Medical Center, 1945 State Rte. 33, Neptune, NJ 07754

yhkuo@jhu.edu

Key Words: *logistic regression, multivariate analysis, composite score, coexistence, prediction*

To summarize the overall clinical status of a patient, many scoring systems have been widely used. These composite scores consider and weight several patient characteristics and clinical variables to form a single value. In clinical studies, a composite score is usually evaluated with other covariates in a multiple logistic regression model to predict the clinical outcomes. Some patient characteristics, such as age and sex, are usually included in the multiple logistic regression models. They are also common components in a composite score. As a consequence, when a multivariate analysis is performed, the coexistence of a composite score and its component has been observed in the medical literature. This presentation will use age as the variable of interest. Cases identified from the medical journals will be used to illustrate the types of coexistence. The impact will be discussed from the statistical and clinical points of view. Solutions will be proposed to prevent this problem.

Estimating Relative Risk Parameters in Presence of Missing Covariates

◆ Jinbo Chen, National Cancer Institute; Nilanjan Chatterjee, National Cancer Institute; Mitchell H. Gail, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd., EPS Room 8089, Rockville, MD 20851

chenjin@mail.nih.gov

Key Words: *missing covariate, logistic regression, pseudo-likelihood, piecewise exponential distribution, EM algorithm, weighted average*

It was of interest to estimate the relative risks of breast cancer for various risk factors, including reproductive factors (R) and mammographic density (MD), using data from BCDDP study. BCDDP followed a cohort of 280,000 women for their breast cancer status for up to five years (screening phase, 1975-1980); then a subsample was selected for long term follow-up (follow-up phase, 1980-1995). For the screening phase, R were available for a nested case-control sample, a portion of which also had MD data. For the follow-up phase, all subjects had R, some of which were updated over time, but only a small subset had the MD data. The missingness of MD information imposed a challenge for our analysis, and it was desired to make efficient use of all available data. We assumed logistic regression model for the screening phase and adopted a pseudo-likelihood approach for estimation. A piecewise exponential model was assumed for the follow-up phase and the maximum likelihood estimates were obtained via the EM algorithm. The two sets of results were then combined by taking their weighted average.

The Effect of Differential Misclassification of Exposure Status in Case-control Study

◆ Tzesan D. Lee, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, NCEH/EHHE, Mail Stop E70, Atlanta, GA 30333

tjl3@cdc.gov

Key Words: *sensitivity/specificity of classification procedure, corrected odds ratio, misclassification errors, resins/lung cancer mortality data*

Substantial misclassification occurs in medical research. Several investigators have examined the effect of misclassification errors on the relative risk in 2 x 2 tables. In the case of nondifferential misclassification, a formula has been developed to adjust for the effect of misclassification errors on the odds ratio. However, two aspects of the current studies on the problem of misclassification still remain inadequately covered. First, the variance of the corrected odds ratio was not given. Second, the assumption of nondifferential misclassification is invalid in most instances in medical screening. This presentation examines the effect of differentially misclassifying the exposure status on the estimation of the true unknown odds ratio in case-control studies. Assuming the classification method classifies the exposure status of both the case and control subjects erroneously and differentially, we present the formula for both the corrected odds ratio and its standard error. The corrected odds ratio depends intrinsically, in a nonlinear way, on the four unknown parameters, namely, the sensitivities and specificities of the classification method.

Analyses of Case-control Studies Using Bayesian Approach

◆ Kuang Fu Cheng, National Central University

National Central University, Graduate Institue of Statistics, Chungli, Taiwan

kfcheng@cc.ncu.edu.tw

Key Words: *Bayesian model, case-control study, logistic regression, profile likelihood, risk parameter*

We developed a model and a simple computation scheme for a Bayesian approach to inference in case-control studies. The model proposed in this paper is based on a profile likelihood of exposure given disease. This profile likelihood is derived from a logistic regression model and the same as the binomial likelihood but with an offset. Further, its intercept parameter is a function of risk parameters.The properties of the suggested priors are discussed and the method can be applied to both discrete and continous covariates.The approach is illustrated with two examples, the first involving a dataset concerning the association between PL-A2 polymorphism and cronary heart disease, and the second based on a dataset comprising two case-control studies of wire codes, magnetic fields, and children's leukemia.

Describing Stratified Multiple Responses for Sparse Data

◆ Ivy Liu, Victoria University of Wellington

Victoria University of Wellington, Wellington, New Zealand

iliu@mcs.vuw.ac.nz

Key Words: *Mantel-Haenszel method, multiple responses, sparse data, stratification*

Surveys often contain qualitative variables for which respondents may select any number of the outcome categories. For instance, for the question "What type of contraception have you used?" with possible responses (oral, condom, lubricated condom, spermicide, and diaphragm), respondents would be instructed to select as many of the outcomes that apply. This type of response is called "multiple responses." Bilder and Loughin (2002) proposed a Cochran-Mantel-Haenszel (MH) type method to test whether the choice of type of contraception is marginally independent of an explanatory variable given a stratification variable (known as conditional multiple marginal independence (CMMI). We apply the generalized MH type estimators to estimate the conditional group effects among the outcome categories and follow the bootstrap method to estimate the variances and covariances for the estimators. It performs well even for highly sparse data.

56 Methods and Applications in Epidemiology ▲

Section on Statistics in Epidemiology

Sunday, August 8, 4:00 pm–5:50 pm

Comparing Restricted Cubic Splines to Fractional Polynomials for Obtaining Reference Percentiles of Clinical Measurements

◆ Stephen D. Weigand, Mayo Clinic; Steven D. Edland, Mayo Foundation & Mayo Clinic; Clifford R. Jack, Jr., Mayo Clinic; Maria M. Shiung, Mayo Clinic

Mayo Clinic, Division of Biostatistics, Harwick 7, Rochester, MN 55905-0001

weigand@mayo.edu

Key Words: *reference percentiles, restricted cubic splines, fractional polynomials, normalizing transformations*

For quantitative clinical measurements, it is often informative to know the patient's estimated percentile relative to sample data from a "normal" or reference population. A useful method to obtain the reference percentiles, which may depend on several covariates, involves first transforming the reference sample measurements to a standard normal using a Box-Cox or other transformation. Using the quantiles of the standard normal distribution and a back-transformation, reference percentiles can be obtained. Transformation parameters can be made to vary as a function of covariates using, for example, restricted cubic splines or fractional polynomials. We compare the performance of these two modeling approaches using human brain measurements obtained from magnetic resonance imaging data. Transformations examined include the Box-Cox and exponential normal.

Using Tensor Product Splines in Modeling Exposure-time-response Relationships: Application to the Colorado Plateau Uranium Miners Cohort

◆ Kiros Berhane, University of Southern California; Michael Hauptmann, National Cancer Institute; Bryan Langholz, University of Southern California

University of Southern California, Dept. of Preventive Medicine, 1540 Alcazar St., CHP-220, Los Angeles, CA 90089-9011

kiros@usc.edu

Key Words: *splines, latency, nested case control, dose-response, excess relative risk*

We propose tensor product spline-based flexible modeling techniques for describing exposure-response relationships for protracted time-dependent exposure histories in epidemiologic studies of occupational exposures. The methods use multidimensional nonparametric estimation techniques based on tensor product splines to jointly model age, time-since-exposure (latency) and exposure-response effects. The main advantage of this proposed methodology is its ability to allow for latency functions that vary by exposure levels and, conversely, exposure-response relationships that are influenced by the latency structure. We present a detailed analysis of data from the Colorado Plateau Uranium Miners cohort. We compare our results with previous analyses of the same dataset via other less flexible techniques that focus on flexible modeling of only the latency function or the exposure response relationship, by keeping the other fixed at a pre-determined functional form.

Bayesian Analysis of Negative Binomial Models Applied to Lyme Disease in New York State

◆ Haiyan Chen, SUNY, Albany and NYSDOH; Howard H. Stratton, SUNY, Albany

SUNY, Albany and NYSDOH, SPH, R632, ESP, Corning Tower, Biometry & Statistics Dept., Albany, NY 12237

haiyan_c@yahoo.com

Key Words: *Bayesian, incidence rate, Lyme disease, Markov chain Monte Carlo, Poisson, negative binomial*

Lyme disease (LD) occurrence in New York State (NYS) appears as an emerging public health problem; however, few studies have investigated its epidemic and spatial dynamics in great detail. Using data from NYS Department of Health Lyme Registry Surveillance System, the authors studied epidemic and spatial characteristics of LD in NYS for the 11-year time period from 1990 through 2000. Because of both overdispersion and excess zeros presented in the NYS LD data related to a Poisson model, a Bayesian negative binomial models (NB) was constructed in order to give a foundation for future state and local health funding plans and education programs. Joint posteriors and predictive densities of future observations are obtained using Markov chain Monte Carlo methods, and performance of the method is evaluated.

Combining Variables by the Use of Principal Components Can Lead to an Index with Low Reliability—Two Examples

◆ John S. Grove, John A. Burns School of Medicine; Clementina Ceria, University of Hawaii, Manoa; J. David Curb, Pacific Health Research Institute

John A. Burns School of Medicine, 1960 East West Rd., University of Hawaii, Honolulu, HI 96822

jgrove@hawaii.edu

Key Words: *principal components, reliability coefficient, intraclass correlation*

Investigators often face having either several imprecise measures of the same underlying variable or a set of variables which are thought to be related. Using them simultaneously as independent variables in a regression model creates estimates which are difficult to interpret and have large standard errors due to intercorrelations. One common reaction is to create an index of the variables from the first eigenvector of principal component (PC) analysis. But since variables are weighted according to their variances, variables with high random error variance may contribute the most to the first component. Two examples are given of this. The first PC of systolic blood pressure (SBP) measured at two exams with different reliabilities resulted in a lower Cox regression coefficient for heart disease incidence on SBP than using the better of the two measurements alone. Using data from the Pacific Genetics Network study on balance measured by an electronic balance platform, the first principal component of eight variables had a reliability coefficient (RC) of only 0.4, whereas two of the input balance variables had RC's > 0.6. The best linear function for balance had a RC of 0.8.

Modeling Variations over Time Using Multilevel Procedures: An Illustration to Model Counts and Number of Protected Coital Acts

◆ Emelita L. Wong, Family Health International

Family Health International, PO Box 13950, Research Triangle, NC 27709

ewong@fhi.org

Key Words: *multilevel modeling, protected coital acts*

To describe patterns of responses over time, regression lines where the explanatory variables consist of time intervals may be fitted. Within a multilevel model framework, such regression line has two components, namely: (1) fixed-effects components, which present mean outcome at time point coded as zero and an overall trend over time, and (2) random effects components, the first set of which measures how the overall mean varies from subject-to-subject and from time point to time point, while the second set measures how the trend (slope) of responses over time varies from subject-to subject and joint variation of the overall mean and overall trends. The data for the illustrative analysis consist of self-reports of number and proportions of protected coital acts in 5,241 occasions, reported monthly for six months during trial participation, and at approximately 14 months after the trial among 964 Cameroonian women. While the fixed effect estimate of trends indicates overall decrease in proportion of protected coital acts over time, the random effects estimates differentiate three types of trends over time: decreasing, no change, and increasing.

An Application of Recurrent Events Analysis

◆ Alexander Cambon, University of Louisville

University of Louisville, Louisville, KY 40241

accamb01@louisville.edu

Key Words: *recurrent events*

In the fields of product engineering reliability and survival analysis, emphasis is often on time-to-event phenomena, particularly time to first event. This is of key importance when evaluating interventions to increase survival time with respect to specific diseases, for example. In product engineering reliability, time to first failure analysis is an important part of improving reliability of consumer appliances. More recently, in survival analysis, increased attention is being given to recurrent or multiple events analysis. The study of multiple or recurrent hospitalization events, for example, can give additional valuable information about an intervention. This paper focuses on an analysis of recurrent hospital events with respect to an intervention for congestive heart failure.

57 New Developments in Time Series Methodology

Business and Economics Statistics Section

Sunday, August 8, 4:00 pm–5:50 pm

Simple and General Tests for White Noise

◆ Ignacio N. Lobato, Instituto Tecnologico Autonomo de Mexico; Carlos Velasco, Universidad Carlos III de Madrid

Instituto Tecnologico Autonomo de Mexico, Av Camino Sta Teresa 930, Mexico City, 10700 Mexico

ilobato@itam.mx

Key Words: *Gaussianity, nonparametric, autocorrelation, periodogram, nonlinear dependence*

This article considers testing that a time series is uncorrelated when it possibly exhibits some form of nonlinear dependence. Contrary to the currently employed tests that require selecting arbitrary user-chosen numbers to compute the associated tests statistics, we consider tests statistics that are very simple since their asymptotic null distributions are standard under general weak dependent conditions, and hence, asymptotic critical values are readily available. We develop asymptotic theory and study finite sample performance.

Stability of Cyclic Threshold Autoregressive Time Series Models

◆ Thomas R. Boucher, Virginia Polytechnic Institute and State University; Daren B.H. Cline, Texas A&M University

Virginia Polytechnic Institute and State University, Dept. of Statistics, Blacksburg, VA 24061-0439

tboucher@vt.edu

Key Words: *ergodicity, Markov chain, nonlinear time series, threshold autoregressive time series*

We investigate the stability of the threshold autoregressive time series model by embedding the time series in a general state Markov chain and deriving conditions under which the Markov chain, and thus the time series embedded in it, is either V-uniformly ergodic or is transient. We concentrate in particular on what we term cyclic threshold autoregressive time series models. These models exhibit asymptotic cyclic behavior, that is, the process cycles through one of a number of sets of subregions of the state space when the process is large. In this case, conditions for V-uniform ergodicity or transience of the process can be determined through analysis of the deterministic skeleton of the process. We apply stochastic drift conditions to demonstrate this. Our methods can be applied to cases where the model has multiple cycles and/or affine thresholds, thereby extending beyond current results.

Recursive Estimation of Misspecified MA(1) Models: General Results

◆ James Cantor, Science Applications International Corporation; David F. Findley, U.S. Census Bureau

Science Applications International Corporation, 1100 North Glebe Rd., Arlington, VA 22201

jcantor@sito.saic.com

Key Words: *time series models, incorrect models, pseudo-linear regression, recursvie maximum likelihood, misconvergence*

We introduce a general algorithm for the recursive estimation of an MA(1) model. The algorithm includes as special cases both pseudo-linear regression (PLR, AML, RML1) and also recursive maximum likelihood estimation (RML2). We consider its application to data generated by an MA(1) model and by several other models. Stimulated by Hannan (1980), and generalizing results of Cantor (2001), we analyze the convergence of the sequence of recursive estimates by showing, under a stability condition, its asymptotic equivalence to a sequence that satisfies a Robbins-Monro recursion. Convergence of the latter recursion is established using results of Fradkov (1980) and Findley (2001). Under moderate restrictions on the coefficients of the generating model, the stability condition is verified for PLR and for a monitored version of RML2. The latter is proved to converge to the mean square optimal parameter in all cases, whereas PLR is proved to converge to a nonoptimal value for non-MA(1) data.

On Multiple Hypotheses Testing and Model Selection

◆ Kasing Man, Syracuse University; Chung Chen, Syracuse University

Syracuse University, Whitman School of Management, Syracuse, NY 13244

ksman@syr.edu

Key Words: *time series, model selection, vector ARMA model*

It is of interest to study the presence or the absence of specific dynamic relations between time series. Chen and Lee (1990) proposed a multiple-hypotheses-testing procedure in a vector ARMA framework to identify the possible dynamic relation (which include independent, contemporaneous relation, unidirectional relation, and feedback relation) between time series. This paper extends this procedure to account for strong form relationship. It is a sequential inference procedure based on the likelihood ratio tests on models with various parametric constraints. On the other hand, viewing models with different parametric constraints as different models, they can be compared using model selection criteria such as AIC and BIC. It will be interesting to see how the two approaches compare and relate to each other. Simulation studies indicate that the performance of AIC is similar to the procedure conducting at a higher significance level; while BIC at a lower significance level. For empirical analysis, we study some economic time series to illustrate the procedures. Robustness of the results over possible alternative model specification will be addressed.

Optimal Filters for Contemporaneously Aggregated Series

◆ Georgios Tripodis, London School of Economics

London School of Economics, Houghton St., London, WC2A 2AE UK

G.Tripodis@lse.ac.uk

Key Words: *Kalman filter, seasonal adjustment, Wiener-Kolmogorov filter*

Time series are often part of a total (e.g., Euro area GDP is derived by aggregating GDP for each member state). Seasonally adjusted series for the total can be derived either directly seasonally adjusting the total or by aggregating the seasonally adjusted estimated for each component of the total. This paper investigates the differences between directly and indirectly seasonally adjusted series. We investigate optimal filters for both the components and the total, which satisfies the additivity constraint for the seasonally adjusted series. We concentrate on final estimation error and on total revisions of seasonally adjusted estimates. We also compare the results with existing seasonally adjusted methods.

Representation of Multiplicative Vector Autoregressive Moving Average Processes

◆ Ceylan Yozgatligil, Temple University; William W.S. Wei, Temple University

Temple University, Speakman Hall 006-00, 1810 North 13th. St., Philadelphia, PA 19122-6083

ceylany@temple.edu

Key Words: *multiplicative vector autoregressive moving average process, VARMA process, estimation, forecasting, causality*

Time series often contains observations of several variables, and multivariate vector time-series processes are used to study the relationship between these variables. There are many studies on vector autoregressive moving average (VARMA) processes, but studies on representation of multiplicative VARMA processes are nonexistent. In multiplicative vector processes, for example, in the seasonal VARMA process, there can be different representations for the process because of the noncommutative property of matrix multiplication. We carefully examine the consequences of different presentations on parameter estimation, forecasting, and causality. From these examinations, we will introduce a summary statistic that is useful in determining the best representation for a multiplicative vector process.

58 Imputation and Missing Values ▲

Section on Survey Research Methods

Sunday, August 8, 4:00 pm–5:50 pm

Comparing Impact of Alternative Approaches for Item Imputation in the Job Openings and Labor Turnover Survey

◆ Mark Crankshaw, Bureau of Labor Statistics

Bureau of Labor Statistics, Postal Square Bldg., 2 Massachusetts Ave., NE, Washington, DC 20212

crankshaw_m@bls.gov

Key Words: *sample survey, labor statistics, nearest neighbor imputation, ratio data*

The Bureau of Labor Statistics produces monthly estimates for job openings, hires, and separations from The Job Openings and Labor Turnover Survey (JOLTS). The JOLTS survey has implemented a hot-deck nearest neighbor imputation algorithm to account for item nonresponse. For this survey, the values taken from imputation donors take the form of a ratio. The distribution of these borrowed ratios varies greatly depending on the size of the donor establishment. We wish to examine how well the profile of our imputation donors matches the profiles of reporting establishments of comparable size. If the profiles do not match we would like to measure the extent of statistical bias. We wish also to examine several variations of our current imputation approach to identify the variation that most reduces bias.

Embedding Logic Checks in an Automated Hot-deck Imputation of Survey Data

◆ Rong Huang, University of California, Los Angeles; Wei Yen, University of California, Los Angeles; Jenny Chia, University of California, Los Angeles

University of California, Los Angeles, Center for Health Policy Research, 10911 Weyburn Ave., Suite 300, Los Angeles, CA 90024

ronghuan@ucla.edu

Key Words: *skip check, hot-deck, imputation*

This paper introduces a method we developed that incorporates skip pattern checks of survey in an imputation process. Frequently, surveys contain missing data that require imputation. In complex surveys in which a variable may be related to many others due to question skipping, imputation of a variable may not possibly include all related variables as controls. The result is that the intervariable relationship is either evaluated post imputation, most often manually, for the imputed values or is simply ignored. When an imputed value is found to be inconsistent with other variables, it is sometimes manually edited to make it consistent with other variables or is set to missing to be imputed again. The system we developed automates the whole process of imputation and logic checks. The resulting imputed values are then consistent with all known skip patterns. A hierarchical sequential hot-decking method is used for the imputation and the program is coded in SAS.

Using Name Information to Impute Respondent Demographics

◆ Darryl V. Creel, Mathematica Policy Research, Inc.; Paul Guerino, Mathematica Policy Research, Inc.; Donsig Jang, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., 600 Maryland Ave. SW, Suite 550, Washington, DC 20024

dcreel@mathematica-mpr.com

Key Words: *imputation, NSRCG, frame information*

In surveys that depend on multiple outside sources to supply frame information, there is often complete and partial nonresponse on stratification variables. In order to draw the sample, it then becomes necessary to impute missing frame information. The purpose of this paper is to test the race/ethnicity and gender imputation schemes utilized in the 2003 National Survey of Recent College Graduates. Race/ethnicity is imputed using a multitier procedure that looks at the racial makeup of institution attended, last name, and first name, whereas gender is imputed based on then gender makeup of the institution attended, first name, and middle name. To test the effectiveness of our imputation scheme, we applied the imputation scheme to known frame cases.

Nonresponse Adjustment in IRS Taxpayer Compliance Studies

◆ Karen C. Masken, IRS

IRS, 1111 Constitution Ave. NW, Washington, DC 20024

karen.c.masken@irs.gov

Key Words: *imputation, compliance rate, tax*

While the general public would not view an audit by the IRS to be a survey, adjusting for “no-shows” in IRS compliance studies presents the same issues as any other survey nonresponse. Traditionally, IRS has treated these no-shows as missing at random and used mean imputation, which is believed to understate the noncompliance rate. Operationally, the examiner determines the adjustment to be made and the taxpayer is assessed accordingly, but using this amount for imputation is believed to overstate noncompliance. While these two methods provide a lower and upper bound on the noncomplinace rate, it is desirable to have a method that estimates the actual rate. This paper will compare alternative methods of imputation and their impact on the overall compliance rate as well as on line-item estimates. It will also make recommendations for imputation in future studies.

A Comparison of the Characteristic Imputation in Census 2000 to the Accuracy and Coverage Evaluation Survey for Matched Persons

◆ Robert D. Sands, U.S. Census Bureau

U.S. Census Bureau, Room 2503-2, Decennial Statistical Studies Div., Washington, DC 20233

robert.d.sands@census.gov

Key Words: *imputation, missing data, coverage survey, census*

In Census 2000, an item imputation operation ensured that each person and housing unit on the census roster had a value for the relationship to householder, sex, age, Hispanic origin, race, and tenure 100% enumeration data items from the Census Short Form. The current study is an attempt to describe how “close” the 2000 Census Characteristic Imputation for these core demographic variables is to the “truth” as represented by data for these same items collected from the same persons in the Accuracy and Coverage Evaluation (ACE) Survey. Closeness of the comparison between the Census and the ACE characteristics will be measured by a percentage of agreement and disagreement.

A Simple Evaluation of the Imputation Procedures Used in NSDUH

Eric Grau, RTI International; ◆ Pete Frechtel, RTI International; Dawn Odom, RTI International; Dicy Painter, Substance Abuse and Mental Health Services Administration

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

frechtel@rti.org

Key Words: *predictive mean matching, nearest neighbor imputation, predictive mean, weighted hot-deck, imputation evaluation*

The National Survey on Drug Use and Health (NSDUH) is the primary source of information on drug use in the U.S. Since 1999, the Predictive Mean Neighborhoods (PMN) procedure has been used to impute missing values for many of the analytical variables. This method is a combination of two commonly used imputation methods: a nearest-neighbor hot-deck and a modification of Rubin’s predictive mean matching method. Although PMN has many practical advantages, it has not been formally evaluated. We propose a simple simulation to evaluate PMN. Using only complete data cases, we will induce random patterns of missingness in the data for selected outcome variables. Imputations will then be conducted using PMN and a weighted nearest-neighbor hot-deck. This process of inducing missingness and imputing missing values will be repeated multiple times. The imputed values using PMN and the weighted hot deck will then be compared with the true values that were found in the complete data, across the repeated iterations. In particular, we will compare the number of matches between the two methods, as well as comparing statistics derived from the data, such as drug prevalence estimates.

59 Invited Poster Session

Section on Government Statistics, Section on Physical and Engineering Sciences, Biometrics Section, Section on Bayesian Statistical Science, IMS, Section on Statistical Computing, WNAR, Section on Statistics and the Environment, Section on Survey Research Methods, Social Statistics Section

Sunday, August 8, 6:00 pm–7:50 pm

The Foundation of Human Rights Statistics in Sierra Leone

◆ Richard Conibere, Benetech

Benetech, Springwell Barn, Coberley, Cheltenham, GL53 9QY UK

rconibere@martus.org

Analyzer is a database application designed specifically for storing human rights violations data. It has been used extensively in Sierra Leone within the Truth and Reconciliation Commission in order to generate statistics from over 7,000 statements giving details of victims and the violations they suffered. Databases have been employed in the field of human rights for many years, but models used in the past have been inadequate for measuring “who did what to whom.” Analyzer, in contrast, has been designed to represent a more accurate model of human rights situations. Where multiple sources describe the same persons and events, the system facilitates record linkage. Stemming from this is direct

support for multiple systems estimation (MSM) and the potential to study recall error. Furthermore, inclusion of measures for inter-rater reliability between matchers can improve matching quality. This poster will describe the underlying structure of Analyzer and some of its more pertinent features, using data from Sierra Leone to demonstrate.

Can Sheer Will Control a Robot Arm?

◆ Valerie Ventura, Carnegie Mellon University

Carnegie Mellon University, 5000 Forbes Ave., Dept. of Statsistics, Pittsburgh, PA 15213

vventura@stat.cmu.edu

Key Words: *particle/Kalman filter, tuning curves, neurons, encoding/decoding*

Neurons in motor cortex areas encode information about movement variables in their firing rate, so that, reversing the problem, it is possible to predict movements given the observed firing rates of a population of neurons. Now that it is possible to record the activity of many neurons simulateously, this paradigm can be used to build prosthetic devices. We propose a statistical model that allows a monkey to use a prosthetic arm for reaching tasks. Issues involve determining what movement or feedback variables cortical neurons encode, building an efficient model that is robust to unpredictable variations in the brain, and producing smooth dynamic real-time arm movements.

Bayesian Methods for Analyzing Speech Recognition Scores

◆ Matthew J. Hayat, National Institutes of Health; Prakash Laud, Medical College of Wisconsin

National Institutes of Health, NIDCD, 6120 Executive Blvd., MSC 7180, Bethesda, MD 20892

hayatm@nidcd.nih.gov

Key Words: *Bayesian, longitudinal, serial correlation, heterogeneity of variance, modeling dependence, Cholesky*

Heterogeneity of variance and serial correlation are often present in measurements taken over time. The most popular parametric dependence models for serial correlation are stationary autoregressive models and other second-order stationary models. In these models, it is assumed that variances are constant over time and correlations between measurements equidistant in time are equal. These assumptions may not be reasonable. Our work considers a class of nonstationary models that allows for heterogeneity of variance and serial correlation. Modeling dependence is difficult for two reasons. First, dimensionality of the problem can be large in many applications and second, the covariance matrix must be constrained to be positive definite. A modified Cholesky decomposition of the precision matrix (inverse of the covariance matrix) allows us to address both of these challenges. It also produces nonstationary analogues of many stationary covariances with special structure that are available in the literature of longitudinal data analysis. We implement full Bayesian inference for several such models. Markov chain Monte Carlo techniques are used.

Pattern Filtering for Acoustic Recognition Using Point Processes

◆ Zhiyi Chi, University of Chicago

University of Chicago, 5734 University Ave., Dept. of Statistics, Chicago, IL 60637

chi@galton.uchicago.edu

Key Words: *pattern recognition, point processes, neuroscience, bioacoustics, simulation*

Many scientific studies require fast and accurate pattern recognition. In the case of acoustic detection, this is challenging because of the variations in acoustic patterns. To address this, an important idea is to construct representations of the inputs that are robust to variations. We find that the structures of bioacoustic signals can be represented by points on the frequency-time domain. The resulting, discrete, representations not only exhibit robustness to random variations, but also significantly reduce the dimension of the data. We then can formulate acoustic detection as detection of global patterns of points. Under a Poisson point process model, the latter can be achieved by linear filtering of point processes. The training of the detector only requires a small sample. It constructs filters based on the structures learned from the sample, and tunes the other parameters by simulation. We have implemented the approach both offline and online. It has enabled neuroscientists to conduct experiments on neurobehavioral interactions with a degree of precision that could not be achieved before.

Graphical Display of Uncertainty and Related Tools for Inference

◆ John L. Eltinge, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 1950/PSB, Washington, DC 20212

Eltinge_J@bls.gov

Key Words: *confidence sets, Current Employment Statistics Program, false discovery rate, small-area estimation, simultaneous inference, triple-goal estimation*

Small-domain estimation methods use regression, hierarchical modeling, and related tools to combine sample survey data and auxiliary information to produce estimators for a relatively large number of subpopulations. Some practical applications of these methods are complicated by two factors. First, anecdotal evidence indicates that in some cases stakeholders attempt to use published small-domain estimates to carry out exploratory analyses and related (informal) simultaneous inference. Second, many stakeholders have relatively limited previous training in statistics. To address these issues, this poster presents some relatively simple graphical methods for display of small-domain estimates and related measures of uncertainty. For cases involving univariate estimates, the triple-goal approach of Shen and Louis (1998) provides a useful framework for development of three types of graphical displays. For cases involving multivariate comparisons, projections of confidence sets and related simultaneous inference methods lead to some alternative graphical approaches.

Modeling Survey Nonresponse in Dichotomous Processes

◆ Jacob J. Oleson, Arizona State University; Chong He, University of Missouri, Columbia

Arizona State University, Dept. of Mathematics & Statistics, Tempe, AZ 85287-1804

Jacob.Oleson@asu.edu

Key Words: *incomplete data, nonignorable, spatial correlation, hierarchical Bayes*

Sampling units that do not answer a survey may dramatically affect the estimation results of interest. The response may even be conditional on the outcome of interest. If estimates are found using only those who responded, the estimate may be biased, known as nonresponse bias. Our objective is to find estimates of success rates from a survey when there may be nonresponse bias. Often, these success rates may be spatially correlated. The response rates may also be spatially correlated. This is particularly true if response is conditional on the outcome. In a Bayesian hierarchical framework, we examine two approaches for treating nonresponse that account for potential spatial correlations. Spatial dependence is induced by a common latent spatial structure. This methodology is appropriate for many surveys including the American Community Survey, the National Health Interview Study, and the National Health and Nutrition Examination Survey. An example will be presented.

Model-based Clustering Methods in Population-genetic Inference

◆ Eric C. Anderson, Southwest Fisheries Science Center

Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Rd., Santa Cruz, CA 95060

eric.anderson@noaa.gov

Key Words: *finite mixtures, MCMC, multilocus data, conservation genetics*

This poster reviews the development of model-based clustering methods in which genetic data are used to cluster individuals (or "portions" of individuals) into groups that represent membership in separately-mating populations. Examples of such inference include: (1) genetic stock identification in fisheries, or, as it has recently been called, "population assignment"; (2) inference of population structure by the identification of separate, panmictic subpopulations; (3) the identification of hybrid individuals or individuals with ancestry from two or more subpopulations; and (4) the inference of recent migration rates between subpopulations. While these problems cover a range of scenarios, the statistical models upon which they are based are very similar. We focus on the progression of such models (presented in terms of their underlying acyclic directed graphs) from the early 1980s to the present. This emphasizes the similarity of current methods, but also details the differences, and suggests areas for future research. In connection with this, I will demonstrate software for visualizing the progress of Markov chain Monte Carlo simulations.

Boosting Extensions to Find Anomaly Structure in Data

◆ Virginia L. Wheway, Boeing Company

Boeing Company, 2100 Lake Washington Blvd., #H408, Renton, WA 98056

virginia.l.wheway@boeing.com

Key Words: *data-mining, noisy, boosting, machine learning, classification, cluster*

Recent advances in data mining have led to the development of a method called "boosting." Instead of building a single model, boosting sees several models being built using weighted versions of the original data. These models are then combined into a single prediction model via voting. Studies have demonstrated that boosting leads to significantly lower prediction error on unseen data. This poster demonstrates how the method of boosting may be extended beyond its original aims of improved prediction. Simple plots of specific boosting statistics may be used as tools to detect noisy data and unearth structure within datasets. Whether or not this "suspect" data occurs in groups or as single observations may also be determined. An industrial dataset will be used to demonstrate the process, and show the power of detecting unknown clusters within datasets. Proposed extensions for this research include testing for the threshold of cluster size able to be detected. Research into the method's extension to continuous data would certainly have many opportunities for existing datasets in many sectors, particularly for data containing rare events and data usually considered to be noisy.

Highly Structured Models and High-energy Astrophysics

◆ David A. van Dyk, University of California, Irvine

University of California, Irvine, CA 92697-1250

dvd@uci.edu

Key Words: *astronomy, MCMC, Bayesian methods, missing data, image analysis, deconvolution*

In recent years, there has been a giant leap in the quality and quantity of new data in observational high-energy astrophysics. Recently launched or soon-to-be launched space-based telescopes that are designed to detect and map ultra-violet, X-ray, and gamma-ray electromagnetic emission are opening a whole new window to study the cosmos. Because production of high-energy electromagnetic emission requires temperatures of millions of degrees and indicates the release of vast quantities of stored energy, these instruments give a new perspective on the hot and turbulent regions of the universe. The new instrumentation allows for very high-resolution imaging, spectral analysis, and time-series analysis. The Chandra X-ray Observatory, for example, produces images at least thirty times sharper than any previous X-ray telescope. The complexity of the instruments, of the astronomical sources, and of the scientific questions leads to a subtle inference problem that require sophisticated statistical tools. This poster describes the statistical methods developed by the California-Harvard Astrostatistics Collaboration to address outstanding inferential problems in high-energy astrophysics.

Spatial Statistics for Modeling Phytoplankton

◆ Leah J. Welty, Johns Hopkins University; Michael L. Stein, University of Chicago

Johns Hopkins University, Bloomberg School of Public Health, Dept. of Biostatistics, 615 N. Wolfe St., Baltimore, MD 20001

lwelty@jhsph.edu

Key Words: *variogram, chlorophyll, spatial, anisotropy, covariance*

Phytoplankton are important elements of lake and ocean ecosystems; they are the base of aquatic and oceanic food chains and may affect global climate through fixation of atmospheric carbon by photosynthesis. Statistical challenges associated with modeling phytoplankton biomass include correction for measurement bias, calibration of indirect measurements, covariance model specification for anisotropic processes, and computational limitations associated with evaluating likelihoods for large datasets. We present potential solutions to these problems that though developed with an eye to understanding phytoplankton dynamics are general enough to be relevant to many spatial, spatial-temporal, and environmental problems.

Stochastic Particle Systems' Approach to Modeling of Viscous Flows

◆ Anna Amirdjanova, University of Michigan

University of Michigan, 439 West Hall, 550 E. University, Dept. of Statistics, Ann Arbor, MI 48109-1092

anutka@umich.edu

Key Words: *hydrodynamics, particle systems, vorticity, stochastic partial differential equation*

For years the Navier-Stokes system has attracted special attention from researchers because of elegant and difficult problems posed by it and its fundamental importance in applications. It is well known that for large Reynolds numbers fluid flow becomes turbulent and requires a stochastic description of the model. Original idea in that direction was to study statistical solutions of the classical Navier-Stokes system. Recently, however, there's been a growing interest in the study of individual solutions to stochastic Navier-Stokes equations. There are several ways to introduce stochasticity into the system. One popular approach is to add external random forces to the classical PDE and treat it as an evolution equation in certain function spaces. Another one comes from vortex methods. The latter originate from a system of randomly moving point vortices (particles, representing centers of rotation in the fluid) and are more tractable from numerical perspective. We present our recent results on the properties of solution to stochastic vorticity equation in R^2 and comment on existing particle systems' interpretations of equations of hydrodynamics.

Statistical Challenges in the Analysis of Mass Extinctions

◆ Steve C. Wang, Swarthmore College

Swarthmore College, Dept. of Mathematics and Statistics, 500 College Ave., Swarthmore, PA 19081

scwang@swarthmore.edu

Key Words: *paleontology, Poisson process, likelihood ratio test, Bayesian model, uniform distribution, fossil record*

Much of our knowledge of the history of life comes from the fossil record. However, the fossil record is notoriously incomplete; in fact, usually more data are missing than are observed. This incompleteness presents interesting challenges for paleontologists and statisticians. We describe approaches for modeling the incompleteness of the fossil record in the context of mass extinctions. These extinctions—such as the end-Cretaceous event in which the dinosaurs perished—have profoundly shaped the course of life on earth. To infer the causes of mass extinctions, it is important to estimate the times of extinction of the species involved. For instance, how can we determine if a set of species went extinct simultaneously or gradually? If they went extinct simultaneously, how can we estimate their common time of extinction? If they went extinct gradually, how long did the extinctions last? We will discuss methods for answering such questions that take into account the incompleteness of the fossil record.

60 Introductory Overview Lecture on Bayesian Methods in Business

ASA, Section on Statistics and Marketing, Business and Economics Statistics Section, ENAR, WNAR, SSC, IMS

Monday, August 9, 8:30 am–10:20 am

Bayesian Methods in Marketing

◆ Robert E. McCulloch, University of Chicago

University of Chicago, Graduate School of Business, 1101 East 58th St., Chicago, IL 60637

rem@gsb.uchicago.edu

Bayesian methods have become widespread in the marketing literature. We review the essence of the Bayesian approach and explain why it is particularly useful for marketing problems. While the appeal of the Bayesian approach has long been noted by researchers, recent developments in computational methods and expanded availability of detailed marketplace data has fueled the growth in application of Bayesian methods in marketing. We emphasize the modularity and flexibility of modern Bayesian approaches. The usefulness of Bayesian methods in situations in which there is limited information about a large number of units or where the information comes from different sources is noted.

Bayesian Methods in Finance

◆ Nicholas Polson, University of Chicago

University of Chicago, 1101 East 58th St., Chicago, IL 60637-1561

ngp@gsb.uchicago.edu

This talk develops Bayesian methods for finance. A number of common finance models such as Black-Scholes and the Merton model are described together with applications to option pricing and portfolio choice. Particular emphasis is given to statistical inference problems where the researcher has to infer latent state variables and parameters from price data. This will include an introduction to Bayesian methods for inference in continuous-time asset pricing models and sequential inference methods for performing online estimation. Markov chain Monte Carlo methods will also provide a tool for exploring these financial econometric models and distributions. Examples include equity price models, option pricing models, term structure models, and regime-switching models.

61 Toward the Gold Standard for Social Measurement in the 21st Century ▲ ⊞

Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section, Section on Health Policy Statistics

Monday, August 9, 8:30 am–10:20 am

From Survey Data to Multiple Types of Data in Historical and Real Time

◆ Juanita T. Lott, U.S. Census Bureau; Fritz J. Scheuren, NORC, University of Chicago; Jay Keller, U.S. Census Bureau; David Banks, Duke University

U.S. Census Bureau, HRD, 3-3024, 4700 Silver Hill Rd., Washington, DC 20233-0001

Juanita.T.Lott@census.gov

Key Words: *real-time data, gold standard, federal statistical system*

This paper addresses one fundamental topic in addressing real-life, real-time, quickly made decisions—the evolution of data requirements for federal statistical systems in relation to heightened expectations of various data users. Our focus is limited to delineating federal data requirements over time in relation to various types of data that are now routinely expected since survey data were heavily introduced in the 1940s. We provide examples of statistical innovations that respond to heightened expectations and close with some thoughts and characteristics that might be part of the effort needed to define a gold standard for social measurement in the 21st century.

Can You Count Them? What Do the Numbers Say?

◆ Rueben Warren, Centers for Disease Control and Prevention; S. Miles Richardson, Centers for Disease Control and Prevention; Gladys Reynolds, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry, 1600 Clifton Rd., NE, MS E-28, Atlanta, GA 30333

Rcw4@cdc.gov

Key Words: *health disparities, environmental justice, institutional racism*

Racial and ethnic health disparities in the United States are well documented. In 1985 there were 60,000 excess deaths among African Americans. Several investigations document associations between hazardous waste sites and race/income status. The U.S. Government Accounting Office found race/ethnicity and low-income were associated with living near hazardous waste sites. The United Church of Christ found that race was the most consistent factor. African Americans and Hispanics are four to five times more likely to live within a one-mile radius of a hazardous waste site. Research is documenting adverse health affects and exposure to toxic and hazardous chemicals. This presentation describes adverse health conditions among racial/ethnic populations; their disproportionate location near hazardous waste sites, toxic chemicals associated with selected health conditions, and plausible physical and behavioral associations. Further consideration should be given to plausible associations between environmental exposures and racial/ethnic health disparities.

Toward a Paradigm for Integrating Information: Applications and Principles

◆ Dean H. Judson, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Suitland, MD 20746

dean.h.judson@census.gov

Key Words: *calibration, information integration, administrative records, estimation, program evaluation*

We describe an emerging synthesis of ideas surrounding the appropriate ways to combine data of different kinds. First, we build a framework in which we describe the typical pitfalls in using such data for research and program evaluation purposes, and how statistical models can be used to "calibrate and augment" one data source with another. Next, we describe several application areas, and illustrate the potential for pitfalls (and use of calibration) in such initiatives. Finally, we point to a potential new applications in the future. We conclude by proposing new uses of multiple data sources for policy-relevant information.

62 Statistical Methods Using Support Vector Machines with Applications in Microarrays and Other Fields

ENAR, Section on Statistical Computing, WNAR, Section on Statistical Graphics

Monday, August 9, 8:30 am–10:20 am

Semparametric Regression for Microarray Data Using Support Vector Machines

◆ Debashis Ghosh, University of Michigan; Xihong Lin, University of Michigan; Dawei Liu, University of Michigan

University of Michigan, Dept. of Biostatistics, School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109-2029

ghoshd@sph.umich.edu

Key Words: *machine learning, gene expression, semiparametrics*

We consider a semiparametric regression model to relate a continuous outcome to clinical covariates and gene expressions, where the clinical covariate effects are modeled parametrically and gene expression effects are modeled nonparametrically using the support vector machine. The nonparametric function allows for the fact that the number of genes is likely to be large and the genes are likely to interact with each other. Equivalences with the linear mixed model will allow for the use of standard mixed model software. Both the regression coefficients of the clinical covariate effects and the support vector estimator of the nonparametric gene expression function can be obtained using the Best Linear Unbiased Predictor in linear mixed models. The smoothing parameter can be estimated as a variance component in linear mixed models. A score test is developed to test for the significant gene expression effects. The methods are illustrated using a prostate cancer dataset and evaluated using simulations.

Support Vector Machines for Polar Cloud Detection

Tao Shi, University of California Berkeley; ◆ Bin Yu, University of California, Berkeley; Eugene Clothiaux, Pennsylvania State University; Amy Braverman, California Institute of Technology

University of California, Berkeley, Statistics Dept., Berkeley, CA 94720

binyu@stat.berkeley.edu

Key Words: *cloud detection, Support Vector Machine, feature selection, MISR*

Cloud detection is crucial to many problems in climate research, but is notoriously difficult in polar scenes where the background scene has similar spectral signatures to those of clouds. The Multi-angle Imaging SpectroRadiometer (MISR), launched in 1999 as part of NASA's Earth Observing System, provides the next generation of high-resolution datasets for climate studies including those related to clouds and their effects on the radiation budget of Earth. MISRs view the Earth at nine view angles and four spectral wavelengths, providing 36 dimensional data at 1.1 km resolution. This is the first instrument in Earth orbit to provide both multi-angle and multispectral information simultaneously. We apply support vector machines with careful feature selection that incorporates our physical understanding of the radiative properties of ice, snow, and clouds to this problem. We compare various methods to develop a practical strategy for obtaining training data to feed the SVM. We then apply the Gaussian kernel support vector machine to a representative area over Greenland and compare the results to an expert-labeled image.

Estimation and Prediction in High Dimension Low Sample Size Settings with Censored Outcomes

◆ Brent A. Johnson, University of North Carolina; Danyu Lin, University of North Carolina; J. Stephen Marron, University of North Carolina

University of North Carolina, Dept. of Statistics, Box 8203, Raleigh, NC 27695-001

bjohnson@bios.unc.edu

Key Words: *microarray, survival analysis*

Support Vector Machines and Distance Weighted Discrimination are two useful methods for discriminant analyses in High Dimension Low Sample Size settings. Classification becomes more difficult when the outcome is censored, however. We first explore various methods that extend SVM and DWD for predicting survival at a particular point in time. Second, we consider a penalized likelihood approach for inferring relationships between survival time and a large vector of predictors. We illustrate the methods with microarray data from two breast cancer studies.

63 Advances in Meta-analysis ▲

Section on Health Policy Statistics, Section on Statistics in Epidemiology

Monday, August 9, 8:30 am–10:20 am

Applying the Law of Iterative Logarithm to Cumulative Meta-analysis

◆ Joseph C. Cappelleri, Pfizer Inc.; Mingxiu Hu, Pfizer Global Research & Development; K.K. Gordon Lan, Aventis Pharmaceuticals

Pfizer Inc., Eastern Point Rd. (MS8260-2222), Groton, CT 06340

joseph_c_cappelleri@groton.pfizer.com

Key Words: *cumulative meta-analysis, meta-analysis, law of iterated logarithm, multiple inspections, sequential analysis, Type I error*

Cumulative meta-analysis typically involves performing an updated meta-analysis every time a new trial is added to a series of similar trials, which by definition involves multiple inspections. This presentation presents an approach—motivated by the Law of Iterated Logarithm—that "penalizes" the Z-value of the test statistic to account for multiple tests and the unstable estimation of the between-study variances at the beginning of the testing process when the number of studies is small (Statistica Sinica 13:1135-1145, 2003). It can also account for the unpredictable nature of information from trials in a cumulative meta-analysis. Our extensive simulation studies show that this method controls the overall Type I error for a very broad range of practical situations for up to 25 inspections for both continuous outcomes and binary outcomes. Examples will illustrate the methodology.

The State of the Art in Diagnostic Test Meta-analysis

◆ Christopher H. Schmid, Tufts-New England Medical Center

Tufts-New England Medical Center, 750 Washington St, Boston, MA 02111

cschmid@tufts-nemc.org

Key Words: *diagnostic tests, meta-analysis, ROC curves, random effects, multilevel model*

Diagnostic testing occupies an increasing share of the focus and costs of health care. This share will only increase as imaging and genetic-screening applications proliferate. Because each technique is usually evaluated in multiple studies, meta-analysis can help combine the results. Nevertheless, relative to reviews of treatment efficacy, new medical tests are infrequently reviewed. We believe that part of the reason is the lack of appropriate methods for combining diagnostic test data. We have surveyed the literature on meta-analysis of diagnostic tests of medical procedures and technologies, collecting every study published since 1990. We describe the methods used, their appropriateness, and whether use of more sophisticated analytic multilevel models might change the conclusions of the studies. Many of the methods employed are simple summaries of sensitivity and specificity or use the summary receiver operating curve. Most are fixed effects methods that do not properly account for between-study heterogeneity. We focus on aspects of study design that introduce heterogeneity between studies and explore whether certain factors may be able to explain this heterogeneity.

Using Selection Models to Deal with Publication Bias

◆ Norma Terrin, Tufts-New England Medical Center; Christopher H. Schmid, Tufts-New England Medical Center; Michael F. Dowd, Tufts-New England Medical Center

Tufts-New England Medical Center, 750 Washington St. Box 63, Boston, MA 02111

nterrin@tufts-nemc.org

Key Words: *publication bias, selection model, trim and fill, funnel plot*

Publication and related biases occur when statistically significant results are more accessible than nonsignificant results. The funnel plot, a popular tool for detecting publication bias, is a scatterplot of studies in a meta-analysis, with a measure of effect on the horizontal axis, and a measure of precision on the vertical axis. Asymmetry in the funnel plot is interpreted as evidence of publication bias. The "Trim and Fill" method imputes studies to make the funnel plot appear symmetric, then pools actual and imputed studies to obtain an overall effect estimate. However, asymmetry may have causes other than publication bias, including study heterogeneity and chance. Selection modeling is a method of bias adjustment that does not use the funnel plot. Study effects are modeled using random effects, and the selection process is modeled by assigning a weight to the estimated effect from each study. The form of the weight function may be parametric or nonparametric, and the estimation method may be maximum likelihood or Bayesian. We describe and compare the performance of several selection models, applied to a wide range of simulated and actual meta-analyses.

64 Evolutionary Graphics for Streaming Data ▲

Section on Statistical Graphics

Monday, August 9, 8:30 am–10:20 am

Data Cleansing and Preparation at the Gates: A Data Streaming Perspective

◆ Don Faxon, George Mason University; R. Duane King, George Mason University; John T. Rigsby, Naval Surface Warfare Center Dahlgren Division; Steve Bernard, George Mason University

George Mason University, MS 4A7, 4400 University Dr., Fairfax, VA 22030-4444

dfaxon@gmu.edu

Key Words: *data-cleaning, streaming data, massive datasets, data preperation, data collection*

Collection of internet traffic data at the gates of a large enterprise necessarily involves data-cleaning, integration, selection, and transformation, especially if data-streaming strategies are employed. The huge quantities of packets that typically cross the enterprise gateway make multiple passes through the data cost-prohibitive. Data-cleansing, customarily perceived as the removal of noise and inconsistent data, is instead seen as a flagging and tagging procedure to facilitate detection of malformed or corrupted IP packets associated with malicious intrusion, or subtle reconnaissance activity as precursor to a massive attack on the enterprise computing infrastructure. Since real-time or near-real-time implementation of data analysis comprising such innovative concepts as data streaming or evolutionary graphics, fast in-line data cleansing and preparation is required. This paper discusses and illustrates the strategies we have incorporated into our data collection and analysis.

Modeling Internet Traffic Data

◆ Karen Kafadar, University of Colorado, Denver; Edward J. Wegman, George Mason University

University of Colorado, Denver, PO Box 173364, Dept. of Mathematics Box 170, Denver, CO 80217-3364

kk@math.cudenver.edu

Key Words: *evolutionary graphics, streaming data, internet attacks, internet traffic modeling*

Graphical displays are valuable tools in the modeling and analysis of data. Internet traffic data pose special challenges due to the high rates at which they are collected and their nonstationary behavior, even within relatively short periods of time. Different displays are indicated depending upon the ultimate purpose of the analysis; e.g., forecasting trends, detecting outliers, identifying possible network attacks. We describe components of internet traffic, propose some methods of visualizing them, and illustrate these methods on data collected at a university network. Some open problems in studying high-volume data in general are mentioned.

Detecting Computer Masqueraders Using Online Monitoring

◆ David J. Marchette, Naval Surface Warfare Center

Naval Surface Warfare Center, Code B10, 17320 Dahlgren Rd., Dahlgren, VA 22448-5100

marchettedj@nswc.navy.mil

Key Words: *computer security, internet traffic, internet attacks, graphics for streaming data*

An important unsolved problem in computer security is that of detecting when an account has been hijacked and the person using the account is not the authorized user. We describe a dataset consisting of the titles of windows used by users, and discuss methods for visualizing the user's session as it progresses.

Intersection graphs are used to relate the current session to past sessions by the user, and the evolution of the graphs are used as a “profile” for the user.

65 Attack Detection, Interdiction, and Response ▲ ⌘

Section on Statistics in Defense and National Security, Social Statistics Section, Cmte on Statisticians in Defense and National Security, Section on Health Policy Statistics

Monday, August 9, 8:30 am–10:20 am

Statistical Challenges in Monitoring Mental Health Effects of Terrorism

◆ Douglas A. Samuelson, InfoLogix, Inc.

InfoLogix, Inc., 8711 Chippendale Court, Annandale, VA 22003

dsamuel@seas.upenn.edu

Key Words: *mental health, psychoactive agents, threat detection, health assessment, health outcomes*

Terrorist attacks, public alerts (including false alarms), and other traumatic events produce adverse effects on individuals and interpersonal relationships. Data are sparse, especially about relationships, but there are credible indications that these effects are surprisingly large: for example, the divorce rate among first responders to the 1995 Oklahoma City bombing quadrupled in the year after the bombing. Detecting such effects and evaluating efforts to respond to them are difficult because of hesitation to seek treatment, stigma, and other incentives against accurate reporting, inconsistencies in diagnostic protocols, uncertainty in linking symptoms to causative events, and privacy issues. Consequently, even a direct attack utilizing psychoactive chemical agents would be hard to identify. This area offers interesting opportunities and pitfalls for innovative statistical analysis.

Discriminant Analysis Models for Unified Damage Prediction Across Failure Modes

◆ Thomas F. Curry, Northrop Grumman Information Technology; Elisabetta L. Jerome, Sverdrup Technology/TEAS

Northrop Grumman Information Technology, 5450 Tech Center Dr., Suite 201, Colorado Springs, CO 80919

tom.curry@ngc.com

Key Words: *predictive modeling, Operation Iraqi Freedom, weapons effects, discriminant analysis*

The development of precision guided munitions and the need to minimize collateral damage have brought damage modeling to the forefront of modern warfare. For years, weapons effects modelers have struggled with the problem of predicting damage modes and damage parameters within each mode. The usual solution was to try to predict the failure mode, and then select a model appropriate for that mode. This resulted in discontinuities in the damage prediction as failure modes changed. Incorrect damage model selection, combined with uncertainties within failure modes, frequently caused large prediction errors. An approach is presented that uses discriminant functions to produce an optimized (minimum error) damage model that is continuous across failure modes. Mode transition points are explicitly defined by mode-weighting. Initial comparisons to older models show the discriminant analysis model can reduce damage prediction error by 75%. Warfighters stated the models were very useful in Operation Iraqi Freedom.

How Do I Know if I’m Sick?

◆ Robert Armstrong, National Defense University; Stephen Prior, Potomac Institute for Policy Studies

National Defense University, Bldg. 20, Suite 3, Fort Lesley J. McNair, Washington, DC 20319-5066

armstrongre@ndu.edu

The fear of a bioterrorism attack has led policymakers to propose various monitoring schemes. The premise for monitoring is that the earlier we detect the fact that a population has been attacked, the greater chance we have to mitigate the event. No attempt has been made, however, to fully evaluate the value of data coming from various monitoring sources—e.g., biosensors, sentinel populations, absenteeism reporting, etc. In particular, no evaluation has been done to estimate the weight that any particular data stream has, with respect to final decisions made by policymakers; for example, whether to vaccinate a population or not, or whether or not to impose a quarantine. Using a modification of a standard military wargame approach, this study has conducted an analysis of 10 separate data streams and their influence on decisions made by “policymakers” in the wargame. Using an analysis of variance, it has sought to determine the most useful combination of data streams and makes recommendations for public funding of an integrated “system of systems” for biodetection.

66 Modeling and Statistics: Tools for Tackling Public Health Threats

Section on Government Statistics, Section on Health Policy Statistics, Cmte on Statisticians in Defense and National Security, Section on Statistics in Defense and National Security, Section on Risk Analysis, Section on Statistics in Epidemiology

Monday, August 9, 8:30 am–10:20 am

Modeling the Impact of HIV-1 Vaccines in an Era of HAART and Potential Change in Risk Behavior

◆ Wasima Rida, Statistics Collaborative, Inc.; Sonja Sandberg, Framingham State University

Statistics Collaborative, Inc.

wasima@statcollab.com

Key Words: *AIDS, epidemic modeling, HIV vaccines, vaccine efficacy*

Using a system of differential equations, we explore the potential effects of HIV-1 vaccination on the prevalence of HIV and AIDS in a population of men who have sex with men. Biological effects of vaccination include reduced susceptibility to infection, delay in progression to AIDS, and reduced infectiousness of vaccine recipients who become infected. Duration of protection and the potential of HIV exposure or re-vaccination to boost immunologic responses are considered. Behavioral effects of vaccination include potential change in the number or duration of sexual partnerships among vaccinees and non-vaccinees as well as the use of HAART following HIV infection. Special attention is given to vaccines with low or moderate efficacy as these vaccines may represent first generation products considered for licensure.

Modeling the Risks of Emerging Threats to the Blood Supply

◆ Susie ElSaadany, Health Canada

Health Canada, Population and Public Health Branch, Statistics and Risk Assessment Section, A.L. 0601E2 Bldg. No 6 - Tunney's Pasture, Ottawa, ON K1A OL2 Canada

Susie_ElSaadany@hc-sc.gc.ca

Key Words: *mathematical modeling, West Nile Virus, public health, infectious disease, transmissible spongiform encephalopathies*

Mathematical modeling is increasingly being used to identify useful interventions to reduce risks to the blood supply from emerging threats. Among those threats is the West Nile Virus. In 2002 several cases of West Nile Virus may have been transmitted via blood and blood products, categorizing the agent as a threat to the safety of the blood supply. Transmissible Spongiform Encephalopathies (TSEs) such as Creutzfeldt Jakob Disease (CJD) and variant Creutzfeldt Jakob Disease (vCJD) may similarly pose risks of transmission through blood products, although TSE transmission to humans via blood transfusion has not been observed. Progress on modeling the potential risks arising from these new emerging infectious agents and the effectiveness of current interventions, such as the blood donor questionnaire, will be presented. Examples will highlight the applications of modeling to addressing emerging infectious disease agents and the difficulty of modeling when data are limited and the resulting uncertainty high.

The Risks and Benefits of Smallpox and Vaccination: Modeling the Impacts on Public Health

◆ Steven A. Anderson, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1401 Rockville Pike, HFM-210, Rockville, MD 20852

andersonst@cber.fda.gov

Key Words: *public health, infectious diseases, smallpox, vaccine, modeling*

Smallpox was eradicated in the 1970s. However, use of the agent as a bioterrorism weapon, while considered remote, is possible. Fortunately, an effective vaccine exists that is capable of preventing the spread of the disease, should it re-emerge. Large scale or mass vaccination campaigns of the entire U.S. population may pose particular challenges to public health. To explore some of these challenges we use mathematical models to explore various "what if" scenarios to determine the impact of vaccination and other interventions. The models can be further used to identify the most effective measures that optimize benefits while reducing the risks associated with vaccination. The structure of the models, various scenarios and their application in evaluating optimal interventions will be discussed.

67 A Tribute to Milton Sobel: His Life and Works ⌘

IMS

Monday, August 9, 8:30 am–10:20 am

Milton Sobel's Contributions to Selection and Ranking

◆ Thomas J. Santner, The Ohio State University

The Ohio State University, 1042 Putney Dr., Columbus, OH 43085

santner.1@osu.edu

Key Words: *selection, screening, ranking*

This talk will present an overview of Milton Sobel's research contributions to selection and ranking methodology. Selection methods are concerned with designing an experiment to choose the "best" of t treatments. Ranking methods, sometimes called screening methods, determine a subset of the t treatments that contains the "best" treatment with a given confidence level. Milton had a keen interest in selection and ranking methods throughout his career. In part, this curiosity was fueled by his associations with Bob Bechhofer, a fellow graduate student at Columbia, and with Shanti Gupta, a coworker at Bell Laboratories. Bob and Shanti were in the beginning stages of formulating selection and ranking methods, respectively. Milton contributed to both theories and proposed several schemes for unifying the two approaches. Along with Bob and Shanti, Milton became one of the three major influences in the development of these topics. Among the themes that pervade Milton's work, we will discuss the use of curtailment, inverse sampling, and sequential methods as techniques to best use scarce experimental resources.

Sobel's Contributions in Sequential Analysis

◆ Yung L. Tong, Georgia Institute of Technology

Georgia Institute of Technology, School of Mathematics, Atlanta, GA 30332-0160

tong@math.gatech.edu

We review Milton Sobel's contributions in the general area of sequential analysis, including his earlier work with A. Wald, his joint work with R.E. Bechhofer and J. Kiefer on an extension of Wald's SPRT, his research on playing-the-winner methods, and other related topics.

Contributions to Group Testing and to the Dirichlet Distribution

◆ Ingram Olkin, Stanford University

Stanford University, Dept. of Statistics, 390 Serra Mall, Stanford, CA 94305-4065

iolkin@stat.stanford.edu

The group-testing problem is the determination of a testing strategy whereby each of N individuals is to be labeled "has hepatitis" or "does not have hepatitis." The key ingredient is that the relevant blood tests can be carried out for groups of individuals. Consequently, if none of the individuals has hepatitis then there is a savings in not having to carry out multiple tests. We here describe Milton Sobel's contrinutions to the solution of this problem. The Dirichlet distribution can be viewed as a multivariate extension of the beta distribution. Just as the beta distribution represents the tail of a binomial distribution, the Dirichlet distribution represents the tail of a multinomial distribution. However, there now is a multitude of stopping procedures for different sampling protocols. Milton Sobel has been one of the leaders in the development of the theory of Dirichlet distributions, and we here provide a review of his contributions.

68 The First Course in Computational Statistics ▲

The American Statistician, Section on Statistical Education, Section on Statistical Computing

Monday, August 9, 8:30 am–10:20 am

Teaching Computing in a Statistics Graduate Program

◆ James E. Gentle, George Mason University

George Mason University, School of Computational Sciences, Fairfax, VA 22030

jgentle@gmu.edu

Key Words: *teaching, statistical computing, computational statistics*

Almost all statisticians spend a large portion of their working days using the computer in various ways. In addition to the standard things that almost everyone does, statisticians' use of computers includes data analysis with prepackaged software, development of algorithms and software to implement new statistical methods, Monte Carlo simulation to study the performance of statistical procedures, and mathematical analysis using symbolic processing software. The required levels of expertise in computing vary widely among statisticians. At whatever level, for many statisticians, the aspects of their jobs that involve computing are largely self-taught. This is sometimes the most efficient way to learn, but it often leaves serious gaps in one's knowledge. For those of us engaged in the education of statisticians, there are important questions of how and when to provide systematic training in computing, and what to include. The answers to these questions have changed over the years, and will continue to change in coming years. The answers also depend to large extent on the level and the orientation of the educational program. I will give some of my answers to these questions.

A Two-pronged Approach to Teaching Computational Statistics

◆ David R. Hunter, Pennsylvania State University

Pennsylvania State University, Dept. of Statistics, University Park, PA 16802

dhunter@stat.psu.edu

Key Words: *computational statistics, graduate education*

Many graduate students in statistics find that mastery of certain topics in statistical computing and/or fluency in programming languages are essential to their research, yet they are typically forced to learn much of this material on their own. The need to better educate graduate students in statistical computing is clear, suggesting that computing courses should be added to the canonical list of requirements for graduate students. Yet it also appears that statistics as a discipline is not ready to replace much of the traditional canon, as exemplified by Penn State University's currently ongoing project of restructuring its graduate curriculum. One solution to this dilemma, which Penn State appears ready to adopt, is a two-pronged approach: find topics in statistical computing that have a natural affinity with subjects already taught, integrate these topics into existing courses, and then design a course in statistical computing that will address important computing topics not covered elsewhere. This talk will discuss potential advantages of this approach as well as specific topics that might be included both in the stand-alone computing course and within other courses.

Teaching Statistical Computing at NC State

◆ John Monahan, North Carolina State University

North Carolina State University, Dept. of Statistics, Raleigh, NC 27695-8203

monahan@stat.ncsu.edu

Key Words: *statistical computing, graduate curriculum*

The diverse topics that make up present a challenge for constructing the appropriate courses in statistics curricula. An undergraduate course covering programming and data management prepares students for most statistics courses. However, many graduate students successfully learn these skills on their own; moreover, squeezing such a skills course into the master's program is difficult. We teach most of the topical statistical software in the courses that use them. For doctoral students, the issue becomes one of requirements and electives: some computing topics should be covered in the required courses, others left to an elective. The selection and placement of these topics depend of the research program of the department. Even for an elective course, the research needs drive the choice of topics, although experience has shown that there can be no substitute for a good foundation in arithmetic and numerical linear algebra.

69 Data Quality and Data Confidentiality of Microdata ▲ ⊞

Section on Survey Research Methods, Section on Government Statistics, Social Statistics Section

Monday, August 9, 8:30 am–10:20 am

Protecting Quality and Confidentiality of Data by MASSC: A Survey-sampling-based Method

◆ David H. Wilson, RTI International; Avinash Singh, RTI International; Feng Yu, RTI International

RTI International, 3040 Cornwallis Rd., PO Box 12194, RTP, NC 27709

dhw@rti.org

Key Words: *disclosure, risk, SDL, MASSC, privacy, confidentiality*

The credibility of a data producer is at stake if the respondents in a database are not assured of the confidentiality of their sensitive information collected by the producer. In a scenario known as “disclosure by response knowledge,” the respondent identifies his own record from the database and is concerned about its disclosure by someone who might know enough about the respondent to identify his record. The MASSC method developed at RTI addresses the above problem using a probabilistic framework that allows for random, but controlled, perturbation and suppression. The MASSC method views a database as a population and relies on the analogy between releasing an untreated database and conducting a census of the population. Survey sampling methods are used in MASSC to provide simultaneous control and measurement of disclosure risk and information loss. The methods used in MASSC consist of four steps: micro agglomeration for partitioning the database into risk strata, probabilistic substitution for perturbation, probabilistic subsampling for suppression, and sampling-weight calibration for preserving estimates for key variables.

GUI Demonstration for MASSC Disclosure Limitation Method

◆ Donghui Wang, RTI International; Feng Yu, RTI International; David H. Wilson, RTI International

RTI International, 3040 Cornwallis Rd., PO Box 12194, Research Triangle Park, NC 27709

dwang@rti.org

Key Words: *MASSC, GUI*

A Microsoft Windows-based application has been developed for MASSC at RTI International. The methodology of the development was based on the idea that the MASSC process can be run on the standard procedures and processes. By focusing on that approach, the backend MASSC running engine was built on the standard SAS procedures and the customized C++ SAS callable packages. A GUI application was built to communicate between the MASSC users and the back-end MASSC engine. The GUI software is very user-friendly. The carefully designed user interface serves the need of saving the user interaction time, causing less confusion, and reducing the possibility of making errors. The GUI is also very flexible in terms of how each step is executed, how the outcome for each step is examined, and the accessibility to the previous steps from the current step. Some sample screens will be presented to show how the GUI works. A treated sample dataset result will also be presented with the GUI report in HTML format while the graphics will be in PDF format.

Combined-year State-level Public Use Files and Single-year Nation-level PUFs from the National Survey of Drug Use and Health Data

◆ Douglas Wright, SAMHSA; Avinash Singh, RTI International

SAMHSA, OAS/HHS, 600 Fishers Lane, Rockville, MD 20857

dwright@samhsa.gov

Key Words: *MASSC method, random substitution, random subsampling*

Since 1999, the Substance Abuse and Mental Health Services Administration (SAMHSA) has provided yearly National PUFs for NSDUH data using a procedure based on the MASSC (Micro Agglomeration, Substitution, Subsampling, and Calibration) system for statistical disclosure limitation. There is a growing demand for state-level data, and SAMHSA is considering providing state-level PUFs based on combining several years of NSDUH data. We explore various concerns and approaches to state-level PUFs and indicate how MASSC could address some of them. Releasing combined-year state-level PUFs along side single-year national PUFs poses several challenges. The most important one is that confidentiality of an individual could be compromised if an intruder is able to match the state-level PUFs with the national PUFs on the basis of various sensitive variables that are typically not perturbed, and thus may succeed in attaching state identifiers to the national PUFs. This problem can be reduced by taking advantage of the randomness in perturbation and suppression used in MASSC.

Application of MASSC to a Survey of Health-related Behaviors among Military Personnel

◆ Vincent G. Iannacchione, RTI International; George H. Dunteman, RTI International; Feng Yu, RTI International; Shijie Chen, RTI International; Donghui Wang, RTI International

RTI International, 1615 M St. NW, Suite 740, Washington, DC 20036

vince@rti.org

Key Words: *disclosure, risk, MASSC*

MASSC will be applied on a randomly selected subsample from the 2001 Department of Defense Survey of Health Related Behaviors Among Military Personnel. This worldwide survey collects data on a wide range of health-related behaviors including alcohol and drug use, tobacco use, mental health status, physical health status, and sexual activity. These are sensitive variables that need to be protected from disclosure. They are also used as outcome variables in various statistical models. Data on personal background variables are also collected. They include branch of service, pay grade, age, gender, education level, race, and ethnicity. These variables are defined as identifying variables since using these and other personal characteristics could enable an intruder to identify a respondent and disclose sensitive information. A wide range of identifying variables and sensitive outcome variables need to be included on a PUF in order for it to have analytic utility. The application of MASSC on a subsample of this survey will illustrate how MASSC can preserve the analytic utility for both descriptive analyses and statistical modeling while, at the same time, minimize disclosure risk.

Application of MASSC to Confidentiality Protection of NCHS-NHIS Data

◆ Feng Yu, RTI International; George H. Dunteman, RTI International; Lanting Dai, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

fyu@rti.org

Key Words: *disclosure risk, information loss, MASSC, NHIS, SDL*

The National Health Interview Survey (NHIS), conducted by NCHS, provides information on health related outcomes in sampled families and individuals. Protecting confidentiality of families and individuals in the family is of great concern in releasing public use files (PUFs), since they contain information on confidential health-related characteristics. We illustrate how MASSC can be used to create PUFs by viewing the existing 2000 NHIS PUF as the original dataset to protect confidentiality of sampled households as well as persons within the household. The disclosure procedure covers the entire process of Micro Agglomeration, Substitution, Subsampling, and Calibration under the MASSC frame work. MASSC ensures that there is control on bias and variance due to treatment while disclosure cost is minimized. The analysis weights are calibrated to the full sample-weighted totals for various subpopulations to improve precision of the estimates for health behaviors across domains. Both confidentiality and analytical diagnostics are presented before and after MASSC treatment of the data.

70 Professional Development of Pharmaceutical Statisticians ⌘

Biopharmaceutical Section

Monday, August 9, 8:30 am–10:20 am

What Does a Pharmaceutical Sponsor Look for in a Statistician?

◆ Christy Chuang-Stein, Pfizer Inc.

Pfizer Inc., 2800 Plymouth Rd., Ann Arbor, MI 48105

christy.j.chuang-stein@pfizer.com

Key Words: *technical skills, professional development, team approach, soft skills*

While graduate schools and many professional workshops provide training opportunities for technical skills, technical skills are only one part of the many skills a statistician needs to possess to be successful in an industry highly committed to the multidisciplinary team approach. We will focus on attributes that pharmaceutical statisticians should cultivate and steps we could take to maximize our effectiveness and success.

Professional Development of Statisticians in the Pharmaceutical Company: A Perspective from a Midsize Company

◆ Everton Rowe, Organon Pharmaceuticals Inc.; Jia-Yeong Tsay, Organon Pharmaceuticals Inc.

Organon Pharmaceuticals Inc., 56 Livingston Ave., Roseland, NJ 07068

e.rowe@organonusa.com

Key Words: *midsize, training, teamwork, resources*

The role of the statistician within the pharmaceutical industry is well established. However, the tasks performed by statisticians are also a function of the size of the organization that they work for. As part of a smaller group, statisticians at a midsize company, as a rule, will have a greater opportunity to interact with a wider range of client groups, have more visibility within the organization, may have less resources at their disposal, and make a greater impact on the organization's bottom line. Thus, their development and training must not only reflect the technical aspects of their jobs, but should also include team building and negotiation skills, problem-solving, and efficient use of resources.

Professional Development for Review Statisticians at the FDA

◆ Nancy D. Smith, U.S. Food and Drug Administration

U.S. Food and Drug Administration, HFD - 200, Room 12B-45, 5600 Fishers Lane, Rockville, MD 20857

smith@cder.fda.gov

The Center for Drug Evaluation at the FDA has developed a program to facilitate the orientation and training of review scientists at the agency, as well as their continued professional development. New scientific reviewers of all disciplines go through the "New Reviewer Workshop" and then are assigned a mentor to help them learn more about the agency and the review process. Core competencies have been developed for statistical reviewers, including communication skills, critical thinking, the drug review process, medical/pharmacologic principles, regulatory law, scientific principles, and technical writing. Learning pathways are in place to guide a statistical reviewer to develop competency in these skills over their first few years at the agency. Advanced training in these and other areas are available for continued professional development.

Continuing Statistical Education for Practicing Professionals

◆ Robert S. Schulman, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Department of Statistics, Blacksburg, VA 24061-0439

schulman@vt.edu

Key Words: *continuing education, professional development, short courses, training*

Practicing professionals must continually improve their skills and remain current with relevant developments, but generally cannot take a semester off work to take standard college or graduate courses. As a result, most continuing education opportunities

involve intensive three- to five-day training sessions. This concentrated format requires a different mindset from participants and a different presentation strategy from the instructor. Unlike familiar college courses, where the students are, say, first-year graduate students in a specific discipline, participants in continuing education seminars typically vary widely in their background, mathematical sophistication, analytical needs, and familiarity with statistical software. The author, an award-winning instructor with 30 years of college teaching experience and 20 years of practice conducting short courses, will address the special considerations associated with statistical training for practicing professionals in the pharmaceutical and other fields. These include such issues as pace and quantity of material presented, use of multiple lecture and interaction styles, and practical demonstration of statistical software.

71 **Technical Issues for Ensuring Confidentiality** ◬

Section on Government Statistics, SSC, Social Statistics Section

Monday, August 9, 8:30 am–10:20 am

Borrowing Strength from Census Data to Assess Survey Disclosure Risk

◆ Christopher Duddek, Statistics Canada

Statistics Canada, 18-O R.H. Coats Bldg., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

christopher.duddek@statcan.ca

Key Words: *disclosure risk assessment, public-use microdata files*

In practice, disclosure risk is often measured using survey data. Due to small sample sizes, however, this information can lead the statistician to alter data more than necessary in the production of public microdata files. In other words, when measuring disclosure risk, it is impossible to know if unique combinations in the sample are unique in the population since population uniqueness is typically unknown. This paper explores the potential of using a publicly available census microdata file to identify the risk of population uniqueness for an unrelated survey. Though the concept of "borrowing strength" is associated with small-area estimation, it is applicable in this context because we can use the great mass of Canadian census data to evaluate disclosure risk for a small-scale survey. Our quarterly survey, the Canadian Changes in Employment survey, is particularly appropriate for this analysis due to the fact that we have a survey frame which allows us, in a limited way, to identify population uniqueness. We compare a census-derived disclosure risk measure to a multiplicity-table measure.

Disclosure Potential in Regression Models: Some Further Results

◆ Arnold P. Reznek, U.S. Census Bureau; T. Lynn Riggs, Chicago Census Research Data Center

U.S. Census Bureau, CECON/CES Room 206 WP2, Washington, DC 20233

arnold.phillip.reznek@census.gov

Key Words: *disclosure risk, regression, confidentiality*

Previous research by the author has demonstrated that disclosure potential can exist in some types of regression models. Ordinary Least Squares (OLS), logit, probit, and Poisson regression models can present disclosure risks when the right-hand side (explanatory) variables consist entirely of fully interacted dummy (0,1) variables. In these cases, the models essentially produce tables involving the left-hand-side (dependent) variable. This paper extends that research by considering possible disclosure risks in other types of regression models. Examples of types of models to be considered include those with ordered or multicategory dependent variables (e.g., ordered probit, multinomial logit); survivial models; models with correlated error terms, including longitudinal data; and multiequation models. The paper also discusses the formation of measures of disclosure risks in regression models, beyond the criteria used for traditional statistical tables.

Iterative Rounding for Large Frequency Tables

Jean-RenÈ Boudreau, Statistics Canada; ◆ Krisztina Filep, Statistics Canada; Lin Liu, University of California, San Diego

Statistics Canada, Ottawa, ON K1A 0T6 Canada

krisztina.filep@statcan.ca

Key Words: *random rounding, controlled rounding, disclosure avoidance, confidentiality*

The Canadian Census of Population has used a rounding algorithm for tables to preserve confidentiality of respondents' answers. Although the method used is unbiased, cells are processed independently and hence, the protection of the statistical secrecy may not be as efficient as it could be if more knowledge about the table was available. It is important to analyze the table if the goal is to achieve better protection of the date. Having done this analysis, we may be in a position to make the rounded table more additive. We have designed a fast rounding algorithm that has the advantage of always producing an additive table. It uses the implicit relationship between rows and columns. However, it may not produce a controlled table. Probabilities of control are assessed. The iterative rounding algorithm takes this "semi-controlled" table as a starting point in a stepwise procedure to achieve more control.

An Application of the Salamander: Estimating the Nonmetro/Metro Difference in Wage Ginis Despite Differential Topcoding

◆ John Angle, Economic Research Service

Economic Research Service, 1800 M St., NW, Room S2060, Washington, DC 20036

jangle@ers.usda.gov

Key Words: *Gini, tail truncation, topcodes, Salamander, wages*

The Salamander is a parsimonious gamma pdf mixture model that explains five dynamic patterns in the wage distribution of the U.S. 1961-2001 by inspection of its algebra and two more numerically. The Current Population Survey, March 1962-1967, whose minimum topcodeable annual wage income, $99,900, was 99.99+th percentile permits the evaluation of the Salamander against the lowering of the minimum topcodeable income. The Salamander’s estimate of the difference between the Ginis of nonmetro and metro wages is robust against a simulated lowering of the minimum topcodeable income, which truncates more of the metro than the nonmetro distribution’s right tail, down to a perhaps surprisingly low quantile.

Comparing Statistical Disclosure Control Methods for Tables: Identifying the Key Factors

◆ Paul B. Massell, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Statistical Res. Div., Rm. 3209-4, Washington, DC 20233-9100

paul.b.massell@census.gov

Key Words: *statistical disclosure control, confidentiality, cell suppression, cell perturbation*

In recent years there has been much research on statistical control methods for tables. In some cases, new developments in mathematical programming and related operations research techniques have led to great speed-ups in the implementation of existing methods (e.g., cell suppression). In other cases, there have been new approaches to the method itself (e.g., cell perturbation and cell rounding). The result is that there are likely to be two or more methods that can be applied to any set of tables that must undergo disclosure control. How does an agency or statistical organization determine which method is best to use? We try to identify the key factors in that decision and discuss how to apply them to a given method.

72 Statistical Methods in Molecular Epidemiology ▲

Section on Statistics in Epidemiology

Monday, August 9, 8:30 am–10:20 am

Estimating the Haplotype Specific Disease Risk in a Case-control Study

◆ E.S. Venkatraman, Memorial Sloan-Kettering Cancer Center; Nandita Mitra, Memorial Sloan-Kettering Cancer Center; Colin B. Begg, Memorial Sloan-Kettering Cancer Center

Memorial Sloan-Kettering Cancer Center, 1275 York Ave., Dept. of Epidemiology & Biostatistics, New York, NY 10021

venkatre@mskcc.org

Key Words: *haplotype, case-control, Hardy-Weinberg equilibrium*

Estimation of the association between haplotypes and disease from a case-control study is considered. Assuming a single “disease haplotype” leads to increased risk, attention focusses on the relative risks associated with a single copy or two copies of the disease haplotype, relative to individuals with no copies. In this setting, case frequencies of the haplotype pairs are in Hardy-Weinberg Equilibrium (HWE) only if the influence of two copies of the disease haplotype on risk is multiplicative. Thus, imputation cannot rely on the assumption of HWE for cases. A method is presented for obtaining unbiased estimates of relative risks, making use of the EM algorithm and the assumption of HWE only for controls. The method accounts for the additional variation in the estimates due to the imputation of expected frequencies of haplotype pairs from ambiguous genotypes. A simulation study shows that the resulting confidence intervals have nominal coverage, and that the methods based on the assumption of HWE for both cases and controls can lead to bias.

Reducing the Genotyping Needed for Case-parents Designs by Pooling DNA Samples

◆ David M. Umbach, National Institute of Environmental Health Sciences; Clarice R. Weinberg, National Institute of Environmental Health Sciences

National Institute of Environmental Health Sciences, Mail-Drop A3-03, PO Box 12233, Research Triangle Park, NC 27709-2233

umbach@niehs.nih.gov

To study disease-related genes, one can genotype affected children and their parents (triads), achieving robustness to genetic population structure. However, for rare alleles, many families carry no copies and are discarded as uninformative. To reduce assays while retaining robustness, we considered design variations in which DNA is pooled before assay. We assume that the assay for each diallelic locus can count the proportion of variant alleles among the 2k present in a pool from k individuals. If DNA from the two parents in each triad is pooled but the child is assayed separately, one can carry out a test that resembles the TDT while reducing assays by one-third. One can also analyze such data via log-linear models. A second design reduces genotyping by two-thirds by randomly matching pairs of triads and assaying one pool of the four parental samples and a second of the two offspring samples. The data can be analyzed with a log-linear modeling approach via the EM algorithm. We study the operating characteristics of these designs under different allele frequencies and modes of inheritance. DNA pooling can be advantageous when specimen volumes are limited or the allele is rare.

A Two-stage Regression Model for Epidemiological Studies with Multivariate Disease Classification Data

◆ Nilanjan Chatterjee, National Cancer Institute

National Cancer Institute, EPS 8038 6120 Executive Blvd., Rockville, MD 20852

chattern@mail.nih.gov

Key Words: *molecular epidemiology, disease classification, semiparametric method*

Advances in clinical and molecular characterization of cancer provide new opportunities to study “etiologic” and “treatment” heterogeneity, i.e., to determine whether effects of exposures or treatments are different for different subtypes of a disease. Polytomous logistic regression, commonly used for analyzing heterogeneity between two or more distinct subtypes of a disease, gives odds ratio estimates for each disease subtype compared to a common group of nondiseased controls, but is not suitable for evaluating the effects of risk factors when disease groupings are defined by multiple overlapping tumor characteristics or markers. We propose a novel two-stage modeling approach for analyzing epidemiologic studies with data on multiple disease markers. The first-stage model involves defining polytomous logistic regression parameters for a set of markers, comparing all possible marker-defined disease subtypes to the common control group. The second-stage model further characterizes the exposure odds ratios for the first-stage disease subtypes in terms of the underlying disease markers in terms of a regression model.

Modeling Measurement Error in a Biomarker on the Pathway from Smoking to Lung Cancer

◆ Sally W. Thurston, University of Rochester

University of Rochester, Dept. of Biostatistics and Computational Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

thurston@bst.rochester.edu

Key Words: *Bayesian, biomarker, DNA adducts, lung cancer, measurement error, smoking*

Molecular biologists have identified specific cellular changes, called biomarkers, which enable them to better characterize the pathway from chemical exposure to initiation of some cancers. In lung cancer, one such biomarker is DNA adducts in lung tissue. Carcinogens derived from cigarette smoke can bind to DNA to form such adducts, and this process is believed to initiate smoking-induced lung cancer. The goal of this work is to incorporate knowledge of such underlying biological mechanisms into a useful statistical framework to improve cancer risk estimates. The model uses measurements of adducts in lung cells and in blood cells; the latter are needed because lung adducts cannot be measured in controls. Adduct measurements in each type of cell are known to vary within individuals. By introducing a latent variable for true lung DNA adducts, I allow for measurement error in both types of observed adduct measurements, but assume greater measurement error in blood adducts. Gibbs sampling is used to obtain the posterior distributions of model parameters. Predicted and observed case status agree for approximately 75% of the sample.

73 Time Series Research at Statistics Canada and the U.S. Census Bureau ▲ ⌘

Business and Economics Statistics Section, Section on Government Statistics, SSC

Monday, August 9, 8:30 am–10:20 am

An Iterated Parametric Approach to Nonstationary Signal Extraction

◆ Tucker S. McElroy, U.S. Census Bureau

U.S. Census Bureau, Statistical Research Division, Washington, DC 20233-9100

tucker.s.mcelroy@census.gov

Key Words: *ARIMA component model, nonstationary time series, seasonal adjustment, signal extraction, Wiener-Kolmogorov filtering, X-11*

Consider the three-component time series model that decomposes observed data (Y) into the sum of seasonal (S), trend (T), and irregular (I) portions. Assuming that S and T are nonstationary, it is demonstrated that widely used Wiener-Kolmogorov signal extraction estimates of S and T are related by simple formulas, and can be obtained through an appropriately initialized iteration scheme applied to optimal estimates derived from reduced two-component models Y^S and Y^T consisting of S plus I and T plus I, respectively; this “bootstrapping” signal extraction methodology is reminiscent of X-11’s iterated nonparametric approach. The contraction-mapping analysis of the iteration scheme provides geometric intuition for the algebraic relationship between full-model and reduced-model signal extraction estimates.

Bayesian Inference for Seasonal Adjustment with Sampling Error: Application to Estimates from Monthly Value of Construction Put-in-place Surveys

◆ William R. Bell, U.S. Census Bureau; Thuy T.T. Nguyen, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Room 3000-4, SRD, Washington, DC 20233-9100

William.R.Bell@census.gov

Key Words: *signal extraction, sampling error model*

The U.S. Census Bureau estimates “value of construction put in place” (VIP) each month using a variety of sources, including sample surveys, administrative records, and trade association data. Recently the VIP estimates were expanded to more levels of detail, but some of the resulting direct estimates have large sampling errors. The short length of the VIP time series (six years), coupled with the large sampling errors for some series, makes seasonal adjustment of these time series difficult. With model-based seasonal adjustment this difficulty manifests itself in large amounts of uncertainty about the model parameters. This paper examines use of Bayesian methods in model-based seasonal adjustment to recognize uncertainty about parameters in both the model for the true unobserved time series and in the sampling error model. The latter accounts for sampling error variances and autocorrelations of the VIP survey estimates. Because of the short length of the VIP time series, accounting for parameter uncertainty is important to judging the accuracy of the seasonal adjustments.

Adjustment of Seasonally Adjusted Series to Annual Totals

◆ Benoit Quenneville, Statistics Canada; Pierre Cholette, Statistics Canada; Guy Huot, Statistics Canada; Kim Chiu, Statistics Canada; Tommaso Di Fonzo, Universita di Padova

Statistics Canada, 120 Parkdale Ave., 17-RHC-BSMD, Ottawa, ON
K1A 0T6 Canada

benoit.quenneville@statcan.ca

Key Words: *benchmarking, generalized least squares, Henderson filters, moving averages, seasonal adjustment*

This paper proposes methods to correct a seasonally adjusted series so that its annual totals match those of the raw series. The proposed methods are designed to preserve period-to-period movement like the Denton type of adjustments, but improve the adjustment at the ends. The methods can be applied to seasonally adjusted series obtained with any seasonal adjustment method such as X-11, X-11-ARIMA, X-12-ARIMA, TRAMO-SEATS and STAMP. The methods are illustrated with a seasonally adjusted series obtained with either X-11-ARIMA or X-12-ARIMA.

Modifications of SEATS Diagnostics for Under- and Overestimation

◆ David F. Findley, U.S. Census Bureau; Kellie Wills, U. S. Census Bureau; Tucker S. McElroy, U.S. Census Bureau

U.S. Census Bureau, Statistical Research Division, Rm. 3000-4, Washington, DC 20233-9100

david.f.findley@census.gov

Key Words: *TRAMO/SEATS, residual seasonality, underadjustment, overadjusment, stability, sliding spans*

We consider SEATS diagnostic to determine whether, for an estimated seasonal decomposition component, there is underestimation or overestimation, meaning inadequate or excessive suppression of frequency components near the targeted suppression frequencies, e.g. the seasonal frequencies in the case of seasonal adjustment. The diagnostic in SEATS depends on variance estimates that assume an infinitely long filter has been applied. For the technically simplest case, estimation of the irregular component, we show that this results in substantial bias toward indicating overestimation. We consider two much less biased modifications of the diagnostic that are calculated from time varying variances associated with the finite-length filters and an associated test for the statistical significance of any indicated misestimation. The results indicate that this approach cannot detect overestimation reliably and is somewhat compromised in the detection of underestimation, because of the asymmetry of the distributions of the diagnostics.

Clustering Time Series: An Application to Seasonal Adjustment

◆ Dominique Ladiray, INSEE

dladiray@yahoo.fr

Key Words: *seasonal adjustment, cluster analysis*

Cluster analysis has been used for decades, not only in Survey Data Analysis, but also in Time Series Analysis especially to improve the quality of forecasts. Algorithms and techniques used in classical cluster analysis must of course be adapted to take into account the nonstationarity of most economic time series. A lot of new techniques have been developed these recent years and are now available. We briefly present these techniques, then show that cluster analysis can be quite surprisingly used in the seasonal adjustment context as a very convenient tool to compare methods and softwares and/or to determine optimal values for the main parameters of seasonal adjustment softwares.

74 Bayesian Methods in Generalized Linear Models

Section on Bayesian Statistical Science

Monday, August 9, 8:30 am–10:20 am

Generalized Mixed Nonlinear Model Using Penalized Splines

◆ Tapabrata Maiti, Iowa State University

Iowa State University, Dept. of Statistics, Ames, IA 50011

taps@iastate.edu

Key Words: *mixed models, penalized spline, additive models*

In generalized mixed linear models, its is assumed that the mean is a known function of a linear combination of covariates and random effects. Often this assumption is not valid. We propose a generalized mixed nonlinear model where this linear combination is replaced by the additive functions of covariates. In particular, we have used penalized spline to model the nonlinearity. The methodology is illustrated using a data from prostate cancer study.

Multiclass Cancer Diagnosis Using Bayesian Kernel Machine Models

◆ Bani K. Mallick, Texas A&M University

Texas A&M University, Dept. of Statistics, TAMU 3143, College Station, TX 77843-3143

bmallick@stat.tamu.edu

Key Words: *microarrays, cancer classification, support vector machine, Bayes classification, MCMC*

Precise classification of tumors is critical for cancer diagnosis and treatment. Using gene expression data to classify tumor types is a very promising tool in cancer diagnosis. In recent years, several works showed successful classification of pairs of tumors types using gene expression patterns. However, the simultaneous classification across a heterogeneous set of tumor types has not been well-studied yet. Usually, this multicategory classification

problems are solved by using a binary classifiers which may fail in a variety of circumstances. We tackle the problem of cancer classification in the context of multiple tumor type. We develop a full probabilistic model-based approach, specifically probabilistic relevance vector machine (RVM), as well as support vector machines for multicategory classification. We develop a hierarchical model where the unknown smoothing parameter is interpreted as a shrinkage parameter. We assign a prior distribution to it and obtain its posterior distribution via Bayesian computation. In this way, we not only obtain the point predictors but also the associated measures of uncertainty.

Semi-automatic Informative Priors via the Relationship between the Power Prior and Hierarchical Models

◆ Ming-Hui Chen, University of Connecticut; Joseph G. Ibrahim, University of North Carolina, Chapel Hill

University of Connecticut, Dept. of Statistics, 215 Glenbrook Rd., U-4120, Storrs, CT 06269-4120

mhchen@stat.uconn.edu

Key Words: *generalized linear model, historical data, power prior, prior elicitation, random effects model*

The power prior has emerged as a useful informative prior for the incorporation of historical data in a Bayesian analysis. We examine formal analytical relationships between the power prior and hierarchical models, and show that the hierarchical models used for combining several datasets are a special case of the power prior. We establish these results for the normal linear model as well as for the class of generalized linear models. These analytical relationships are quite novel as they unify the theory of the power prior, demonstrate the generality of the power prior, shed new light on benchmark analyses using the power prior, and provide key insights to the estimation of the power parameter in the power prior. Several key theorems are presented establishing these formal connections, as well as a formal methodology for estimating a guide value for the power parameter. Several examples are given to illustrate the proposed methodology.

Bayesian Hierarchical Linear Mixed Models for Additive Smoothing

◆ Dongchu Sun, University of Missouri, Columbia; Paul L. Speckman, University of Missouri, Columbia

University of Missouri, Columbia, 146 Middlebush Hall, Columbia, MO 65201

dsun@stat.missouri.edu

Key Words: *smoothing spline, additive models, hierarchical models, variance components, noninformative priors*

Bayesian hierarchical models have been used for smoothing splines, thin-plate splines, and L-splines. In analyzing high-dimensional datasets, additive models and backfitting methods are often used. A full Bayesian analysis for such models may include a large number of random effects, many of which are not intuitive, so researchers typically use noninformative improper or nearly improper priors. We investigate propriety of the posterior for these cases. Our findings extend known results for normal linear mixed models to certain cases with Bayesian additive smoothing spline models.

Bayesian Inference about Relative Importance of Variables

◆ Ehsan S. Soofi, University of Wisconsin, Milwaukee; Joseph J. Retzer, Maritz Research

University of Wisconsin, Milwaukee, School of Business Administration, PO Box 742, Milwaukee, WI 53201

esoofi@uwm.edu

Key Words: *Bayes factor, entropy, mutual information, uncertainty reduction, variable selection*

Comparison of the relative importance of explanatory variables appear in many disciplines. Various measures of relative importance have been proposed for regression, ANOVA, logit, and survival analysis. Attempts have been made to define a framework for and requirements of relative importance measures. However, little attention has been paid to characterizing the more general, underlying notion of "importance." It is natural to measure the importance of an explanatory variable by the extent to which its use reduces uncertainty about predicting the outcome of the dependent variable, namely, its information content. Information importance measures are applicable to GLM with qualitative variables and quantitative variables. For normally distributed variables and for variables that can be transformed to normality, the information measures are functions of simple and multiple correlations. These information measures are equivalent to the sequential Bayes factors for the normal regression. For the normal regression, inference about the relative importance of explanatory variables and about the differences between their relative importance measures are performed.

75 Statistical Literacy 2004

Section on Statistical Education

Monday, August 9, 8:30 am–10:20 am

Measuring the Health of Populations: Teaching Statistical Principles Using Epidemiology

◆ Donna F. Stroup, Centers for Disease Control and Prevention; Richard A. Goodman, Centers for Disease Control and Prevention; Ralph Cordell, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 4770 Buford Hwy. NE, Mail Stop K40, National Center for Chronic Disease Prevention, Atlanta, GA 30341

dstroup@cdc.gov

Key Words: *epidemiology, statistical education, public health, electronic teaching materials, quantitative literacy, numeracy*

Recent developments in education have addressed the need to enhance the teaching of statistics to improve the statistical literacy and scientific reasoning of students. Nonetheless, many students perceive their statistics instruction as unrelated to decisions in their lives, and data show that students in the United States lag behind students in other parts of the world in

mathematics and science achievement. We argue that epidemiology, the scientific basis for public health, provides a useful and motivating context for teaching statistical principles and methods and suggest that examples from this and other public health sciences be used in teaching. First we describe resources developed by the Centers for Disease Control and Prevention to aid in implementing epidemiology in statistics education, and present evaluative evidence of the effectiveness of these resources when implemented at the high school level. Next, we illustrate how these resources address selected college mathematics and high school statistics education standards. Finally, we show how teaching statistics in the context of epidemiology responds directly to challenges in statistical teaching.

Planning a Statistical Literacy Program at the College Level: Musings and a Bibliography

◆ Robert W. Hayden, Plymouth State University

Plymouth State University, 82 River St., Ashland, NH 03217

hayden@mail.plymouth.edu

Key Words: *statistical literacy, education*

Much has been written in recent years on statistical literacy, but what do we mean by “statistical literacy”? I will take a pragmatic approach and provide resources to help you to define this term for yourself and implement your idea of statistical literacy in the classroom. This paper includes a bibliography of relevant resources. The text provides notes on the bibliography and ruminates on the issues involved. Even if my conclusions differ from yours, the bibliography and discussion may still prove useful in defining, defending and implementing a statistical literacy program.

A Practical Approach to the Introductory Political Science Statistics Course

◆ Gary M. Klass, Illinois State University

Illinois State University, 4600 Political Science, Normal, IL 61790-4600

gmklass@ilstu.edu

Key Words: *education, graphical display, multivariate regression, political science, social indicators*

The methodological pluralism of the political science discipline presents unique challenges to the design of an introductory research methods and statistics course. Conventional approaches to social science statistics education stress statistical concepts and applications that are often of little practical use to students in their subsequent coursework or careers. Political scientists typically use numerical data in their research and writing in one of two ways that traditional methods courses do not teach. They present and interpret political and social indicator data, often time series data, in tabular or graphical format, or they use some form of multiple regression. This presentation describes an approach to a statistics course more relevant to the actual practice of political science. It begins by teaching students how to obtain, interpret, and present political and social indicator data, skills involving a variety of applications across even “nonquantitative” fields of political science. It ends by trying to get students to the point where they can at least understand the regression-based research of quantitative political science.

Influence of Binary Confounders on Associations Using a Noninteractive Model

◆ Milo Schield, Augsburg College; Thomas Burnham, Cognitive Consulting

Augsburg College, 2211 Riverside Dr., Minneapolis, MN 55410

milo@pro-ns.net

Key Words: *epidemiology, causation, significance*

The defining and sufficient conditions for a binary confounder to nullify or reverse an association involving two binary factors have been previously identified using a noninteractive model. This paper investigates various models or summaries of these binary confounder conditions or families using confounder size. The relationship between confounder size and the association that can be nullified or reversed thereby is determined. An association that is confounder-resistant to nullification or reversal by confounders below a given size can be viewed as “confounder significant” against confounders up to this size. This approach allows researchers, editors, and journalists to set a minimum standard for the confounder significance of associations in the same way that alpha is used to set a minimum standard for statistical significance. Being confounder significant is argued to have an increasing importance as the size of datasets increase and the effect sizes that are statistically significant decrease. This paper also extends the analysis of outcomes from binary to continuous.

76 Issues in Weighting and Estimation for Household Surveys

▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Monday, August 9, 8:30 am–10:20 am

Earned Income and Longitudinal Attrition in the SIPP and SPD

◆ David W. Hall, U.S. Census Bureau; Smanchai Sae-Ung, U.S. Census Bureau; Jan Tin, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Room 3725-3, Mailstop 8700, Suitland, MD 20747

david.warren.hall@census.gov

Key Words: *longitudinal survey, nonresponse, attrition bias, earned income, survey data quality, weighting*

Longitudinal surveys, like the Survey of Income and Program Participation (SIPP) and the Survey of Program Dynamics (SPD), experience growing attrition over the life of a survey panel. With each attriter, we lose information on variables such as earned income, which is commonly regarded as an important characteristic that is indicative of the socioeconomic well being of an individual and his/her family. We conduct a statistical evaluation of the longitudinal difference in earned incomes between the attriters and the continuers in the SPD and the 1996 SIPP panel surveys using Social Security Administration (SSA) earnings records. We also analyze the difference between the earned income data collected from each survey and those from SSA earnings records.

Evaluation of Variance Estimation Methods Based on a SPD Longitudinal File

◆ Smanchai Sae-Ung, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver HIll Rd., Room 3725-3, Mail Stop 8700, Suitland, MD 20747

smanchai.sae.ung@census.gov

Key Words: *variance estimation, replicate weights, balanced repeated replicate, successive difference replicate, residual linearization, complex survey*

The Survey of Income and Program Participation (SIPP) traditionally uses Fay's balanced repeated replication method for direct variance estimation, while the Current Population Survey (CPS) uses the successive difference replication method. The DataFerrett system (through the U.S. Census Bureau web site) may use the residual technique of linearization as outlined by Deville for variance estimation in the future due to its expediency in interactive computation. We evaluate variance estimates produced by these three methods for key characteristics of the Survey of Program Dynamics (SPD) based on one of its longitudinal files. We determine the differences in the variance estimates produced by the three methods and the causes of the differences, and assess the effects of the differences on statistical inferences of the SPD's key characteristics.

The Effects of Oversampling on the National Crime Victimization Survey

◆ Katrina Washington, U.S. Census Bureau; Barbara Blass, U.S. Census Bureau; Karen King, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233

katrina.t.washington@census.gov

Key Words: *NCVS, oversampling, crime*

In an ongoing effort to find more efficient and cost-effective ways to conduct the NCVS, the Bureau of Justice Statistics (BJS) has asked the Bureau of the Census to research the implications of oversampling within Primary Sampling Units (PSUs) that have high Uniform Crime Reporting (UCR) rates. The 2000 sample design for the NCVS is a self-weighting design. If oversampling is done, this design would have to be manipulated prior to the sample being released in the field. Since BJS is not interested in increasing their overall housing unit sample size, the process of oversampling is being explored as a type of reallocation, where two different methods of reallocating sample from PSUs with low crime rates to PSUs with high crime rates are explored. Both of these methods of reallocation are being simulated using NCVS data from 1999-2001. These simulations are being used to draw conclusions about the effects that oversampling may have if implemented as part of the 2000 sample design for the NCVS. This paper will present what effects this change in design could have on the crime estimates, variances, and CVs for the NCVS.

Weighting Alternatives to Compensate for Longitudinal Nonresponse in the Survey of Income and Program Participation

◆ Leroy Bailey, U.S. Census Bureau

U.S. Census Bureau, Statistical Research Division, Rm. 3134-4, Washington, DC 20233

leroy.bailey@census.gov

Key Words: *longitudinal weighting, nonresponse mechanism, logistic regression*

In the longitudinal estimation for the Survey of Income and Program Participation (SIPP), a weighting-cell adjustment approach is currently employed to account for nonresponse. This paper presents the results of a study of the impact of selected nonresponse weighting schemes on SIPP longitudinal estimates. Emphasis is initially placed on the modeling of the nonresponse mechanism. A general framework for the weighting alternatives is presented, and empirical comparisons and evaluations of the procedures are provided, based on panel data for several of the survey's principal items.

Multivariate Modeling and Standard Error Estimation Using the Current Population Survey: An Examination of Various Approaches

◆ Arthur F. Jones, U.S. Census Bureau; Gestur Davidson, University of Minnesota; Lynn A. Blewett, University of Minnesota; Michael Davern, University of Minnesota; James Lepkowski, University of Michigan

U.S. Census Bureau, Dept. of Commerce, Washington, DC 22303

arthur.jones.jr@census.gov

Key Words: *variance, income, poverty, CPS-ASEC*

Multivariate statistical analysis of the Current Population Survey (CPS) is widely used in the social science and public health literature as a source of evidence. However, the statistical evidence cited from the CPS not be comparable because researchers use a variety of statistical methods to produce standard errors used in significance tests. We examine five common methods of standard error estimation: (1) the simple random sample estimator, (2) the generalized variance estimator, (3) first order Taylor series robust estimator, (4) a survey design based Taylor Series estimator using the public use file, and (5) a survey-design-based Taylor Series estimator using the internal Census data file. We restrict our analysis to the 2003 Annual Demographic Supplement and we estimate a basic multivariate statistical model to predict personal income, health insurance coverage and poverty. Our findings will help researchers who use the public CPS data for multivariate statistical modeling know how well various approaches to standard error estimation works.

77 Medical Expenditure Panel Survey ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics, Social Statistics Section, Section on Health Policy Statistics

Monday, August 9, 8:30 am–10:20 am

Overview of Methodology for Imputing Missing Expenditure Data in the Medical Expenditure Panel Survey

◆ Steven R. Machlin, Agency for Healthcare Research and Quality; Deborah D. Dougherty, Westat

Agency for Healthcare Research and Quality, CFACT, 540 Gaither Rd., Rockville, MD 20850

smachlin@ahrq.gov

Key Words: *hot-deck imputation, MEPS, health expenditures*

In the Medical Expenditure Panel Survey (MEPS), expenditures are defined as payments from all sources (including individuals, private insurance, Medicare, Medicaid, and other sources) for health care services during the year. Data on expenditures are collected for sample persons in the Household Component of the survey, and from a sample of their health care providers responding to the Medical Provider Component of the survey. In the absence of payment information from either component, expenditure data are completed through weighted hot-deck imputation procedures. The MEPS collects a wide variety of data about individuals and health care events that are correlated with expenditures and a selected set of these variables are used in the imputation processes. Several hot-deck iterations are run for medical event type category in the survey (e.g., doctor visits, hospitalizations, etc.) based on factors such as whether partial payment information was reported and whether payments for the event covered multiple visits. This paper provides an overview of the methodological approach to impute MEPS expenditure data and how class variables for the hot deck procedures were determined.

Overview of Class Variables Used to Impute Selected Missing Expenditures in the Medical Expenditure Panel Survey

◆ Marc W. Zodet, Agency for Healthcare Research and Quality; Diana Z. Wobus, Westat; Steven R. Machlin, Agency for Healthcare Research and Quality

Agency for Healthcare Research and Quality, 540 Gaither Rd., Room 5058, Rockville, MD 20850

mzodet@ahrq.gov

Key Words: *MEPS, expenditures, hot-deck imputation*

The Medical Expenditure Panel Survey (MEPS) collects data on health care utilization, expenditures, sources of payment, insurance coverage, and health care quality measures. The survey was designed to produce national and regional estimates for the U.S. civilian noninstitutionalized population. The data on medical expenses are collected from both household respondents in the Household Component and from a sample of their health care providers in the Medical Provider Component. In the absence of payment information from either component, expenditure data are derived for sample persons through an imputation process. Missing expense data are imputed at the event level for each medical event type using a weighted hot-deck procedure. This process utilizes individual and event level data collected in MEPS that are correlated with medical expenditures. Bivariate analyses and linear regression models were utilized to assess the current class variables used for imputation. This paper details the methodology used to select, prioritize, and categorize the class variables used to impute missing expenditures for three event types: doctor visits, hospitalizations, and home health visits.

Investigation of the Impact of Imputation on Variance Estimation in the Medical Expenditure Panel Survey

◆ Robert M. Baskin, Agency for Healthcare Research and Quality; Lap-Ming Wun, Agency for Healthcare Research and Quality; John Sommers, Agency for Healthcare Research and Quality; Marc W. Zodet, Agency for Healthcare Research and Quality; Steven R. Machlin, Agency for Healthcare Research and Quality; Trena M. Ezzati-Rice, Agency for Healthcare Research and Quality

Agency for Healthcare Research and Quality, 540 Gaither Rd., Rockville, MD 20850

rbaskin@ahrq.gov

Key Words: *multiple imputation, Rao-Shao adjustment, hot-deck, Bayesian bootstrap, imputation cell, missing at random*

The Medical Expenditure Panel Survey (MEPS, a national probability sample survey sponsored by the Agency for Healthcare Research and Quality), is designed to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS, like most sample surveys, experiences unit and item nonresponse despite efforts to maximize response. At the present, a form of a weighted sequential hot-deck is used to impute missing values of health care expenditures, along with other variables, but currently it is difficult to assess the impact that imputation has on the variance estimates produced from MEPS data. The purpose of this study is to begin an assessment of the impact of this imputation on estimates of variance for estimates of health care expenditures as well as other key variables that are imputed. The assessment involves an evaluation of estimated variances using the Rao-Shao adjustment for imputation within replicates and variances produced by multiple imputation versus the "naïve" estimate of the variances that treat the imputed values as if they were observed.

Examination of Skewed Health Expenditure Data from the Medical Expenditure Panel Survey

◆ William W. Yu, Agency for Healthcare Research & Quality; Steven R. Machlin, Agency for Healthcare Research and Quality

Agency for Healthcare Research & Quality, 540 Gaither Rd., Room 5054, Rockville, MD 20850

wyu@ahrq.gov

Key Words: *medical expenditures, skewness, sample size, confidence interval*

The Medical Expenditure Panel Survey Household Component (MEPS-HC) is designed to provide nationally representative annual estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. The expenditure data from MEPS have been shown to exhibit a marked positive skewness, with a few high expenditure respondents and many low or zero expenditure respondents. As a consequence of this departure from the normal distribution, the frequency with which a conventional confidence interval for a MEPS expenditure estimate will not capture the true population parameter may be higher than the probability stated for the confidence interval. Based on repeated sample simulations using data from the 1996 to 2001 MEPS-HC, this paper evaluates and compares the "actual" probability achieved for confidence intervals derived from expenditure data by types of expenditure and varying sample sizes. The results are also compared to estimated confidence probabilities obtained from repeated sample simulations for other types of variables that do not exhibit as marked a positive skewness as health care expenditures.

Predicting Medical Expenditures Using Survey Data

◆ Joel W. Cohen, Agency for Healthcare Research and Quality

Agency for Healthcare Research and Quality, 540 Gaither Rd., Rockville, MD 20855

jcohen@ahrq.gov

Key Words: *medical expenditures, predicting expenditures, persistence of expenditures, Medical Expenditure Panel Survey*

Research on the distribution of health care expenditures among the U.S. population has shown that a small proportion of individuals accounts for a disproportionate share of total expenses. Work by Berk and Monheit using the 1977 National Medical Care Expenditure Survey (NMCES), 1987 National Medical Expenditure Survey (NMES) and the 1996 Medical Expenditure Panel Survey (MEPS) indicates that the top 1% of that distribution accounts for more than one-quarter of all expenses, and the top 5% accounts for more than half. Additional work on the persistence of high expenditures, however, suggests that in general being in the top of the expenditure distribution is not something that is highly persistent over time. Although expenditures in one year are correlated with expenditures in the next, there are a number of other factors that determine individuals' levels of spending, including his or her health conditions. The ability to accurately predict expenditures is critical for both ensuring adequate coverage of high expenditure individuals in survey samples, and for determining efficient treatment regimens for these individuals in both private and public health plans.

78 Statistical Methodology in the Analysis of Brain Imaging Data ◬ ⌘

Biometrics Section

Monday, August 9, 8:30 am–10:20 am

Finding Consistent Activation Across fMRI Tasks: Using pFDR to Test a Union of Nulls

◆ Thomas E. Nichols, University of Michigan

University of Michigan, 1420 Washington Heights, Biostatistics, Ann Arbor, MI 48109

nichols@umich.edu

Key Words: *multiple comparisons, medical imaging, functional magnetic reasonance imaging, false discovery rate*

Psychologists studying memory use Functional Magnetic Resonance Imaging (fMRI) to understand how information is encoded, stored and retrieved in the brain. Short-term, or working memory, can be "probed" in a number of ways, say. By asking a subject to remember a list of words, or a list of digits, or even even a collection of shapes (triangles, squares, etc.). The goal is to identify regions of the brain that support working memory generically, but which are not specialized to words or numbers or shapes. This requires testing a union of nulls (no effect in one or more of the three tasks) versus an intersection of alternatives (effects in all three tasks). Worsley and Friston proposed using the maximum p value to test for a intersection of effects, but their inference is based on the intersection of nulls. We propose using Storey's Positive False Discovery Rate (pFDR) to make inference on the union of nulls. Storey shows that pFDR can be interpreted as the posterior probability of the null given that a statistic lies in a rejection region. Our method is simple, and amounts to making inference on the sum of q values. We demonstrate the method on simulated and real data.

Autocorrelation Model Selection in fMRI Analysis

◆ Jeanette Mumford, University of Michigan; Thomas E. Nichols, University of Michigan; Wen-Lin Luo, University of Michigan

University of Michigan, Dept. of Biostatistics, 1000 Island Dr. Ct. Apt 104, Ann Arbor, MI 48105

jmumford@umich.edu

Key Words: *fMRI time series, correlation model identification, goodness of fit*

A functional magnetic resonance imaging (fMRI) data exhibit temporal autocorrelation. It is crucial to identify an accurate autocorrelation model so that efficient inferences can be obtained. In time series analysis, ACF and PACF are usually used to explore the possible autocorrelation model. In neuroimaging, this work breaks down since it is not practical to look at 100,000 plots, one for each volume element. Therefore we need a way to efficiently identify the best autocorrelation model over the entire brain. We apply model selection techniques to choose the fMRI autocorrelation model. We use simulation to understand the performance of different model selection criteria (AIC, BIC, MDL, ICOMP, and NURE) in selecting the autocorrelation model. Based on the simulation results and combined with the graphical exploration in time and frequency domain, we develop a systematic approach to modeling fMRI autocorrelation. With our previous work on linear model diagnosis, we can assess the goodness of fit of both the mean and autocorrelation model. We demonstrate our method by a real fMRI data analysis.

Functional Logistic Regression Models in Brain Imaging

◆ Todd Ogden, Columbia University

Columbia University, Dept. of Biostatistics, 6th Floor, 722 W. 168th St., New York, NY 10032

to166@columbia.edu

Key Words: *brain imaging, functional data, logistic regression*

Dynamic PET scans are administered to a sample of depression subjects and in controls in a study to measure neuroreceptor density throughout the brain. The resulting co-registered brain maps are regarded as functional predictors in a logistic regression model with depression status as the outcome measure.

Recursive Partitioning for Changing Kriging Models in Brain Regions

◆ William R. Schucany, Southern Methodist University; Patrick S. Carmack, Southern Methodist University; Richard F. Gunst, Southern Methodist University; Wayne A. Woodward, Southern Methodist University; Jeff Spence, Southern Methodist University

Southern Methodist University, Dept. of Statistical Science, Dallas, TX 75275-0332

schucany@smu.edu

Key Words: *changepoint, regression trees, spatial correlation, PET, fMRI, voxels*

A brain image analysis package like SPM yields a comprehensive analysis of the entire brain, but it does so at the cost of a special global approximation for spatial correlations. On the other hand, spatial kriging models have been fit in manually selected spatially homogenous regions. While these models take advantage of the regional correlation structure, they fail to yield a comprehensive analysis of the entire brain. Recursive partitioning promises to automate the placement of localized spatial kriging models. These models exploit the correlation present in imaging data and potentially yield a spatial analysis of the entire brain. Demonstrations by examples and simulations will be presented.

Accommodating Spatial Correlations in Brain-imaging

◆ Richard F. Gunst, Southern Methodist University; Patrick S. Carmack, Southern Methodist University; William R. Schucany, Southern Methodist University; Jeff Spence, Southern Methodist University ; Wayne A. Woodward, Southern Methodist University

Southern Methodist University, Dept. of Statistical Science, Dallas, TX 75275-0332

rgunst@smu.edu

Key Words: *kriging, spatial modeling, SPECT, fMRI*

Brain-imaging produces hundreds of thousands of measurements throughout the brain for each subject in one or more test groups. Group comparisons are often based on the analysis of t- or F-statistics calculated from each of these many locations in each group of subjects. Since these statistics are highly correlated for neighboring locations in the brain, random field theory has been used to highlight areas of activation that differ among the groups. When interest is in specific regions of the brain, spatial statistical modeling can be an effective alternative. Classical spatial modeling methods can be combined with factor-effects models to provide more powerful group comparisons within these specified regions of interest. Experience gained from an analysis of SPECT images of Gulf War Syndrome subjects will be detailed.

79 Survey Analysis ⌘

Section on Survey Research Methods, General Methodology

Monday, August 9, 8:30 am–10:20 am

Estimation Challenges Associated with the Increase of Use of the Tax Data for the Unified Enterprise Survey

◆ Claude Nadeau, Statistics Canada; Sylvie Gauthier, Statistics Canada

Statistics Canada, R.H. Coats 11-F, Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

claude.nadeau@statcan.ca

An agreement with the Canadian Customs & Revenue Agency allows Statistics Canada to use tax data coming from the incorporated and unincorporated businesses. In 2002, a project was launched with the objective to use more tax data in order to reduce the response burden of businesses, reduce the cost of survey programs and to improve the data quality. So, for reference year 2003, the Unified Enterprise Survey (UES) will use tax data as replacement data for about 50% of the surveyed simple-single establishments. This presentation will describe the issues and the difficulties at the estimation stage when using survey data and tax data. More specifically, modeling and adjustment factors to take into account the frame deficiencies resulting of the lost of information of survey data, will be discussed.

Response Models in RDD Surveys

◆ Stephen Williams, Mathematica Policy Research, Inc.; John W. Hall, Mathematica Policy Research, Inc.; Ronghua Lu, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., PO Box 2393, Princeton, NJ 08543

swilliams@mathematica-mpr.com

Key Words: *nonresponse, propensity modeling, RDD telephone surveys, Community Tracking Study*

For random digit dialing (RDD) household surveys, for example, the resolution of whether or not a number identifies a household is becoming more challenging. Likewise, obtaining household participation in a survey is becoming more difficult. This paper

investigates the use of logistic regression models to predict for each sampled RDD number in a recently completed household survey, the likelihood of resolving the number and the likelihood, if the number is a household telephone number, that the household will complete the screening questions, and finally that the household will complete the questionnaire. It builds on earlier analyses by the same authors. We apply these methods to compute weights for round three of the Community Tracking Study (CTS) Household Survey. The RDD numbers in this example are generated through the GENESYS system, which provides a substantial amount of auxiliary demographic information at the telephone-exchange level. Other information includes prior responses from the longitudinal panel for some numbers. We develop these models for a single stage national survey sample and also for samples in a number of the communities studied in the CTS.

Comparing Matchers to Enhance Upstream Capture of Duplicates on the MAF

◆ Mark K. Moran, U.S. Census Bureau

mark.k.moran@census.gov

Key Words: *record linkage, MAF, duplicate*

The Master Address File (MAF) inventories addresses nationwide for the Decennial Census and various surveys. Expanding coverage sometimes introduces duplicates—costly in the field. Best practice catches duplicates upstream. While we currently prevent duplication via a sophisticated "matcher," the tool has never been compared to a commercial probabilistic matcher. Using an administrative records file, we compared how both matchers linked addresses to the MAF. To account for handling of units at same address (e.g., apartments), three samples were drawn from discrepant cases and masked for clerical review: one lacking units, another from identical units, and a third from unlike units. Clerical reviewers classified links as confirmed, rejected, or unresolved. We found strong overall agreement between matchers (98% of confirmed links), while 1% were uniquely linked only by the production matcher and another 1% uniquely linked only by the probabilistic matcher. Similar results held in the unit absent universe, as well as in the equals universe. Unlike units, as expected, had smaller proportions of consensus.

Comparison of the American Community Survey Three-year Average Estimates with the Census 2000 Sample Estimates for 36 Counties

◆ Gregg J. Diffendal, U.S. Census Bureau; Andre L. Williams, U.S. Census Bureau; Rita J. Petroni, U.S. Census Bureau

U.S. Census Bureau, PRED RM. 1107-2, Washington, DC 20003

gregg.j.diffendal@census.gov

Key Words: *multiyear estimates, profile data tables, Census Long Form, survey comparisons*

The American Community Survey (ACS) plans call for it to replace the Census Long Form for the 2010 Census. This paper compares the ACS estimates at the county and tract levels for a group of 36 diverse counties with the corresponding Census sample estimates. The profile tables containing summary data on demographic, social, economic, and housing statistics are used to give a broad level of typical data for comparison. The ACS data is the cumulation of the data collected in 1999, 2000, and 2001 which represents the equivalent ACS data with five years of samples under full implementation. The Census data are the usual census sample estimates. Graphical techniques are used to demonstrate the similarities and differences between the ACS data and the Census data for this group of small areas. This is the first data using multiple-year estimates from the ACS that are directly compared to the Census long form estimates.

Minimum Change Edit and Imputation for the 2006 Canadian Census

◆ Darryl Janes, Statistics Canada; Michael Bankier, Statistics Canada

Statistics Canada, 15-A R.H. Coats Bldg., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

darryl.janes@statcan.ca

Key Words: *CANCEIS, decision logic tables, hot-deck imputation, deterministic imputation*

In the 2001 Canadian Census of Population, the Canadian Census Edit and Imputation System (CANCEIS) performed minimum change hot-deck imputation for nearly half of the census variables. An earlier version of this software was also used in the 2001 Brazilian and Swiss Censuses. In the 2006 Canadian Census, CANCEIS will perform hot-deck imputation as well as deterministic imputation on all census variables. New tools are being added to make CANCEIS easier to use such as the Decision Logic Table Editor and a Graphical User Interface. Improvements to the imputation methodology include additional distance measures to allow numeric edit and imputation for a wider range of variables.

A Hybrid Estimation Approach to State-level Estimates in the Survey of Industrial Research & Development

◆ John G. Slanta, U.S. Census Bureau; Jeri M. Mulrow, National Science Foundation

U.S. Census Bureau, Room 2225 Bldg. 4, Washington, DC 20233

john.gregory.slanta@census.gov

Key Words: *small-area estimates, synthetic estimates, model-based estimates*

The Census Bureau annually conducts the company-based Survey of Industrial Research and Development (R&D) in conjunction with the National Science Foundation (NSF). The survey provides national data on the amount of basic and applied research and development conducted within the United States. In addition, state-level estimates of industrial R&D are of great interest. It is estimated that between 2% and 5% of all U.S.-based companies perform R&D. A new national sample is selected each year and stability across years in the state estimates is a concern given the rare event nature of R&D. This paper explores the use of an estimator that is a hybrid between a design-based and a model-based estimate that also meets constraints required for data publication. This paper will present the estimation and variance estimation methodologies and show results from the survey.

80 Simulation

Section on Statistical Computing

Monday, August 9, 8:30 am–10:20 am

Identifying the Input Region of Complex Queuing Models for a Preset Performance Threshold—An Efficient Simulation Methodology

◆ Ying-Chao Hung, National Central University

National Central University, Graduate Institute of Statistics, Taiwan

hungy@stat.ncu.edu.tw

Key Words: *network simulation, input region, performance threshold, support vector regression, space-filling design*

The development of network simulations has been largely devoted to improving the network performance over the last decade. In order to meet service level commitments, a procedure called Network Capacity Planning is often used to help adequately select input parameters for the simulation. The shortcoming of this type of simulation methodology is that, for large systems, it is extremely expensive both in terms of CPU time and use of available resources (e.g., processors). We present a methodological framework for designing efficient simulations for complex networks. The goal is to (1) adequately identify an input region such that the performance measure of interest (such as delay, backlog, jitter, etc) does not exceed a preset threshold, and (2) minimize the required number of simulation runs (experiment trials). Our approach works in sequential and combines the methods of SV (Support Vector) regression and space filling designs. A generalized switch model is used to illustrate the proposed methodology.

A Method for Simulating Multivariate Non-normal Distributions with Specified Intraclass Correlations

◆ Todd C. Headrick, Southern Illinois University, Carbondale; Bruno Zumbo, University of British Columbia

Southern Illinois University, Carbondale, Section on Statistics, Dept. of EPSE, 222-J Wham Bldg., Mail Code 4618, SIUC, Carbondale, IL 62901-4618

headrick@siu.edu

Key Words: *intraclass correlation, Monte Carlo, multivariate data, non-normality, simulation*

Intraclass correlations are commonly used indices in the fields of biometrics, quality control, longitudinal data analysis, measurement theory, and survey research. There is a paucity of computationally efficient algorithms for simulating intraclass correlations under conditional non-normality. As such, a procedure is developed for simulating multivariate non-normal continuous distributions with specified (a) standardized moments, (b) intercorrelations, and (c) intraclass correlations. The theoretical models specified are associated with two-factor designs with either fixed or random effects. A numerical example will be worked to demonstrate the methodology. The results of a Monte Carlo simulation will also be provided to demonstrate that the proposed procedure generates the specified population parameters, intercorrelations, and intraclass correlations.

Performance of Likelihood-based Estimation Methods for Multilevel Binary Regression Models

◆ Marc Callens, Population and Family Research Centre; Christophe Croux, Katholieke Universiteit Leuven

Population and Family Research Centre, Centrum voor Bevolkings- en Gezinsstudie, Markiesstraat 1, Brussels, B-1000 Belgium

Marc.Callens@cbgs.be

Key Words: *binary regression, fractional factorial experiment, Monte Carlo simulation, multilevel analysis*

By means of a fractional factorial simulation experiment, we compare the performance of penalized quasi-likelihood, nonadaptive Gaussian quadrature and adaptive Gaussian quadrature in estimating parameters for multilevel logistic regression models. The comparison is done in terms of bias, mean squared error, numerical convergence, and computational efficiency. It turns out that, in terms of mean squared error, standard versions of the quadrature methods perform relatively poor in comparison with penalized quasi-likelihood.

Simulation Ranking and Selection Using Transient Data

◆ Huaiyu Ma, Rensselaer Polytechnic Institute; Thomas R. Willemain, Rensselaer Polytechnic Institute

Rensselaer Polytechnic Institute, Decision Sciences & Engineering Systems, 110 Eighth St., Troy, NY 12180-3590

mah@rpi.edu

Key Words: *bootstrap, simulation, ranking and selection, transient, sequential analysis*

One potentially valuable application of discrete event simulation is to rapid comparison of competing system designs. Speed is important when choosing among alternative ways to respond to disruptions in the operation of mission-critical systems. It is also important when sorting through the large number of alternatives generated by combinatorial design methods. Traditional statistical analyses that discard the initial transient (warm-up period) waste useful data due to a mistaken assumption that only steady state data have relevance to the ranking of alternatives. We describe techniques to exploit transient data, including the use of the bootstrap as a data multiplier, and compare the performance of transient-based methods to the best currently available sequential methods for ranking and selection. We focus on the simple special case of comparing two alternative designs. Performance comparisons are based on the number of simulation observations required to arrive at a reliable identification of the better alternative.

Inference and Filtering for Partially Observed Diffusion Processes via Sequential Monte Carlo

◆ Edward Ionides, University of Michigan

University of Michigan, Dept. of Statistics, 550 East University, Ann Arbor, MI 48109

ionides@umich.edu

Key Words: *discrete time sampling, inference for stochastic processes, nonlinear diffusion, particle filter, stochastic differential equation*

Stochastic models involving discrete time sampling of an underlying continuous time process, possibly with observation error, arise in many fields. When the underlying process is Markov and has continuous sample paths, the model is called a partially observed diffusion process. Questions of interest include reconstruction of the unobserved process and estimation of unknown model parameters. Sequential Monte Carlo (SMC), also known as the Particle Filter, is a technique which has been applied to various partially observed Markov models. This work investigates a new class of SMC algorithms which takes advantage of special properties of diffusion processes. Transition densities of nonlinear diffusions and conditional nonlinear diffusions are hard to calculate, however diffusion processes are easy to simulate from. Likelihood ratios are also easy to calculate. Questions are raised concerning the mixing properties of the "particles" for which the new algorithms differ from standard SMC, and the somewhat surprising ability of the new algorithms to function when the observation error is low or singular.

First Hitting Time Analysis of the Independence Metropolis Sampler

◆ Romeo Maciuca, University of California, Los Angeles; Song-Chun Zhu, University of California, Los Angeles

University of California, Los Angeles, 8130 Math Sciences Bldg., Box 951554, Los Angeles, CA 90095-1554

rmaciuca@stat.ucla.edu

Key Words: *fundamental matrix, Metropolized Gibbs Sampler, Independence Metropolis Sampler, eigenanalysis, expectation, variance*

The focus of the presentation will be a special case of the Metropolis algorithm, the Independence Metropolis Sampler (IMS), for finite state spaces. The IMS is often used in designing components of more complex Markov chain Monte Carlo algorithms. We present new results related to the first hitting time of individual states for the IMS. These results are expressed mostly in terms of the eigenvalues of the transition kernel. We derive a simple form formula for the mean first hitting time and we show tight lower and upper bounds on the mean first hitting time with the upper bound being the product of two factors: a "local" factor corresponding to the target state and a "global" factor, common to all the states, which is expressed in terms of the total variation distance between the target and the proposal probabilities. We also briefly discuss properties of the distribution of the first hitting time for the IMS and analyze its variance. We conclude by showing how nonindependence Metropolis-Hastings algorithms can perform better than the IMS.

Statistical Quasi-Newton: A New Look at Least Change

◆ Scott Vander Wiel, Bell Laboratories; Chuanhai Liu, Texas A&M University

Bell Laboratories, Rm 2C-2767, 700 Mountain Ave., Murray Hill, NJ 07023

scottyv@bell-labs.com

Key Words: *BFGS, DFP, negative Broyden family, Wishart model*

A new quasi-Newton method is derived by modeling curvature uncertainty with a Wishart distribution and by following a certain least change principle to determine a Hessian estimate that preserves accuracy across iterations. The new update is in the Broyden class but uses a negative parameter, outside the convex range usually regarded as the "safe-zone" for Broyden updates. Although Newton steps based on this update tend to be too long, optimal step sizes can be estimated from the Wishart model. In numerical comparisons to BFGS, the new algorithm converges with about 20% fewer iterations and gradient evaluations and about 10% fewer function evaluations on a suite of standard test functions. Our statistical framework provides a simple way to understand differences among various Broyden updates such as BFGS and DFP and shows that these methods do not preserve Hessian accuracy. In fact, BFGS, DFP and all other updates with non-negative Broyden parameters tend to inflate Hessian estimates. Numerical results on three new test functions validate these conclusions.

81 Modeling Strategies for Risk Assessment ▲

Section on Risk Analysis

Monday, August 9, 8:30 am–10:20 am

A Decision-tree-model Based on Roc Analysis Predicting Progression to Type 1 Diabetes Mellitus

◆ Shui Yu, University of Pittsburgh; Sati Mazumdar, University of Pittsburgh; Nancy B. Sussman, University of Pittsburgh; Vincent C. Arena, University of Pittsburgh; Dorothy J. Becker, University of Pittsburgh

University of Pittsburgh, Dept. of Biostatistics, Graduate School of Public Health, Pittsburgh, PA 15261

shuiyu@hotmail.com

Key Words: *diabetes, prediction, tree-model*

A novel decision tree analysis was used to select the factors that best predict the risk of Type 1 diabetes (T1DM) in first-degree relatives of T1DM probands. The uniqueness of this approach lies in the application of the receiver operating characteristic (ROC) curve analysis to identify the cutpoints for the predictors included in the model. This approach can be easily adjusted to give desired importance to sensitivity or specificity depending on the clinical relevance of the risk. The study was carried out in a cohort of 1484 first-degree relatives (FDRs) of T1DM probands followed in the greater Pittsburgh area for 22 years. We compared the model from this analysis with those from logistic regression and Cox model.

A Survey and Comparison of Statistical Approaches to Credit Decisions

◆ William G. Platt, San Francisco State University

San Francisco State University, 99 Stonegate Rd., Portola Valley, CA 94028

jplatt@sfsu.edu

Key Words: *credit risk, Markov chains, classification trees*

There are many methods to making credit decisions. There has been a recent flurry of activity relating to credit scoring systems, including the ge of discriminant models, classification trees, neural nets, genetic algorithms, and Markov chains. Commercial enterprises are reluctant to identify which models work best. This paper will present the alternative models currently used, and compare their performance.

A Nonlinear Pharmacokinetic-Weibull Model for Low-risk Dose Estimation for Carcinogenic Agents

◆ Abdul S. Al-Khalidi, Insurance Bureau of Canada

Insurance Bureau of Canada, 2235 Sheppard Ave. East, Attria II, Suite 1100, Toronto, ON M2J 5B5 Canada

salkhaldi@ibc.ca

Key Words: *low-risk dose, renewal Weibull process, tolerance, pharmacokinetic model*

This paper introduces a statistical model to estimate low risk doses of carcinogenic substances from data in which time to tumor onsets are recorded or estimated. The low-risk dose estimator is derived from a parametric probability distribution for the time to tumor onset to provide exposure duration effect, as well as dose levels. The derived probability distribution for the tumor onset time assumes: (1) the effective dose is a nonlinear pharmacokinetic transform of the exposed (or administered) dose; and (2) the probabilistic model underlying the waiting time of occurrence of a tumor is a Renewal Weibull Process. In the area of toxicology and environmental health, the damaging effects of environmental chemical (such as gaseous toxins) or other stimuli are often studied in animal and microbial systems. Data from such experiments will be analyzed via the model developed in this work. The literature on the topic is full with various models, using different assumptions. Most of these models consider the administered dose as effective dose, the derived models are built on this basis, and the low-risk doses are estimated using extrapolations.

The Partial Regression Approach to Fit GAMs in Air Pollution Research

◆ Shui He, University of Pittsburgh; Sati Mazumdar, University of Pittsburgh; Vincent C. Arena, University of Pittsburgh

University of Pittsburgh, 303 Parran Hall Dept. of Biostatistics, Graduate School of Public Health, Pittsburgh, PA 15261

shh10@pitt.edu

Key Words: *generalized additive model, partial residual, air pollution, time series*

Generalized Additive Models (GAMs) in S-plus, the most common statistical approach in time-series studies of air pollution and health, can seriously overestimate the fitted model parameters and underestimate their variances in the presence of concurvity. We present an alternate way to fit GAM to this type of data using partial residuals. Our simulation study indicates that this method gives better slope and corresponding variance estimates compared to the standard approach.

Model Uncertainty and Risk Estimation for Quantal Responses

◆ A. John Bailer, Miami University; Robert Noble, Miami University

Miami University, Dept. of Mathematics & Statistics/NIOSH, Oxford, OH 45056

baileraj@muohio.edu

Key Words: *unit cancer risk, Bayesian model averaging, quantal multistage models*

Observational human and experimental animal studies often serve as the basis for predicting risk of adverse responses in humans exposed to occupational hazards. A statistical model is applied to exposure-response data and this fitted model may be used to obtain estimates of the exposure associated with a specified level of adverse response. Unfortunately, a number of different statistical models are candidates for fitting the data and may result in wide ranging estimates of risk. Bayesian model averaging (BMA) offers a strategy for addressing uncertainty in the selection of statistical models when generating risk estimates. This strategy is illustrated in the context of applying the quantal multistage model. An animal carcinogenicity study dataset is used to illustrate how BMA provides excess risk estimates that reflect both sampling variability and model uncertainty.

A Five-color Scheme for "Directed Perception" with Maps

◆ Turkan K. Gardenier, Pragmatica Corporation

Pragmatica Corporation, 246 Maple Ave. East, Suite 206, Vienna, VA 22180

drgarden@erols.com

Key Words: *bivariate, geographical data, contingency table, uncertainty, color schemes*

A method will be presented to illustrate the need for incorporating uncertainty into bivariate representations of mapped data. The five-color scheme isolates data elements from the four quadrants of a contingency table and allows for a middle region where delineations are less certain. Comparisons among the quadrants employ consideration of data HL/LH (high-low and low-high) data elements to supplement inferences based upon data within quadrants of concurrence.

82 Longitudinal and Correlated Discrete Data Analysis

Biometrics Section

Monday, August 9, 8:30 am–10:20 am

Parameter-Driven Models for Time Series of Count Data

◆ Rachel M. Altman, University of Washington; Brian G. Leroux, University of Washington

University of Washington, Dept. of Biostatistics, Health Sciences Bldg., Stop 357232, Seattle, WA 98195

rachel@stat.ubc.ca

Key Words: *count data, generalized linear mixed models, latent variables, hidden Markov models*

Modelling correlated count data is a challenging problem. Unlike the situation for continuous data, for which the multivariate normal distribution is available, for count data there is no convenient and flexible class of multivariate distributions that can capture the shape of the distribution and the autocorrelation. Furthermore, techniques which have been developed for assessing the fit of models for normally distributed data do not extend readily to the count data setting. We propose a general class of parameter-driven (latent variable) models for count data. This class includes the generalized linear mixed model, hierarchical generalized linear model, and the hidden Markov model. We consider the interpretation of these models and discuss a parameter estimation method which yields estimates of the regression coefficients that are both efficient and robust to misspecification of the latent process. We apply these ideas to the analysis of multiple sclerosis and polio incidence data.

Analysis of Clustered Count Data with Excessive Zeros

◆ On Yee Tang, University of Hong Kong

University of Hong Kong, Dept. of Statistics and Actuarial Science, Pokfulam Rd., Hong Kong, China

toyee@hkusua.hku.hk

Key Words: *clustered data, count data, frailty, multiple imputation, Poisson regression, zero-inflated*

In many medical or public health investigations, the count data encountered often exhibit an excess of zeros, and very frequently this type of data are collected on clusters of subjects or items. Poisson regression model with frailty is adopted to analyze this type of clustered zero-inflated count data. The noncentral chi-square distribution with zero degrees of freedom is proposed to model the random effects, which not only account for the subject specific heterogeneity, but also the dependency among subjects within a cluster. The use of this special distribution can provide more flexibility on the relationship between the covariates and the random effects. A simple multiple imputation approach is proposed for the parameter estimation of this model. The proposed methodology are motivated and illustrated by the public health survey conducted in Indonesia, which employs a multilevel cluster sampling scheme on the number of days of missing primary activities due to illness in a four-week period and the number of days in beds due to illness in a four-week period.

Transition Models for Multivariate Longitudinal Binary Data

◆ Leilei Zeng, University of Waterloo; Richard J. Cook, University of Waterloo

University of Waterloo, Dept. of Statistics and Actuarial Science, 200 University Ave. West, Waterloo, ON N2L 3G1 Canada

lzeng@uwaterloo.ca

Key Words: *association parameters, longitudinal data, estimating functions, Markov model, multivariate process, transition probability*

In many settings with longitudinal binary data interest lies in modeling covariate effects on transition probabilities. When interest lies in tracking how two processes change together, one may examine the degree to which changes in one process are correlated with changes in another process. In such settings, the associations between the transition occurrences for the two processes are the scientific focus. Under Markov assumptions, use of marginal transition models permits separate modeling of covariate effects on the transition probabilities for univariate longitudinal binary data, but no insight into the associations can be obtained. While time-dependent covariates may be constructed for one process based on the other, the two processes are then treated asymmetrically. We propose a method of estimation and inference for joint transitional models for bivariate longitudinal binary data based on GEE2 or alternating logistic regression. This approach enables one to model covariate effects on marginal transition probabilities as well as on the association parameters between the two processes.

A Comparison of Goodness-of-fit Tests for the Logistic GEE Model C

◆ Scott R. Evans, Harvard University; Lingling Li, Harvard University

Harvard University, FXB-513/HSPH, 651 Huntington Ave., Boston, MA 02115

evans@sdac.harvard.edu

Key Words: *goodness-of-fit, GEE, clustered binary data, logistic regression*

The logistic Generalized Estimating Equations (GEE) model has become a widely used method of analysis for clustered binary data. The model is easily fit with readily available software. Output is easily interpreted and can be used to estimate probabilities and/or odds ratios. With this increase in application has been an increase in the development of methods to assess the adequacy of the fitted model. Assessment of the adequacy of the fitted GEE models has been a challenging area of research since no likelihood exists and residuals within cluster are correlated. Recently several goodness-of-fit statistics have been proposed; however, a comparison and evaluation of these methods is lacking. We review, compare, and evaluate these methods.

Hypothesis Testing with Clustered Binary Pairs

◆ Chaya S. Moskowitz, Memorial Sloan-Kettering Cancer Center; Mithat Gonen, Memorial Sloan-Kettering Cancer Center; E.S. Venkatraman, Memorial Sloan-Kettering Cancer Center

Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology and Biostatistics, 307 E 63rd St, 3rd Floor, New York, NY 10021

moskowc1@mskcc.org

Key Words: *design effect, marginal homogeneity, McNemar's test, sensitivity, symmetry*

Clustered binary pairs arise naturally in many applications. For example, in the context of comparing two diagnostic tests for detection of lesions, when both tests are applied to individuals who have multiple lesions the test results form clustered binary pairs. Focusing on hypothesis testing, we propose a class of weighted test statistics and derive optimal weights that minimize the variance among this class of test statistics. We show that two existing methods, one based on the design effect and one based on the generalized score test from GEE are equal to each other and members of this class. Another method based on the intracluster correlation, while not a member of this class, is asymptotically equivalent in certain situations. We present results of simulation studies comparing the properties of the test statistics.

Comparison of Variance Estimation Approaches in a Two-state Markov Model for Longitudinal Data with Misclassification

◆ Rhonda J. Rosychuk, University of Alberta; Xiaoming Sheng, University of Utah School of Medicine

University of Alberta, Dept. of Pediatrics, 9423 Aberhart Centre, Edmonton, AB T6G 2J3 Canada

rhonda.rosychuk@ualberta.ca

Key Words: *jackknife, bootstrap, variance, Markov process, misclassification*

We examine the behavior of the variance-covariance parameter estimates in an alternating binary Markov model with misclassification. Transition probabilities specify the state transitions for a process that is not directly observable. The state of an observable process, which may not correctly classify the state of the unobservable process, is obtained at discrete time points. Misclassification probabilities capture the two types of classification errors. Variance components of the estimated transition parameters are calculated with three estimation procedures: observed Fisher information, jackknife and Monte Carlo bootstrap techniques. Simulation studies are used to compare variance estimates and reveal the effect of misclassification on transition parameter estimation. The three approaches generally provide similar variance estimates.

Bayesian Mixture Models for Complex High-dimensional Count Data

◆ Yuan Ji, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd. - Unit 447, Houston, TX 77030

yuanji@mdanderson.org

Key Words: *MCMC, Gibbs sampler, Metropololis-Hastings, Bayesian inference*

Phage display is a very useful method to study the behavior of a very large number of peptides and proteins on the surface of a small bacterial virus, called a phage. The resulting count data from phage display experiments usually possess high dimensionality and a complex correlated structure. Statistical modeling of these data are therefore challenging. The main issues involve multiple comparisons and, probably more importantly, modeling the complex correlation structure in the data, which is of major interest. We develop a class of Bayesian mixture models for such complex high-dimensional count data and propose a selection methodology for identifying peptides with distinct ascending display patterns in their counts. We construct Bayesian hierarchical priors for the parameters that are specifically designed for this type of data. Our simulation results indicate that the proposed mixture models and priors are very suitable for the count data. We present a case study in detail to demonstrate the proposed methodology.

83 Longitudinal and Hierarchical Modeling for Health Policy ▲ ⌘

Section on Health Policy Statistics, Social Statistics Section, Section on Survey Research Methods

Monday, August 9, 8:30 am–10:20 am

Single Level Mediation: Estimation and Sampling Distribution of the Mediated Effect when X is Dichotomous and Y are Continuous

◆ Oumar S. Sy, Kansas State University

Kansas State University, 1603 Denison Ave. # 23, Manhattan, KS 66502

osy@ksu.edu

Key Words: *mediation, estimation*

Researchers commonly ask whether relationships between predictors X, and outcomes Y, are mediated by a third set of variables, M. Although the evaluation of the effect of X on Y that is mediated by M has been the focus of much research in social science, formal significance tests of the mediated effect are rarely conducted. The purpose of this article is to develop statistical procedures that provide point estimate, standard error estimate, and exact sampling distribution of the mediated effect when X is dichotomous and Y and M are both assumed to be continuous. This is an extremely common case since in applied research in the social and behavioral sciences where prevention programs (X) have only two components: control (X=0) and intervention (X=1).

Emergency Department Visits Among COPD Patients: An Example of Repeat Single and Recurrent-event Analysis

◆ Jianglin Xu, Northwestern University; Todd Lee, Northwestern University; Whitney Witt, Northwestern University; Ramon A. Durazo-Arvizu, Northwestern University

Northwestern University, Chicago, IL 60612-4209

j-xu3@northwestern.edu

Key Words: *COPD, ED, VA, HR*

There is increasing interest in applying survival analysis techniques to datasets with recurrent events in health services research. The objective of this analysis was to evaluate the association between follow-up office visits and repeat ED visits using single and recurrent event methods. A cohort of VA patients with a COPD-related ED visit was identified (n=27,416). Cox proportional hazard and Anderson-Gill models were used to compare the risk for repeat ED visits within 90 days of the index ED visit between patients with and without a follow-up physician visit within 30 days. Models were adjusted for age, gender, and health care utilization. When only the first repeat ED visit was considered (3,845 ED visits), patients with a follow-up office visit were at higher risk (HR=1.53; 95% CI, 1.42-1.63). When all repeat ED visits were considered (5,364 ED visits), the HR was 1.49 (95% CI, 1.40-1.58). The recurrent event model used more information; however, results were similar between models. Among COPD patients with an ED visit, those with a follow-up office visit within 30 days were more likely to have a repeat ED visit within 90 days.

A Markov Chain Model of Functional Decline in Activities of Daily Living Among Nursing Home Patients Using MDS Quarterly Assessment Data

◆ Richard C. Gardiner, New York Association of Homes and Services for the Aging; Colene Byrne, New York Association of Homes and Services for the Aging; Hailing Li, New York Association of Homes and Services for the Aging

New York Association of Homes and Services for the Aging, 150 State St., Suite 301, Albany, NY 12207-1698

dgardiner@nyahsa.org

Key Words: *Markov chain, nursing homes, functional decline, MDS data*

A Markov chain is used to model decline in functional independence of residents in Nursing Homes. The model transition matrix is adjusted for risks that uniquely effect individual resident transition probabilities. Multinomial logistic regression is used to construct the risk adjusted transition matrices. Five stages of functional independence were identified using 11 MDS assessment variables measuring activities of daily living. MDS assessment records for 1.1 million New York State nursing home residents for the period April 1, 2000, to December 31, 2002, were used to derive the normative quarterly transition probabilities between the five stages of functional independence. Separate models are needed to accurately reflect functional decline in residents afflicted with Alzheimer's disease or dementia and all other residents. Model outcome vectors very closely replicate actual outcomes even after the passage of 10 quarterly transition intervals. The models have application in a number of health policy areas including, monitoring nursing home quality of care, program evaluation of nursing care strategies to deter functional decline, and simulation of current nursing home practice.

Optimal Designs for Bayesian Hospital Report Cards

◆ Peter C. Austin, Institute for Clinical Evaluative Sciences

Institute for Clinical Evaluative Sciences, G1 06, 2075 Bayview Ave., Toronto, ON M4N 3M5 Canada

peter.austin@ices.on.ca

Key Words: *hospital report cards, Bayesian models, provider profiling, decision theory, quality of care*

There is an increasing interest in measuring the quality of medical care that patients receive. Report cards publishing hospital-specific mortality rates for specific conditions or procedures are increasingly common. There is growing interest in the use of Bayesian methods in report cards. These allow the determination of the probability that the mortality rate at a specific hospital exceeds a specific threshold. The choice of probability level required to classify a hospital as having higher than acceptable mortality has not been justified. In profiling, misclassification can occur: hospitals that truly have acceptable mortality rates can be classified as having higher than acceptable mortality rates, while hospitals that truly have unacceptably high mortality can be classified as having acceptable mortality. We place the design of hospital report cards in a Bayesian decision-theoretic framework, by introducing loss functions to explicitly describe costs incurred by patients and providers when hospitals are misclassified. We then use Monte Carlo simulations to explicitly determine optimal probability levels that minimize the mean posterior cost of misclassification.

A Dirichlet Multinomial Logistic Regression Model: Physician Assistant Autonomy

◆ Jeffrey R. Wilson, Arizona State University; Eugene Schneller, Arizona State University; Baohui Zhang, Arizona State University

Arizona State University, W.P. Carey School of Business, School of Health Administration and Policy, Tempe, AZ 85287-4506

jeffrey.wilson@asu.edu

Key Words: *nested, dependency, extra variation, clustering*

A Dirichlet-Multinomial logistic regression model is used to model data pertaining to physician assistants' opinions over time. The data is part of a larger study where researchers were interested in the beliefs of physician assistant about autonomy over time. The physicians were surveyed in 1972, 1977, and 1985 to see how their views have changed. The Dirichlet logistic regression model is used since two sources of variation were inherent in the data beyond what is expected in a multinomial model were found to be present.

Random Effect Models for Medical Costs in the Presence of Death

◆ Lei Liu, University of Michigan; Robert A. Wolfe, University of Michigan; Jack Kalbfleisch, University of Michigan

University of Michigan, 1420 Washington Heights, Dept. of Biostatistics, Ann Arbor, MI 48109-2029

liulei@umich.edu

Key Words: *proportional hazard model, informative censoring, mixed model, changepoint, survival analysis*

Statistical analysis of medical cost data has prompted increasing interest recently. In this paper we focus on medical costs recorded in certain time intervals, e.g. every month, which can be treated as repeated measures in the presence of a terminating event like death. We propose a joint random effect model for medical cost and survival time. In particular we incorporate the fact that there is a spike of monthly medical cost from some fixed time (say three months) before death. We propose a "turn back time" scheme to move back the censoring time for medical cost by three months while keep that for survival unchanged. Maximum likelihood estimation and inference are carried out through Monte Carlo EM algorithm with Metropolis-Hastings sampler in the E-step. An analysis of medical cost data for dialysis patients is presented to illustrate the proposed methods.

84 Statistical Methods in Finance

▲ ⌘

Business and Economics Statistics Section

Monday, August 9, 8:30 am–10:20 am

Inference for Stochastic Volatility Models with Semi-Markov Regime Switching

◆ Zhaohui Liu, University of Connecticut; Nalini Ravishanker, University of Connecticut

University of Connecticut, 1 Northwood Rd., Apt. 85, Storrs, CT 06268

Zhaohui.Liu@UConn.edu

Key Words: *stochastic volatility, Bayesian inference, financial time series, hidden semi-Markov process*

We discuss stochastic volatility models with hidden semi-Markov regime switches. In particular, the models analyzed include the univariate stochastic volatility model for financial return time series, and a bivariate model for the return and transaction volumes. In addition to the AR structure of the volatility process, it is assumed that the unobserved volatility follows a semi-Markov regime switching process, with the duration time at each state assuming different distributions. Under this framework, the mean and variance as well as duration times of the volatility at different states could vary. These assumptions afford a richer model which is expected to better describe the underlying volatility process due to its influence from different economic forces. Statistical inference for such models using the sampling based Bayesian approach will be described. Model selection and prediction will be described and illustrated.

MCMC Estimation of Multiscale Stochastic Volatility Models

◆ German Molina, Credit Suisse-First Boston; Chuan-Hsiang Han, University of Minnesota; Jean-Pierre Fouque, North Carolina State University

Credit Suisse-First Boston, Flat 77, Ionian Bldg., 45 Narrow St., London, E14 8DW UK

german@alumni.duke.edu

Key Words: *MCMC, Stochastic Volatility Model, time scales, foreign exchange*

We propose to use Monte Carlo Markov Chain methods to estimate the parameters of Stochastic Volatility Models with several factors varying at different time scales. This approach, in contrast with classical one-factor models, identifies and estimates well-separated time scales. This is tested with simulated data as well as foreign exchange data.

MCMC Estimation of Multifactor Affine Term-Structure Models

◆ He Hu, University of California, Los Angeles

University of California, Los Angeles, 8130 Math Sciences Bldg., Box 951554, Los Angeles, CA 90095

huhe@stat.ucla.edu

Key Words: *MCMC, term-structure models*

A number of methods have been proposed in the past years to study term-strcture models. It includes maximum likelihood and various moment-based estimations. Markov chain Monte Carlo (MCMC) methods have been applied only in recent years and have proven dramatically more efficient than other methods in high-dimensional models. Continuous time process is discretized via the Euler approximation. In order to utilize information of an arbitrary number of zero-coupon bonds, measurement errors are incorporated to the observation equation. Metropolis within Gibbs algorithm is used as the general framework. The simulation of state variable vectors greatly increases the bias of parameter estimates, which also rise with number of state variable factors. With the high computation capacity of MCMC sampler, this approach is also applicable to affine term-structure models with correlated state factors.

Statistical Estimation Error and Risk Aversion in Portfolio Formation

◆ Andrew F. Siegel, University of Washington

University of Washington, Dept. of Management Science, Box 353200, Seattle, WA 98195-3200

asiegel@u.washington.edu

The presence of statistical estimation error in the parameters of the asset returns distribution leads to loss of von Neumann Morgenstern utility in the portfolio formation problem with n assets because the optimal portfolio cannot be constructed using available information. This economic loss may be interpreted as the appropriate statistical loss function that should be used when choosing a statistical estimator to be used in conjunction with portfolio formation. Asymptotic properties of this approach are studied using the delta method in the case of quadratic utility.

Modeling Market Returns with a Speculative Bubble

◆ Leonard C. MacLean, Dalhousie University

Dalhousie University, 6152 Coburg Rd., Halifax, NS B3H 3J5 Canada

lmaclean@mgmt.dal.ca

Key Words: *securities, bubbles, shocks, returns*

We define a model for market returns, which, under certain conditions, determines the emergence of a speculative bubble in the economy, and drives more generally bond and equity returns. The dynamic equations in the model contain a shock term, whose size and intensity depend on yield differentials. A methodology for fitting parameters and shocks to actual returns is presented.

Fitting the shocks requires solving an integer programming problem. The methods are applied to data on U.S. bonds and securities.

On Examining Asymmetric Behavior of Price Limit Moves

◆ Ping-Hung Hsieh, Oregon State University; J. Jimmy Yang, Oregon State University

Oregon State University, 200 Bexell, College of Business, Corvallis, OR 97331-8527

hsiehph@bus.oregonstate.edu

Key Words: *delayed price discovery, MCMC, spillover effect, tail index*

The development of statistical model in modeling asset returns with price limits has progressed to capture the dynamic price generating structure as well as the leftover effect that occurs when the equilibrium price is censored. In addition, recent empirical studies have observed asymmetric behavior between upper and lower price limits. That is, the price process hits one limit more often than the other. This observation has motivated us to develop a model that accounts for this additional feature. We consider a stochastic volatiity process that drives the underlying equilibrium returns and tailor the tails of the conditional distribution to allow asymmetric tail behavior. The model provides a measure that quantifies the asymmetric effect of price limits. Its financial implications such as the volatility spillover and delayed price discovery are addressed. A simulation study is conducted to gain insight into the performance of the proposed model with respect to several existing models. The proposed model is demonstrated with an application to stock market data.

Scenario-based Conditional Nonlinear Hedge Fund Model

◆ Bernard Lee, Imperial College, London; Youngju Lee, Allianz Hedge Fund Partners

Imperial College, London, South Kensington Campus, London, SW72AZ UK

bernardl@alumni.princeton.edu

Key Words: *hedge fund, nonlinear conditional model, hedge fund-style allocation, bootstrap*

Hedge fund strategies typically generate option-like returns. Linear-factor models using benchmark asset indices have difficulty explaining them. There were several attempts to explain hedge fund returns using traditional asset classes with nonlinear characteristics. Even though a hedge fund with a certain type of style will focus on the movement of equity markets, bond markets, currency markets, or commodity markets, we believe sophisticated hedge fund managers will consider the interactions between those traditional asset classes and those markets that interact with each other. Our research tries to model this traditional market interaction with consideration of nonlinear characteristics of hedge fund returns using conditional factor model which developed based on the previous work of Markowitz and Perold (1981). This model generates several different scenarios depending on market situation. Hedge funds have a short history and lack of data is one of most challenging parts of hedge fund research. We introduced a bootstrap method to test the hedge fund model in this research.

85 Statistical Models for Air Pollution and Health Effects

Section on Statistics and the Environment

Monday, August 9, 8:30 am–10:20 am

Overlap Bias in the Case-crossover Design, with Application to Air Pollution Exposures (ENVR Student Paper Award Winner)

◆ Holly Janes, University of Washington

University of Washington, Dept. of Biostatistics, F-600 Health Sciences, Campus Mail Stop 357232, Seattle, WA 98195

hjanes@u.washington.edu

Key Words: *overlap bias, case-crossover design, estimating equations, air pollution*

The case-crossover design uses cases only, and compares exposures just prior to the event times to exposures at comparable control, or "referent" times, to assess the effect of short-term exposure on the risk of a rare event. It has commonly been used to study the effect of air pollution on the risk of various adverse health events. Careful referent selection is important to control for time-varying confounders, and in order to ensure that the distribution of exposure is constant across referent times, a key assumption of this method. Yet the referent strategy is important for a more basic reason: the conditional logistic regression estimating equations commonly used are biased when referents are not chosen a priori and are functions of the observed event times. We call this bias in the estimating equations overlap bias. We propose a new taxonomy of referent selection strategies in order to emphasize their statistical properties. We give a derivation of overlap bias, explore its magnitude, and consider how the bias depends on properties of the exposure series. We conclude that the bias is usually small, though highly unpredictable, and easily avoided.

A Bayesian Hierarchical Approach for Relating PM2.5 Exposure to Cardiovascular Mortality in North Carolina

◆ Christopher H. Holloman, Battelle; Steven M. Bortnick, Battelle; Michele Morara, Battelle; Warren J. Strauss, Battelle Memorial Institute; Catherine A. Calder, The Ohio State University

Battelle, 505 King Ave., Columbus, OH 43201

hollomanc@battelle.org

Key Words: *fine particulate matter, spatial modeling, SHEDS-PM, exposure simulator*

Over the past several years, studies have established that elevated PM2.5 levels and negative health effects are related; however, considerable uncertainty remains regarding the nature of the association. Since the EPA began widespread monitoring of PM2.5 levels in 1999, the epidemiological community has performed numerous observational studies modeling mortality and morbidity responses to PM2.5 levels using Poisson Generalized Additive Models (GAMs). While these models are useful for relating ambient PM2.5 levels to mortality, they give little information about the strength of the effect between exposure to PM2.5 and mortality. In order to address this question, we propose a three-stage Bayesian hierarchical model as an alternative to the classical Poisson GAM. Fitting our model in seven North Carolina counties using data covering the years 1999-2001, we find that an increase in PM2.5 exposure is linked to increased risk of cardiovascular mortality in the same day and following two days. In addition, we compare the results obtained from our model to those obtained by applying Frequentist and Bayesian versions of the classical Poisson GAM to our study population.

Threshold Effects in Air Pollution Mortality Associations

◆ Lise A. Tole, Leicester University; Gary M. Koop, Leicester University

Leicester University, Dept. of Economics, University Rd., Leicester, LE7 1RH UK

lat7@le.ac.uk

Key Words: *nonlinear time series, Markov chain Monte Carlo, dose-response relationship, Bayesian model averaging*

Important policy decisions hinge on whether air pollution has significant effects on human health. A topic of particular interest is whether there are thresholds below which pollutants are harmless. Numerous studies have used daily time series data on a pollutant (often particulate matter) and a health outcome (often mortality) to investigate possible thresholds in the dose-response relationship. The statistical methods used in previous studies typically use splines or non- or semiparametric methods such as the general additive model. We argue that such approaches may not be suitable when a myriad of possible explanatory variables (and lags) could trigger threshold effects. We introduce a flexible parametric model that allows for the possibility that threshold effects could occur in one (or more) of many possible ways. Multiple potential thresholds are defined according to levels and growth rates of a pollutant (or a weather variable or an interaction between a weather variable and a pollutant) on the current day or the recent past (or averaged over the recent past). We develop Bayesian methods for statistical inference and apply our methods to an extensive dataset.

Spatial-temporal Hierarchical Poisson Model to Study Association between Speciated Fine Particles and Human Health Effects from Pollution

◆ Hae-Ryoung Song, North Carolina State University; Montserrat Fuentes, North Carolina State University

North Carolina State University, Dept. of Statistics, 2501 Founders Dr., Campus Box 8203, 220 Patterson Hall, Raleigh, NC 27695-8203

hsong3@stat.ncsu.edu

Key Words: *mortality, generalized Poisson regression model, Bayesian hierarchical model, space-time series*

Many epidemiologic studies have shown the evidence for statistically significant effects of particular matter (PM) on mortality across many areas of the country. The main objective of our research is to quantify uncertainties about the impacts of fine PM exposure on mortality. We develop a multivariate spatial regression model for better estimation of the mortality effects from fine PM and its components across the coterminous U.S. We consider a flexible Bayesian hierarchical model for a space-time series of (mortality) counts by constructing a likelihood based version of a generalized Poisson regression model. We apply these methods to daily mortality county counts, measurements of total and several components of fine PM from national monitoring networks in the U.S.

Space-time Modeling of Daily Sulfate Levels Combining Observations and CMAQ Output

◆ Mikyoung Jun, University of Chicago; Michael L. Stein, University of Chicago

University of Chicago, Statistics, 5734 S. University Ave., Chicago, IL 60637

jun@galton.uchicago.edu

Key Words: *space-time covariance function, air quality model, model evaluation, space-time process, separable covariance function*

New statistical procedures to evaluate the Models-3/Community Multiscale Air Quality (CMAQ), a numerical, deterministic model which produces multiple air pollutants' concentration and deposition levels over U.S. and some parts of Canada, using observed data are introduced. Certain space-time correlations are used to assess dynamic aspects of CMAQ and to compare the space-time structure of CMAQ to that of observations. We show that developing statistical models of the difference of CMAQ output and observations should be easier than developing models for the observations or CMAQ output directly, which has direct implications on strategies for developing maps of sulfate concentrations using observations and model output. Statistical evidence against separable covariance functions, which are often used for space-time or multivariate spatial processes, are given. Space-time modeling results for the difference of CMAQ output and observations based on various space-time covariance functions will be compared. The development of space-time covariance function on a sphere across time will be discussed.

A Simple Statistical Air-dispersion Modeling with Spatially Correlated Data in Field Experiment

◆ Tae-Young Heo, North Carolina State University; Jacqueline M. Hughes-Oliver, North Carolina State University

North Carolina State University, Dept. of Statistics, 2501 Founders Dr., Campus Box 8203, Raleigh, NC 27695-8203

theo@stat.ncsu.edu

Key Words: *air-dispersion model, uncertanity, field experiment, clustered conditional autoregressive model*

We suggest a method for combining air-dispersion models with statistical models in order to test, evaluate, and reduce uncertainty due to point source effects. In environmental science, there are a variety of computational air-dispersion models. These models usually only provide deterministic predictions or estimates without measures of prediction or estimation uncertainty. By introducing error components in air-dispersion models, we obtain improved prediction of concentrations, we reduce prediction uncertainty, and we are able to statistically test the impact of a point source. One of our goals is to incorporate the information from deterministic air-dispersion models into the statistical modeling of these processes in order to validate and improve the deterministic models and to

provide better predictions by adding spatial process. Application is made to three real datasets on experimental trace fields, namely, the Kincaid, Indianapolis, and Prarie Grass projects.

A Case Study of the Implications of Atmospheric Sonde Data Pre-processing Using "Correction Factors"

◆ Wendy Meiring, University of California, Santa Barbara

University of California, Santa Barbara, Statistics and Applied Probability, Santa Barbara, CA 93105

meiring@pstat.ucsb.edu

Key Words: *environmental, data quality, ozone, atmospheric*

Atmospheric processes frequently are measured by several instruments, corresponding to a variety of space-time resolutions and sampling schemes. Types of measurement instruments include surface-based, satellite, balloon-based, and airplane-based instruments. Each instrument exhibits its own measurement error processes, with biases that may depend on atmospheric conditions. Atmospheric observations frequently are pre-processed prior to further analysis. For some atmospheric measurements, this pre-processing may involve "correcting" the data from certain instruments to improve agreement with observations from other measuring instruments. In the context of a stratospheric ozone study, we illustrate some of the resulting statistical challenges. We discuss the vital role that analyses of "correction factors" may play in studying data quality, concentrating on a case study of balloon-based ozonesonde data which have been "corrected" based on other instruments. Data pre-processing of this type also has implications for attempts to combine data from different measurement instruments.

86 Testing and Simulation Methods ⌘

General Methodology

Monday, August 9, 8:30 am–10:20 am

Conditional and Unconditional Confidence Intervals Following a Group Sequential Test

◆ Xiaoyin Fan, Merck & Co., Inc.; David L. DeMets, University of Wisconsin, Madison

Merck & Co., Inc., 785 Jolly Rd. Bldg. C, UN A-102, Blue Bell, PA 19422

xiaoyin_fan@merck.com

Key Words: *group sequential, confidence interval, conditional coverage, stopping time*

The naïve confidence interval (CI) of the parameter following a group sequential study is biased in terms of the coverage probability. There are several available exact CI methods based on different ways of ordering the two-dimensional test statistics sample space. While those exact CI's maintain the overall coverage probability, the conditional coverage probability at any given stopping stage could be well below the target level. A conditional CI (CCI) method can provide correct conditional coverage using the conditional likelihood inference at the stopping stage. However, the pure CCI method has serious drawbacks such as an unreasonably wide interval and inconsistency with the main hypothesis testing result. A two-step restricted conditional CI (RCCI) is proposed which takes advantage of both the stopping time and the test value at that point. Compared with available methods, numerical results show that the RCCI not only improves the conditional coverage considerably from the exact CI's, but also avoids the major undesirable properties of the CCI. The repeated confidence interval (RCI) method is compared as well but also found less appealing than the RCCI.

Modified Simes' Critical Values under Dependence

◆ Gengqian Cai, Temple University; Sanat K. Sarkar, Temple University

Temple University, 7605 B Massey Way, Elkins Park, PA 19027

gcai@temple.edu

Key Words: *Simes' test, step-up procedure, MTP2, false discovery rate*

A modification of the critical values of Simes' test is suggested when the underlying test statistics are multivariate normal with a common nonnegative correlation, yielding a more powerful test than the original Simes' test. A step-up multiple testing rocedure with these modified critical values provides an FDR-controlling procedure, and is more powerful than the traditional Benjamini-Hochberg (BH) procedure in terms of number of rejections. Simulations were carried out to compare this modified BH procedure with the BH and other modified BH procedures in terms of False Nondiscovery Rate (FNR) and average power. The present modified BH procedure is observed to perform well compared to others when the test statistics are highly correlated and most of the hypotheses are true.

Hypothesis Testing of Functional Data with Applications to Longitudinal Studies

◆ Kun Nie, BayesSoft, Inc.; Xiaowei Yang, BayesSoft, Inc.; Gang Li, University of California, Los Angeles; Steven Shoptaw, BayesSoft, Inc.

BayesSoft, Inc., 3340 Sawtelle Blvd. #310, Los Angeles, CA 90066

knie@ucla.edu

Key Words: *longitudinal study, functional linear regression, adaptive Neyman test, thresholding test, Fourier transform, wavelets transform*

Longitudinal studies in certain fields contain large number of repeated measures on each subject. The high dimensionality and temporal correlations of data erode the power of traditional testing methods seriously. Within the frame of functional linear regression models where the observations of the same subject are viewed as a sample from a functional space, we develop a new strategy for hypothesis testing problems. First, we perform the Fourier/wavelet transforms to each set of observation, which serve to reduce temporal correlations and compress useful signals into a few components; second, linear regressions are applied in the Fourier/wavelet domain; then the significance of the regression coefficients are tested based on the adaptive Neyman or thresholding test statistics. The simulation results showed that both the

adaptive Neyman and thresholding approaches significantly enhance sensitivities over traditional methods such as pointwise regression with Bonferonni corrections or univariate analysis where each set of observations is collapsed into a composite score. The methods are illustrated with an example of a smoking cessation study.

The Application of Fractional Polynomials and Support Vector Machines in Transportation Analysis

◆ Elke A. Moons, Limburgs Universitair Centrum; Marc Aerts, Limburgs Universitair Centrum; Geert Wets, Limburg University Centre

Limburgs Universitair Centrum, Universitaire Campus, Bldg. D, Diepenbeek, 3590 Belgium

elke.moons@luc.ac.be

Key Words: *fractional polynomial models, support vector machines, logistic regression, goodness of fit*

Transportation analysis has gone through some major changes over the last decade. Activity-based models became more and more important when modeling travel demand. Until now, these activity-based models have mostly been linear in nature. Over the last decennia, though, flexible linear and nonlinear statistical models became popular in other fields such as medicine (pharmakokinetics, survival analysis, etc.) while also in the field of machine learning, some nonlinear models have become standard; e.g., classification and regression trees, support vector machines. The idea of this paper is to investigate what flexible linear and nonlinear models, such as fractional polynomial models and support vector machines can add in a logistic regression setting in transportation analysis. Do they perform worse than the widely used logit model or do they yield better results? Are the parameters as interpretable as in linear regression models? We will apply these models to an important aspect of the activity-based process, i.e., to the selection of the transport mode fo work-related trips. A comparison in goodness of fit among models will be illustrated on real-world activity-diary data.

Markov Models for Financial Time Series

◆ Stanley L. Sclove, University of Illinois, Chicago

University of Illinois, Chicago, Information & Decision Sciences Dept. (MC 294), 601 S. Morgan St., Chicago, IL 60607-7124

slsclove@uic.edu

Key Words: *stock prices, EM algorithm, rates of return, hidden Markov model*

In the classical geometric Brownian motion model for stock prices the difference of the logarithm of the price series (the continuous rate of return, ROR) is a sequence of independent identically distributed normal random variables. Problems in applying this model are (1) successive values of ROR may not be independent, and (2) the ROR may not be normally distributed. The Markov model (MM) used here postulates states and transitions between them. This MM can model dependence, and the states can have different distributions, e.g., different normal distributions. Both visible-state and hidden-state MMs are treated. In the hidden-state model the EM algorithm estimates both the states and the distributional parameters. A run of one state can be considered a segment. Segmenting macroeconomic series by MMs can provide objective definitions of phases of the economy: recession, recovery, and expansion. Segmenting stock indices can provide objective definitions of bull and bear markets. Fitting MMs to stock prices can provide alternative scenarios for portfolio optimization, trading strategy, and options pricing.

Simulation as an Alternative to the Delta Method

◆ Patrick R. Johnston, Abt Associates Inc.

Abt Associates Inc., 55 Wheeler St., Cambridge, MA 02138

patrick_johnston@abtassoc.com

Key Words: *delta method*

This paper addresses the situation where the data comprise a collection of parameter estimates along with their standard errors. The goal of the analysis is to obtain a confidence interval for a function of these parameters. Problems of this nature are commonly solved via the delta method, whereby the function of interest is approximated by a linear function and a confidence interval is obtained for the latter. The appropriateness of the resulting confidence interval depends on the appropriateness of the linear approximation. We propose a simple alternative based on simulating the density of the function and obtaining a Bayesian probability interval which may then be interpreted in a frequentist fashion. The resulting interval estimate is attractive from frequentist, likelihood, and Bayesian perspectives. The method was applied to a problem involving the estimation of cocaine production from South America.

The Refined Positive Definite and Unimodal Regions for the Gram-Charlier and Edgeworth Series Expansions

◆ Fred Spiring, University of Manitoba; Bartholomew P.K. Leung, Hong Kong Polytechnic University; Anthony Yeung, University of Manitoba

University of Manitoba, Dept. of Statistics, 941 Kilkenny Dr., Winnipeg, MB R3T 3Z4 Canada

spiring@cc.umanitoba.ca

Key Words: *series expansion, positive definite, Sturm's Theorem, Gram-Charlier, Edgeworth*

Gram-Charlier and Edgeworth Series Expansions are used in the field of statistics to approximate probability density functions. The expansions have proven useful, but have experienced limitations due to the values of the moments that admit a proper probability density function (pdf). An alternative approach in developing the boundary conditions for the boundary of the positive region for both series expansions is investigated using Sturm's theorem. This alternative approach is used to develop the boundary of the positive and unimodal regions for both series expansions. Maple and Mathematica codes that allow verification of the moment ratio combinations that result in a proper pdf are provided. This method accurately determines the positive and unimodal regions of the expansions. Tabulated values are provided and plots included that illustrate the proposed method. The results can then be compared with the regions developed by others, including Barton & Dennis (1952).

87 Models and Methods for Infectious Disease Data

Section on Statistics in Epidemiology, Biometrics Section, Section on Health Policy Statistics, ENAR

Monday, August 9, 8:30 am–10:20 am

Multivariate Surveillance Schemes for Infectious Diseases on Multiple Locations

◆ Bo Hong, University of Alabama

University of Alabama, 430 Grace St., #222, Tuscaloosa, AL 35401

bhong1@bcc.cba.ua.edu

Key Words: *multivariate, surveillance, infectious diseases, control chart, time series, spatial*

A majority of the statistical research efforts on monitoring infectious disease surveillance have focused on univariate techniques, typically implemented on aggregated national disease data. However, most surveillance systems produce multivariate data: several reporting units are monitored simultaneously. The objective of this research is to design multivariate statistical surveillance schemes that simultaneously monitor incidence level of infectious diseases on multiple geographic locations, and provide early signals for the potential onset of epidemics in time and space. We examine multivariate control charts for autocorrelated processes by modeling the processes with multiple time-series method and then implementing multivariate control charts on forecast errors. Our proposed multivariate schemes outperform the univariate ones when applied to the weekly reported mortality of influenza and pneumonia from 1996 to 2004 for the United States. A systematic simulation study was also conducted to evaluate the performance of proposed surveillance schemes.

The Effective Reproduction Number and Control of Severe Acute Respiratory Syndrome

◆ Shenghai Zhang, Health Canada

Health Canada, Centre for Infectious Disease Prevention and Control, Tunney's Pasture, AL:0900B1, Ottawa, ON K1AOK9 Canada

Shenghai_Zhang@hc-sc.gc.ca

Key Words: *SARS, basic reproduction number, incubation period, severe acute respiratory syndrome*

Researchers has recently studied a new disease—severe acute respiratory syndrome (SARS). Although many results for the disease have been obtained, the mechanisms by which the virus spread are still not fully understood. A statistical methodology has been developed in this paper to estimate the distribution of incubation period and and effective reproduction number for the disease. The information from these estimations can be used to deal with the emergence of the disease and to help protect humans from the SARS virus in the future.

Immunity to Plague for Gerbils in Kazakhstan is Seasonal

◆ Siyun Park, University of Iowa; Kung-sik Chan, University of Iowa; Nils C. Stenseth, University of Oslo

University of Iowa, Dept. of Statistics and Actuarial Science, 241 Schaeffer Hall, Iowa City, IA 52242-1409

sypark2@hotmail.com

Key Words: *binomial distribution, continuous-time Markov chain, mixed-effect generalized nonlinear model, penal of time series*

The gerbil population in Kazakhstan is a natural focus of plague where the disease can be transmitted to humans by vectors, mainly, fleas. A long-term monitoring study of this natural system was undertaken from 1949-1995, for tracking the prevalence of plague in the gerbil population. In particular, in both spring and fall of each year, samples of gerbils were given bacteriological and serological tests for plague symptoms. While bacteriological tests detect for the presence of plague bacteria and hence the plague disease in the animal at the sampling time, serological tests detect for the presence of antibodies to plague bacteria. Consequently, the serological test data are indicative of past infections, and may shed light on the structure of immunity to plague. Using a continuous-time Makov chain model, we deduce a mixed-effect generalized nonlinear model for analyzing the serological test data, and find that for this gerbil system, immunity to plague is seasonal. Specifically, gerbils appeared to lose immunity more quickly in spring than in fall.

A Model to Estimate Risk of Infection with Human Herpesvirus 8 Associated with Transfusion from Cross-sectional Data

◆ Ruth M. Pfeiffer, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd., EPS 8030, Rockville, MD 20852

pfeiffer@mail.nih.gov

Key Words: *current status data, mixture models, multistate models*

In cross-sectional studies of infectious diseases, the data typically consist of: age at the time of study, status (presence or absence) of infection, and a chronology of events possibly associated with the disease. Motivated by a study of how human herpesvirus 8 (HHV-8) is transmitted among children with sickle cell anemia in Uganda, I have developed a flexible parametric approach for combining current-status data with a history of blood transfusions. I model heterogeneity in transfusion-associated risk by a child-specific random effect. I present an extension of the model to accommodate the fact that there is no gold standard for HHV-8 infection and infection status was assessed by a serological assay. The parameters are estimated via maximum likelihood. Finally, I present results from applying various parameterizations of the model to the Ugandan study.

Nonlinear Mixed Effects Models in the Prediction of S. Pneumoniae Antimicrobial Resistance Rates

◆ Robertino M. Mera, GlaxoSmithKline

GlaxoSmithKline, 1250 S. Collegeville Rd., UP-4335, Collegeville, PA 19426

robertino.m.mera@gsk.com

Key Words: *mixed models, antimicrobial resistance, epidemiology, prediction models*

Streptococcus pneumoniae infections are the most significant bacterial cause of respiratory illness in young children, the elderly and persons with chronic medical conditions. High-level penicillin resistance rates in the U.S. have been increasing from levels of 5.6% in 1992 to 24.3% in 2001. Although compartment-transmission mathematical models have improved our understanding of resistance dynamics at the individual level, prediction of resistance in broad populations requires a different approach. The present study implements a nonlinear mixed effects model that is consistent with the insights provided by the transmission models. The model takes advantage of a large amount of surveillance data; antibiotic prescriptions and other established risk factors at the state and regional level in the U.S. in the period 1996 to 2001. The application shows that penicillin resistance rates will reach a plateau from 2006 to 2008. The two most important factors that affect prediction are antibiotic consumption and the uptake of the conjugate pneumococcal vaccine. Penicillin consumption changes the timing of the plateau, and 90% plus vaccine coverage immediately reduces the resistance rate.

Bayesian Hierarchical Models for Estimating Vaccine Efficacy on Susceptibility and Infectiousness

◆ Xiaohong M. Davis, Centers for Disease Control and Prevention; Michael J. Haber, Emory University; Lance A. Waller, Emory University

Centers for Disease Control and Prevention, 4770 Buford Hwy. NE, Mailstop K-21, Atlanta, GA 30341

xmao@cdc.gov

Key Words: *vaccine efficacy, vaccine efficacy for susceptibility, vaccine efficacy for infectiousness, Bayesian hierarchical models, household random effect*

Likelihood methods have been used to estimate vaccine efficacy on susceptibility (VES) and on infectiousness (VEI) from household data assuming fixed transmission probabilities and vaccine effects. In real disease epidemic, the transmission probabilities and vaccine effects may not be the same across households, communities or even individuals. We propose Bayesian hierarchical models for estimating VES and VEI using final size (outbreak) data from a household study. Our method allows for both between- and within-household heterogeneities in transmission probabilities due to random effects at both household and individual level. We performed extensive Markov chain Monte Carlo (MCMC) simulations to evaluate the performance of the estimators of VES and VEI for different models and vaccination designs.

88 Bioinformatics

Biopharmaceutical Section

Monday, August 9, 8:30 am–10:20 am

Constructing a System for Quantitative Proteomics Based on Mass Spectrometry Techniques: A Data Analysis Challenge

◆ Matthew Westlake, RTI International; Georgiy Bobashev, RTI International; Benjamin J. Cargile, RTI International; James L. Stephenson, Jr., RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

mwestlake@rti.org

Key Words: *spectrometry, quantitative proteomics*

Measuring the differential expression of genes through mass spectrometry based techniques is a way to perform quantitative proteomics. In a typical experiment an isotopic tag is used to differentiate two samples whose protein content is extracted, possibly separated into discrete groups by electrophoresis or other methods, and the results are subjected to mass spectrometry analysis. The researcher is presented with a large number of spectra to compare. These spectra often contain noise and the identification of families of peaks representing peptides can be tedious and difficult, especially as the size of the proteome increases. Matching these peak families to proteins and comparing spectra is another challenge. The growth of mass-spectrometry-based techniques for measuring gene expression has prompted a need for tools to analyze the results. We present an approach developed to address these issues and automate the comparison process. Our process includes methods to separate signal from noise, identify peaks and peak families, identify shifts in peak location, identify incomplete peptide separation in adjacent spectra, and bring the results together for sample comparison.

Genome-wide Index of Communication Function for Measuring Distance between Functional Gene Groups

◆ Tianwei Yu, University of California, Los Angeles; Wei Sun, University of California, Los Angeles; Ker-Chau Li, University of California, Los Angeles

University of California, Los Angeles, 8130 Math Sciences, Dept. of Statistics, Los Angeles, CA 90095-1554

tyu@stat.ucla.edu

Key Words: *gene expression, microarray, Saccharomyces cerevisiae*

Microarray data provide valuable information about the co-expression of genes. The degree of association between two genes can be easily measured using Pearson's correlation or similar methods. To further the understanding of expression regulation at the functional module level, we devised a method to assess the degree of association between two subunits of cellular process represented by multiple genes associated with two GO terms. A genome-wide index of communication function (GIOC) for each module is defined on the discrete state space of all genes whose expression profile is available, by transforming the highest correlation between each gene and the module. Distance between modules is then measured by Kullback-Leibler divergence between their GIOC functions. This algorithm overcomes two obstacles: (1) not all genes from the same term are tightly co-expressed in general, and (2) the landscape of gene expression profiles is structured heterogeneously in a space of high dimension. When applied on representative yeast modules selected from the gene ontology system, the algorithm yields meaningful grouping and some interesting new putative relationships.

Data-adaptive Tuning of Multicategory Support Vector Machines

◆ Eway Y. Zhan, Rosetta Biosoftware

Rosetta Biosoftware, 12220, Suite 210, 113th Ave. NE, Kirkland, WA 98034

yihui_zhan@rosettabio.com

Key Words: *gene expression profiles, microarray measurements, kernel machine, supervised learning, cancer classification, diagnostic prediction of cancer*

We propose a criterion for selecting hyperparameters in multicategory support vector machines (MSVM) and apply it to microarray data analysis. The MSVM classifiers are based on a recent extension that inherits the optimal property of the binary SVM that treat all classes simultaneously. The tuning criterion is an extension of the radius-margin bound in binary case that has better smoothness properties than the GACV (generalized approximate cross-validation) criterion and is optimized by a direction set method with respect to both kernel parameters and a balancing parameter between data fit and the complexity of the decision function. Computational examples involving data in public domain are given to illustrate the application of the criterion in gene expression profile classification and gene selection.

U-Statistics for Microarrays: Normalization, Signal Value Estimation, Gene Expression Profiles

◆ Asifa Haider, Rockefeller University; Mayte Suarez-Farinas, Rockefeller University; Knut M. Wittkowski, Rockefeller University

Rockefeller University, 1230 York Ave. Box 21, New York, NY 10021

haidera@rockefeller.edu

Many of the currently used statistical algorithms for processing and analyzing microarray data are heuristically motivated adaptations of the linear model, tunable by a wealth of scaling and cut-off parameters. While these algorithms are widely used, the assumption of the linear model do typically not apply, the underlying heuristics are often flawed, and the choice of the parameters is highly subjective. We have recently demonstrated the use of u-statistics for multivariate data as a nonparametric alternative to linear models for the analysis of gene expression profiles. We are now applying the same approach to the lower levels of microarray analysis. Compared to methods such as "background subtraction with noise correction" and the replacement of observed mismatches by "ideal" mismatches in MAS 5.0, the proposed approach avoid often substantial biases, which further reduce the low sensitivity of microarrays. We apply this approach to correlating activity of anti-inflammatory drugs along genomic pathways with disease severity of psoriasis based on both clinical and histological parameters.

MAOSA: A New Procedure for Detection of Differential Gene Expression

◆ Greg Dyson, University of Michigan; C.F. Jeff Wu, Georgia Institute of Technology

University of Michigan,

gdyson@umich.edu

Key Words: *multiplicity, microarray, shrinkage, data mining*

The Multiplicity-Adjusted Order Statistics Analysis (MAOSA) is a new technique to determine differential expression in gene expression data. The MAOSA technique assumes the normality of the middle 1-delta of the distribution of the test statistics. The test statistics are adjusted by adding a constant c to the denominator to ensure that genes with low variability are not erroneously called significant. Then, after a transformation of these test statistics to the uniform scale, known features of differences in uniform order statistics are used to facilitate analysis. The inherent multiplicity problem (thousands of genes to be tested) will be eased by performing a Bonferroni correction on a small number of effects. This is a reasonable approach since the majority of genes are not differentially expressed. The MAOSA method can incorporate prior biological knowledge in the analysis with the selection of c. Datasets derived from a spotted glass array and a gene chip are analyzed using MAOSA and the results compared to other techniques.

Identification of Alternative Splicing Events Using Expression Data from Tiling Microarray

◆ Lisa H. Ying, Merck & Co., Inc.; Zhengyan Kan, Merck & Co., Inc.; John Castle, Merck & Co., Inc.; Jason Johnson, Merck & Co., Inc.; Vladimir Svetnik, Merck & Co., Inc.

Merck & Co., Inc., RY33-300, PO Box 2000, Rahway, NJ 07065

lisa_ying@merck.com

Key Words: *microarray, alternative splicing, Principle Component Analysis, clustering*

A great challenge in the post-genomic era is to decipher how exons from the same genes are assembled differently into different transcripts through a mechanism called alternative splicing. It is well known that alternative splicing plays important roles in the regulation of gene functions and makes significant contribution to protein complexity. Recent studies suggest that more than half of human genes are alternatively spliced. A tiling array experiment was performed using custom ink jet microarrays, with probes tiled across 200 known genic regions. In total, RNA samples from 52 diverse tissues were hybridized. Several statistical methods/techniques such as robust Principle Component Analysis, Mahalanobis distance, clustering, etc., were utilized to identify the alternative splicing events. Our results were evaluated against known alternative splicing events based on transcript analysis (RefSeq, mRNA, and EST).

Improving Estimation of the Conditional False Discovery Rate

◆ Stanley B. Pounds, St. Jude Children's Research Hospital; Cheng Cheng, St. Jude Children's Research Hospital

St. Jude Children's Research Hospital, Dept. of Biostatistics (MS 262), 332 N. Lauderdale St., Memphis, TN 38105

stanley.pounds@stjude.org

Key Words: *false discovery rate, q value, spacings, density estimation, loess, microarray*

Recent attempts to account for multiple testing in the analysis of microarray data have focused on controlling the false discovery rate (FDR). However, rigorous control of the FDR at a preselected level is often impractical. Consequently, it has been suggested to use the q value as an estimate of the proportion of false discoveries among a set of significant findings. However, this interpretation may be unwarranted considering that the q value is based on an unstable estimator of the positive false discovery rate. Another method proposes estimating the FDR by modeling p values as arising from a beta-uniform mixture (BUM) distribution. Unfortunately, the BUM approach is reliable only in settings where the assumed model accurately represents the actual distribution of p values. A method called the spacings loess histogram (SPLOSH) is proposed for estimating the conditional FDR, the expected proportion of false positives conditioned on having k "significant" findings. SPLOSH is designed to be more stable than the p value and applicable in a wider variety of settings than BUM.

89 Bayesian Variable and Model Selection and Hierarchical Models

Section on Bayesian Statistical Science

Monday, August 9, 8:30 am–10:20 am

Variable Selection and Covariance Selection in Multivariate Regression Models

◆ Christopher K. Carter, Commonwealth Scientific and Industrial Research Organization; Ed Cripps, University of New South Wales; Robert Kohn, University of New South Wales

Commonwealth Scientific and Industrial Research Organization, Locked Bag 17, North Ryde NSW, Mathematical and Information Sciences, Sydney, 1670 Australia

Chris.Carter@csiro.au

Key Words: *cross-sectional regression, longitudinal data, model averaging, Markov chain*

This article provides a general framework for Bayesian variable selection and covariance selection in a multivariate regression model with Gaussian errors. By variable selection we mean allowing certain regression coefficients to be zero. By covariance selection we mean allowing certain elements of the inverse covariance matrix to be zero. We estimate all the model parameters by model averaging using a Markov chain Monte Carlo simulation method. The methodology is illustrated by applying it to four real datasets. The effectiveness of variable selection and covariance selection in estimating the multivariate regression model is assessed by using four loss functions and four simulated datasets. Each of the simulated datasets is based on parameter estimates obtained from a corresponding real dataset.

Choices of Priors and Predictive Performance of Bayes Variable Selection Procedures

◆ Wen Cui, Texas State University

Texas State University, Dept. of CIS and QMST, 601 University Dr., San Marcos, TX 78666

jcui@txstate.edu

Key Words: *Bayesian, variable selection, priors, predictive performance*

Bayesian approach to the variable selection problem is straightforward and promising. However, specification of priors is often difficult and guideless. In this study, the focus is on searching for a guideline of prior specifications for Bayesian variable selection procedures and obtaining the insights about the sensitivity of Bayesian variable selection procedures to various choices of priors. For this purpose, in the context of normal linear regression, variety of priors on both model parameter and model specific parameters were examined. Their ability to lead to close-form posteriors and their impact on the predictive performance of Bayesian procedures were investigated through extensive simulations.

Robust Bayesian Variable Selection and Its Application to QTL Studies

◆ Cheongeun Oh, Yale University

Yale University, 60 College St., New Haven, CT 06520

cheongeun.oh@yale.edu

Key Words: *Bayesian, variable selection, QTL problems*

Variable selection is a crucial problem in multiple regression. From a Bayesian prospective, George and McCulloch (1993) proposed a procedure that uses Gibbs sampling for selection promising subsets, called Stochastic Search Variable Selection. It entails embedding a hierarchical normal mixture model in the regression setup where latent variables are used to identify subset choices. However, as pointed by several authors, results of their procedure are highly sensitive to priors. The purpose of this paper is to develop a robust and efficient Bayesian variable selection method. We propose two important modifications. First, we modify a prior structure on the coefficients to make the procedure computationally efficient, especially when a large number of candidate factors are considered. Second, instead of looking at the posterior distribution of each candidate model, we focus ranking the candidate variables according to their marginal posterior probability, which is shown to be more robust. We also suggest how to deal with highly correlated factors which have been known as a major challenge for Stochastic Search Variable Selection.

Hierarchical Models for Assessing Variability among Functions

◆ Sam Behseta, California State University, Bakersfield; Garrick L. Wallstrom, University of Pittsburgh; Robert E. Kass, Carnegie Mellon University

California State University, Bakersfield, 9001 Stockdale Hwy., Dept. of Mathematics, Bakersfield, CA 93311

sbehseta@csub.edu

Key Words: *functional data analysis, Bayesian statistics, variance estimation, neuronal data analysis*

We present a method for summarizing functional variation based on fits, with taking account of the estimation process. We show, for example, that the proportion of variance associated with the first principal component of a sample covariance matrix computed from estimated functions will be biased upward. Alternatively, we introduce three Bayesian methods of accounting for estimation variation, all based on hierarchical models using free-knots splines. Our approach extends the Bayesian Adaptive Regression Spline method of DiMatteo, Genovese and Kass (2001). We rely upon the notion of a hierarchical Gaussian Process model which uses the approximate normality of the estimated function values. We apply our Hierarchical Gaussian Process (HGP) model to neuronal data obtained from the primary motor cortex area of the brain. We demonstrate a considerably lower estimation of the proportion of variability attributed to the first principal component.

Bayesian Inference in Generalized Additive Mixed Models with Normal Random Effects

◆ Yisheng Li, University of Texas M.D. Anderson Cancer Center; Xihong Lin, University of Michigan

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030

ysli@mdanderson.org

Key Words: *Bayesian generalized additive mixed model, integrated Wiener prior, smoothing spline, generalized linear mixed model, Gibbs sampling, adaptive rejection sampling*

We propose Bayesian generalized additive mixed models for correlated data, which arise frequently in studies involving clustered, hierarchical, and spatial designs. The models allow for additive functional dependence of an outcome variable on covariates by using nonparametric regression and account for correlation between observations using random effects. Partially improper integrated Wiener priors are used for the nonparametric functions and the resulting estimators are cubic smoothing splines. Systematic inference on model parameters can be made within a modified generalized linear mixed model framework. Computation is carried out using Gibbs sampling. We illustrate the proposed approach by analyzing an infectious disease dataset and compare its performance with that of the double penalized quasi-likelihood approach through simulation.

Bayesian Models and Model Selection in Closed Population Capture-recapture Experiments

Sujit K. Ghosh, North Carolina State University; ◆ Ross M. Gosky, North Carolina State University

North Carolina State University, Dept. of Statistics, Campus Box 8203, Raleigh, NC 27695-8203

rmgosky@ncsu.edu

Key Words: *Bayesian methods, model selection, capture-recapture*

Capture-recapture models are used to estimate the unknown sizes of animal populations. When the population is closed, with constant size during the study, eight standard models exist for estimating population size. These models allow for variation in animal capture probabilities due to time effects, heterogeneity among animals, and behavioral effects after the first capture. Using Bayesian statistical modeling, we present versions of these eight models. We explore the use of Akaike's Information Criterion (AIC), and the Deviance Information Criterion (DIC) as tools for selecting the appropriate model for a given dataset. Through simulation, we show that AIC performs well in model selection.

Using Subpopulation Membership as a Bayesian Diagnostic Technique

◆ Guofen Yan, Cleveland Clinic Foundation; J. Sedransk, Case Western Reserve University

Cleveland Clinic Foundation, 9500 Euclid Ave., Wb4, Cleveland, OH 44195

gyan@bio.ri.ccf.org

Key Words: *discrepancy, model assessment, posterior predictive p value, test statistic*

We investigate the ability of posterior predictive p values to detect unknown hierarchical structure. Since there is limited research in this area we study the simple, but important case where the data come from a two-stage hierarchical regression model while the fitted model does not include this feature. We consider a general class of discrepancy measures (i.e., minimum, maximum, 10th and 90th percentiles, mean, median and standard deviation), which are not motivated by knowledge of the hierarchical structure. We show that diagnostics based on some of the discrepancy measures listed above are not effective when applied to the entire dataset, but such diagnostics are useful when applied to important subpopulations.

90 Regular Contributed Posters

ENAR, SSC, General Methodology, Biometrics Section, Biopharmaceutical Section, Section on Bayesian Statistical Science, Business and Economics Statistics Section, Section on Statistical Computing, Section on Statistics in Epidemiology, Section on Statistical Graphics, Section on Statistics in Sports

Monday, August 9, 9:00 am–10:50 am

Modeling Tumor Growth in Cancer Clinical Trial

◆ Jun Wu, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, DC6054, Indianapolis, IN 46285

wuju@lilly.com

Key Words: *cancer, time-to-tumor progression, clinical trial, growth curve, survival time*

In clinical trial settings, tumor growth is followed by measuring the size of tumors by common diagnostic imaging method on a regular basis. The sum of all tumors at post-baseline measurement is compared with that at baseline. Tumor progression is determined if the sum increases by a clinically significant proportion over that at baseline. Time from randomization to tumor progression (TTP) is a common clinical endpoint that will be compared between treatment arms with time-to-event method. Malignant tumor progression tends to occur in short time frame relative to the interval between measurements, which may affect the analysis. Effect of augmenting the TTP analysis with growth curve (linear, exponential and Gompertz) of individuals' tumors on hypothesis testing will be studied on simulated data. The correlation of modeling tumor growth with survival time will also be discussed.

Graphical and Clustering Methods for Exploring Regions of Misfit in Latent Variable Models

◆ Todd E. Bodner, Portland State University; Daniel Wilson, Portland State University

Portland State University, Dept. of Psychology, PO Box 751, Portland, OR 97207

tbodner@pdx.edu

Key Words: *clustering methods, exploratory data analysis, factor analysis, graphical displays, goodness of fit, structural equation modeling*

In practice, latent variable model misfit is assessed with scalar statistics at the level of the overall model (e.g., chi-square statistics or fit indices like RMSEA) or at the level of missing individual paths (e.g., modification indices). The present paper discusses graphical and clustering methods that might be useful for exploring model misfit between these two extremes. These approaches focus on information available in the discrepancy matrix (i.e., the matrix of differences between the observed and model-implied covariances). Looking for patterns of misfit in a discrepancy matrix by inspecting its scalar elements becomes more difficult as the number of indicator variables increases. Using readily available software, we propose creating a 2- or 3-D picture of the discrepancy matrix where colors or bar-lengths represent the direction and magnitude of the discrepancies. Descriptively, one can then see where regions of misfit lie, if anywhere. Permutation of the rows and columns of the discrepancy matrix to cluster large discrepancies can further aid exploration. The efficacy of these approaches are explored and illustrated with real and simulated data.

A Multivariate Statistical Analysis of Stock Trends

◆ James B. Lawrence, Miami University; April Kerby, Alma College; Vasant Waikar, Miami University

Miami University, 208 Bishop Hall, 300 E Spring St., Oxford, OH 45056

lawrenjb@muohio.edu

Key Words: *discriminant, analysis, stocks*

The stock market is a financial game of winners and losers. Is your stock tried and true or here today, gone tomorrow? How can one pick out a golden nugget like Microsoft from the hundreds of dot-coms that went bust after an all-too-brief moment of glory? It may seem like there is really no way to tell—a seemingly uncountable number of variables influence our markets and companies; how can one take all of them into account? Is there a simpler way of looking at the market madness? This paper seeks to use statistical methods to survey and analyze financial and economic data to discover such a method of simplification. Using principal component analysis, we will combine related factors into a smaller number of key components largely responsible for the variations observed. Then, using discriminant analysis, we will develop a model for separating companies into two categories based on their predicted stock performance: good and poor investment choices.

A Multivariate Statistical Analysis of the NBA

◆ Lori Hoffman, The Ohio State University; Maria Joseph, Kentucky State University; Vasant Waikar, Miami University

The Ohio State University, 2676 3B Cannon Point Ct., Columbus, OH 43209

hoffler23@hotmail.com

Key Words: *discriminant, analysis, sports*

Will your favorite National Basketball Association (NBA) team make it to the playoffs this year? What variables affect a team's postseason outcome? In an attempt to determine which teams will make the NBA playoffs, we will collect and analyze team data using multivariate statistical methods including principal components analysis and discriminant analysis.

Statistical Models for the Influence of Death of Spouse on the Depression Status of the Elderly Twins

◆ Maria Iachina, University of Southern Denmark; Ivan Iachine, University of Southern Denmark

University of Southern Denmark, Epidemiology, Institute of Public Health, Campusvej 55, 5230 Odense M, DK-5230 Denmark

miachina@health.sdu.dk

Key Words: *depression symptoms, death of spouse, missing data, multiple imputation, GLM, longitudinal data*

The Longitudinal Study of Aging Danish Twins included information on depression symptoms which were combined with data on death of spouse between consequent interviews. The aim was to study the short- and long-term influence of death of spouse on depression status taking into account the serial and intrapair twin correlation and incompleteness of depression data due to drop-outs. A logistic regression analysis indicated that observed depression score was a strong predictor of future drop-out, suggesting the inappropriateness of the complete case analysis. To deal with these issues we developed a marginal generalized linear model which related the mean depression changes with lagged death of spouse data. The incomplete data problem was handled using multiple imputation. Standard errors were obtained using a clustered version of the Huber-White sandwich estimator. The analysis shows a significant effect of the death of spouse on the depression status of the surviving person, the depression score increases by about 10%. The average depression score returns to normal in about 12 months for both males and females. Alternative likelihood-based approaches are also discussed.

A Generative Approach to Dendrograms with Application to Functional Biomechanical Data

◆ Jeffrey A. Bakal, Queen's University ; Glen Takahara, Queen's University; John T. Smith, Queen's University; Joan Stevenson, Queen's University

Queen's University, Room 310, Jeffery Hall, Kingston, ON K7L 3N6 Canada

jbakal@mast.queensu.ca

Key Words: *cluster analysis, functional data analysis, biomechanics*

Classical hierarchical clustering is a common method for the unsupervised classification of data. An attractive feature of this method is the graphical representation of the clustering, the dendrogram, which facilitates estimation of the number of clusters as well as providing certain information about the clusters. Euclidean distance is not always appropriate for representing dissimilarity between objects and clusters. We propose a dissimilarity measure based on the joint probability of membership in the same cluster, computed in terms of an estimated multivariate normal mixture model. The advantages of this method are pertinent to the analysis of functional data and are illustrated in the context of functional data arising in biomechanics.

A Bootstrapping Approach to Adjust for Unequal Distribution of Compounds in the Chemical Space When Assessing the Performance of Three Toxicity Prediction Programs

◆ Robert H. Gallavan, Jr, Pfizer Inc.; Richard Mueller, Pfizer Inc.

Pfizer Inc., 700 Chesterfield Pkwy. W, Mail Stop T2W, Chesterfield, MO 63017-1732

robert.h.gallavan@pfizer.com

Key Words: *bootstrap, in silico, toxicology, DEREK, CASETOX, TOPKAT*

This work was part of a study to evaluate the ability of three computer programs (TOPKAT, DEREK for Windows, and CASETOX) to predict mutagenicity for 520 drug candidates. Sensitivity, specificity, and accuracy were determined for each program based on the predicted and observed results of the Ames test. Initial results showed a marked difference in sensitivity (DEREK: 28%; CASETOX: 50%; TOPKAT 63%). To determine if the results were skewed by an unequal distribution of compounds throughout the chemical space, the drug candidates were assigned to one of 188 classes based on chemical structure and 1,000 datasets of 188 compounds were generated by randomly selecting one compound from each class. The performance of each program was evaluated as before for each of these datasets and the median and upper and lower 95% confidence intervals for sensitivity, specificity, and accuracy were determined using the empirical distribution functions. The median sensitivity of DEREK under these conditions (52%) was within the 95% confidence intervals for CASETOX and TOPKAT [(50%, 69%) for both], indicating that the three programs performed similarly, though poorly, across the chemical space.

A Mixed Models Approach to Nest Site Fidelity of Leatherback Turtles at Playa Grande

◆ Eric Nordmoe, Kalamazoo College; Samantha Knapman, Kalamazoo College; Paul Sotherland, Kalamazoo College; James Spotila, Drexel University; Frank Paladino, Indiana-Purdue University; Richard Reina, Monash University

Kalamazoo College, 1200 Academy St., Kalamazoo, MI 49006

enordmoe@kzoo.edu

Key Words: *mixed models, multilevel models, ecology*

This study applies hierarchical linear mixed models to investigate whether observed nesting patterns of leatherback turtles (Dermochelys coriacea) support the presence of individual heterogeneity in nest site selections. Data were collected on the nest site selection of leatherback turtles at Playa Grande in Parque Nacional Las Baulas, Costa Rica, for eight seasons from 1993-94 to 2000-01. Using these data, models were constructed to describe the nest site selections of leatherbacks along the coastal axis of Playa Grande. Using a mixed model approach, we find small but significant variability in nest site selections both between individuals ($P< 0.0001$) and within-individual between seasons ($P=0.0091$). Individual differences between turtles yield estimated model-based intraclass correlations between same-season nest site selections of individual turtles ranging from 0.097 to 0.142 for the eight seasons studied. Similarly, the estimated intraclass correlations between individual nest site selections in different seasons ranged from 0.058 to 0.085. The analysis suggests small but significant between-turtle heterogeneity in nest site preferences.

The Use of Repeated Cross-sections Markov Model to Estimate Overweight Transition Probabilities among Brazilian Children and Adolescents

◆ Fernando A.B. Colugnati, Federal University of São Paulo; Francisco Louzada-Neto, Federal University of São Carlos; José Augusto A.C. Taddei, Federal University of São Paulo

Federal University of São Paulo, R. Loefgreen, 1647, UNIFESP, São Paulo, 04040-032 Brazil

fcolugnati@uol.com.br

Key Words: *Markov models, pseudo-panel, repeated cross-sections, transition probabilities, overweight, obesity*

Overweight in childhood and adolescence is epidemic in emerging urban societies. The knowledge on overweight onset dynamics, in terms of population characteristics, plays an important role to plan and implement preventive programs and policies, identifying social and regional strata presenting higher transition rates. However, the difficulties in conducting prospective studies, due to high costs and long-term results, make repeated cross-sections (pseudo-panel) the best approach to estimating overweight growing rates for different population strata. Aggregate data from various comparable sources are utilized to estimate the transition probabilities, where individual information is substituted by population profiles that are considered as population cohorts. Finally, the results are compared with the traditional dynamic Markov Model approach using data from a seven-year prospective study carried out among Brazilian students.

Smoothing of Multimodal Histograms to Attain Unimodality: Application to Telomere Lengths and Clonal Heterogeneity in Leukemia Patients

◆ Martin L. Lesser, North Shore LIJ Research Institute; Julia Y. Tai, North Shore LIJ Research Institute; Rajendra Damle, North Shore LIJ Research Institute; Nicholas Chiorazzi, North Shore LIJ Research Institute

North Shore LIJ Research Institute, 1129 Northern Blvd. Suite 302, Manhasset, NY 11030

marty@nshs.edu

Key Words: *multimodality, histograms, smoothing, clonal heterogeneity*

Do observers agree on when a multimodal histogram (hgm) becomes unimodal after repeated smoothing? Does the point at which unimodality occurred result in clusters of hgms? Telomeres (tels) are the distal ends of chromosomes that become shorter as cells age. By studying tel lengths in leukemia patients (pts), patterns of clonal heterogeneity may be identified. Hgms of tel lengths were created for each of 70 leukemia pts. Median no. tels in the hgms was 1,693. Hgms showed one or two large modes and multiple local maxima in the tails. Each hgm was then repeatedly smoothed with a kernel density estimator with increasing bandwidths (BW). Four observers recorded the minimum BW at which the hgm appeared unimodal. The intraclass correlation for the four observers was 0.8, indicating generally good agreement on a highly subjective criterion. After mean BW across observers was computed for each hgm, the distribution of the 70 mean BWs suggested clusters of three or four groups. Multiple observers can agree as to when a smoothed histogram becomes unimodal. Some leukemia pts may have multimodal hgms, suggesting clonal heterogeneity.

Matrix Valued General Linear Modeling of Music Instrumentation Digital Interface Data

◆ Lisa Cannon, Sam Houston State University; Cecil R. Hallum, Sam Houston State University

Sam Houston State University, Dept. of Mathematics and Statistics, Huntsville, TX 77341

mth_crh@shsu.edu

Key Words: *linear models, multivariate analysis, MIDI, generalized inverses, matrix differentials*

Music Instrumentation Digital Interface MIDI) data is music data in multivariate form. Each musical tune is expressible as a matrix with each column comprising the frequencies of play for each instrument for each clock beat in time. This paper utilizes multivariate statistical techniques and linear algebraic structures to attain unique insights into MIDI music. Matrix differentials and generalized inverses provide opportunities in matrix-valued linear modeling of these MIDI tunes to investigate the relationship between differing tunes, artists, music categories, etc. Applications resulting from converting vector-valued data into sound/music forms are included and discussed particularly as relates to assisting seeing-impaired individuals "hear" critical information in vector-valued data.

Developing an Outcome of Adult Respiratory Distress Syndrome in a Trauma Population

◆ Christina A. Gaughan, University of Pennsylvania School of Medicine; Jason D. Christie, University of Pennsylvania School of Medicine; Paul N. Lanken, University of Pennsylvania; Barbara B. Finkel, University of Pennsylvania; Barry Fuchs, University of Pennsylvania; Robert Gallop, West Chester University; J. Richard Landis, University of Pennsylvania School of Medicine

University of Pennsylvania School of Medicine, Center for Clinical Epidemiology and Biostatistics, 525B Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104

cgaughan@cceb.upenn.edu

Key Words: *biostatistics, outcome classification, sensitivity analysis*

An outcome is not always defined as straightforward categorical presence or absence of a disease or syndrome. This problem is particularly poignant when a syndrome has several defining criteria with variable precision in their definitions. Such is the case when identifying ARDS. By consensus, ARDS is defined as meeting three criteria within a 24-hour period: (1) bilateral pulmonary infiltrates on chest x-ray consistent with pulmonary edema; (2) absence of evidence of left atrial hypertension; and (3) poor systemic oxygenation (PaO2/FiO2 ratio <=200). Within a prospective cohort study of 273 trauma patients aimed at understanding the genetic and biological mechanisms of ARDS, we sought to assess the time of onset of ARDS as well as generate methods to minimize outcome misclassification. The following challenges were addressed: (1) agreement of chest radiograph interpretation; (2) interpretation of longitudinal changes in PaO2/FiO2 ratio; (3) classification of subjects meeting partial or nonsimultaneous criteria; (4) timing of classification relative to other time-varying covariates; (5) robustness of ARDS definition to variations in PaO2/FiO2 thresholds (sensitivity). NHLBI SCOR acknowledgment.

Decision Analysis and Substitutions in the Boston Red Sox

◆ Elaine Allen, Babson College; Tanya Mercure, Babson College; Jared Campbell, Babson College

Babson College, Math/Science Division, Wellesley, MA 02457

allenie@babson.edu

Key Words: *decision analysis, sports, baseball*

It's the bottom of the eighth inning, you're the manager for the Boston Red Sox facing the rival New York Yankees in Game 7 of the 2003 American League Championship. What do you do? With a laboring Pedro Martinez on the mound, Manager Grady Little decided to keep Martinez in for the bottom of the eighth. Comparing historic data on all pitchers available to the Red Sox before the bottom of the eighth inning to the batters scheduled to bat we used decision analysis to determine the best possible outcome of the game. Additionally, looking back at the 2003 season, we examined the batting performance and transitions to "slumps" by the Red Sox to determine the best course of action for batting substitutions over the season.

A Gap-statistic-based Method of Determining the Number of Clusters

◆ Mingjin Yan, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, 1776 Liberty Lane Apt. C34, Dept. of Statistics, Blacksburg, VA 24060

myan@vt.edu

Key Words: *Gap statistic, K-means, clustering*

Cluster analysis has attracted increasing attention as an exploratory tool in the “unsupervised” learning of gene expression data. Currently, the major concerns in clustering analysis lie in two aspects: what clustering method should be implemented and how to determine the number of clusters in a dataset. In spite of its importance, making inference about the correct group number in a dataset is not an easy task due to no clear definition of cluster.” Widely used clustering algorithms (e.g., K-means or hierarchical) give heuristic but no determined results about the number of groups. Tibshirani et al. proposed a method which formalized the estimation of the number of data clusters via the Gap statistic. A new technique based on the Gap statistic is developed to find the optimal cluster number. Performances of this new method and the Gap statistic are compared with both simulated and real-world datasets.

Design and Analysis of Split-mouth Dental Trials

◆ Donald E. Mercante, LSU Health Sciences Center

LSU Health Sciences Center, School of Dentistry, Box 137, 1100 Florida Blvd, New Orleans, LA 70119

dmerca@lsuhsc.edu

Key Words: *split-mouth design, incomplete block design, cross-over, correlated data*

Studies in which the person is subdivided into several smaller experimental units are common in dentistry. In particular, regions of a mouth, such as opposing upper quadrants, are randomly assigned a treatment or intervention and the response measured. Sampling plan may include obtaining subexperimental level information at several specific sites within each quadrant or experimental unit. The overall goal of these trials is to provide increased efficiency with respect to whole-mouth (i.e., parallel group) trials. These designs are similar in nature to cross-over designs but without washout periods. A possible advantage being that the trials can be conducted in less time, or with fewer patients than either a parallel or crossover trial. However, if carry-over effects are present, then treatment effects are potentially biased. Additionally, in many dental trials there are more treatments than available experimental units, thus necessitating the use of an incomplete block design. We focus on the design and analysis of split-mouth dental trials when using incomplete block designs. Issues of estimability, confounding, and efficiency of these designs are addressed.

What Accounts for the Goals Scored in the NHL?

◆ Norean R. Sharpe, Babson College; Brett Adamczyk, Babson College; Andrew Mullin, Babson College

Babson College, Division of Math & Science, Babson Hall, Babson Park, MA 02457

sharpen@babson.edu

Key Words: *sports data, hockey, intervention model, regression model, time series analysis*

This study examines the average number of goals scored per game over time since 1917 using a unique dataset containing both league-based and team-based variables. The league variables include changes in rules and equipment, while the team-based variables include penalty minutes, schedule length, and several performance measures. An intervention model, based solely on when the league changes occurred, is developed in conjunction with a model using the continuous performance measures both for the entire time frame of the dataset, as well as for a more recent period in the NHL. In addition, traditional time series models, such as moving averages, exponential smoothing, and ARIMAs are developed as a source of comparison with the intervention and regression models. All models are evaluated on the basis of retrospective accuracy, as well as accuracy against a hold-out sample. The impact of equipment changes, such as the size of goalie masks and goalie pads, on the frequency of goals scored is discussed.

A Comparison of Two Fully Bayesian Changepoint Models for Early Detection of Prostate Cancer

◆ Wonsuk Yoo, Medical University of South Carolina; Elizabeth H. Slate, Medical University of South Carolina

Medical University of South Carolina, 135 Cannon St., Suite 303, Charleston, SC 29425

yoow@musc.edu

Key Words: *Prostate Specific Antigen (PSA), Bayesian changepoint model, reversible jump Markov chain Monte Carlo (MCMC), conditional predictive ordinate (CPO), pseudo-Bayes factor, ROC curve*

This research compares two models for longitudinal prostate specific antigen (PSA) readings according to their ability to detect prostate cancer (Pca) onset. Model I is a fully Bayesian hierarchical changepoint model, similar to that of Slate and Clark (1999), in which cancer onset is represented as a changepoint in the men’s PSA trajectories and all men are presumed to experience onset eventually. Model II postulates a mixture for the PSA series, for which one component contains a change point and the other does not. Both models are fit using Markov chain Monte Carlo methods, with a reversible jump implementation for model II. We apply these models to data from the Nutritional Prevention of Cancer Trials, and investigate the effects of covariates (smoking, alcohol ge, and body mass index) on PSA growth by examining credible regions for covariate parameters and by computing conditional predictive ordinate values and pseudo-Bayes factors. We then use ROC curves to compare the performance of diagnostic rules for Pca onset derived from the posterior distributions of the changepoints for each model.

Visualization Techniques to Display Pairwise Rankings

◆ Cathleen A. Platt, San Francisco State University
San Francisco State University, 99 Stonegate Rd., Portola Valley, CA 94028
jplatt@sfsu.edu

Key Words: *pairwise rankings, visualization, hockey*

There are many circumstances, ranging from sports rating systems and post-season selections to consumer preference models, that rely upon pairwise rankings and supporting statistical models. This paper develops visual displays that capture and reveal the underlying structure of pairwise rankings. The techniques are applied to college hockey rankings. There are 58 teams in U.S. Division I NCAA hockey, each playing between 20 and 35 regular-season games. Sixteen teams are selected for post-season play. Five of the selections are automatic bids to conference champions, whereas the remaining 11 bids are determined by pairwise rankings. To complicate matters, the current selection criteria now include a “bonus” to a team’s ratings percentage index (RPI) for “quality wins,” defined as nonconference wins against a team in the Top 15 of RPI. More bonus points are awarded for neutral-site wins than home wins, and even more for Rd. wins. Because the number of pairwise comparisons is large and the bonus-points uncertain, visual displays that capture the essence of the comparisons and rankings, and that easily can be updated following each game, should prove helpful.

A Convenient Kernel Fisher Discriminant

◆ Mei-Hsien Lee, National Taiwan University
National Taiwan University, 6F No. 24 Gongyuan Rd., Banciao, Taipei, 220, Taiwan, R.O.C.
meihsien@stat.sinica.edu.tw

Key Words: *classification, Fisher linear discriminant analysis, regularized discriminant analysis, reproducing kernel, reproducing kernel Hilbert space, support vector machine*

The Fisher linear discriminant has long been a popular tool for classification. Recently, it has been extended via kernel machine to nonlinear discriminant. Such an extension of the classical Fisher linear discriminant to the kernel Fisher discriminant involves maximizing a regularized Rayleigh coefficient, which requires some computational efforts. We propose a convenient version of kernel Fisher discriminant (KFD), which does not involve any optimization, but only simple operations of addition and subtraction of kernel values. We also show that this convenient version of KFD is a criterion based on likelihood ratio of two Gaussians on a reproducing kernel Hilbert space. The proposed method is test on simulated data as well as real data.

91 Introductory Overview Lecture on Sequential and Adaptive Methods in Modern Statistics—Real-life Applications

ASA, ENAR, IMS, SSC, WNAR, Section on Survey Research Methods
Monday, August 9, 10:30 am–12:20 pm

Applications of Sequential Methodologies in Some Modern Statistical Areas

◆ Nitis Mukhopadhyay, University of Connecticut
University of Connecticut, Dept. of Statistics, U 4120, CLAS Bldg, 215 Glenbrook Rd., Storrs, CT 06269-4120
mukhop@uconnvm.uconn.edu

Abraham Wald and his colleagues pioneered the broad area of sequential analysis in the mid-1940s. Wald was definitely motivated to launch this exciting branch of statistics in order to deal with many important issues of sampling inspection arising from defense related practical projects of national importance. In the past two decades, it seems that the area of clinical trials has been one of the primary beneficiaries of many sophisticated sequential methodological research. It must also be said that the area of sequential analysis has come to maturity in many respects because of the advancement of clinical research. This presentation will begin with a review of some of the areas and types of modern statistical problems (other than those in clinical research) where innovative sequential methodologies have played major roles during the past 20-plus years. Modern statistical areas dealing with significant problems arising, for example, in surveillance, tracking, data-mining, agriculture, integrated pest management, psychology, entomology, finance, will be emphasized.

Clinical Trials, Medical Ethics, and Statistical Reasoning

◆ Pranab K. Sen, University of North Carolina, Chapel Hill
University of North Carolina, Chapel Hill, Depts. of Biostatistics and Statistics, Chapel Hill, NC 27599-0001
pksen@bios.unc.edu

The Helsinki 1997 declaration by the World Medical Association has generated a number of practical issues connected with the conduct of clinical trials, especially in the third-world countries, where, in addition to debatable medical ethics (for using human subjects under discretionary environment), the affordability as well as cost-benefit issues are of primary concern. The conventional Placebo-Controlled Trials (PCT), for which statistical resolutions have been in the offing since 1970s, need a critical appraisal, and Activity-Controlled Equivalence Trials (ACET) have been advocated. This change in scenario signals a number of statistical issues that need to be addressed in the light of practical adaptability as well as methodologic support. Some of these issues are appraised in the light of time-sequential nonparametrics (TSN), and the needed modifications of the otherwise existing methodology are discussed in detail. The issue of incorporating interim analysis schemes is a vital part of this appraisal.

92 New Advances of Generalized Linear Models in Industrial Experiments

Business and Economics Statistics Section, Section on Quality and Productivity

Monday, August 9, 10:30 am–12:20 pm

Applying Dual Response Methodology to Robust Parameter Design for a Generalized Linear Model

◆ William A. Brenneman, Procter & Gamble Company; William R. Myers, Procter & Gamble Company; Raymond H. Myers, Virginia Polytechnic Institute and State University

Procter & Gamble Company, 8700 Mason-Montgomery Rd., Mason, OH 45040

brenneman.wa@pg.com

Key Words: *robust parameter design, generalized linear models, dual response approach, dispersion effects, noise variables, response surface methodology*

Robust parameter design (RPD) has been used extensively in industrial experiments, since its introduction by Genichi Taguchi. RPD has been studied and applied, in most cases, assuming a linear model under standard assumptions. More recently, RPD has been considered in a generalized linear model (GLM) setting. We apply a dual response approach when using RPD in the case of a GLM. We motivate the need for exploring both the process mean and process variance, by discussing situations when compromise between the two is necessary. Several examples are provided to further motivate the need for applying a dual response approach when applying RPD in the case of a GLM.

Graphical Methods for Design Assessment for Designs Involving Generalized Linear Models

◆ Christine Anderson-Cook, Virginia Polytechnic Institute and State University; Ayca O. Godfrey, Virginia Polytechnic Institute and State University; Raymond H. Myers, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics, Blacksburg, VA 24061-0439

candcook@vt.edu

Key Words: *design optimality, scaled prediction variance, fraction of design space plots*

Fraction of design space plots have been used to assess designs in a number of design situations. Understanding the influence of initial parameter guesses on the choice of design locations is an important aspect that can be considered with the FDS plots. An important difference between the linear models and generalized linear models is the need for two separate measures, the scaled prediction variance and the penalized prediction variance, to summarize variance properties. Various types of robustness to parameter misspecification are considered and illustrated with several examples.

Monitoring Mean Shifts for Multistage Processes Using Generalized Linear Models

◆ Douglas Montgomery, Arizona State University; Duangporn Jearkpaporn, Arizona State University; George Runger, Arizona State University; Connie Borror, Arizona State University

Arizona State University, Tempe, AZ 85287

doug.montgomery@asu.edu

Key Words: *model-based control charts, generalized linear models, multistage process*

This paper develops a monitoring scheme for detecting a mean shift of a multistage manufacturing process for a mixture of normally and non-normally distributed variables (and gamma-distributed responses). The procedure is based on a deviance residual obtained from a generalized linear model. The deviance residual is shown to be a likelihood ratio statistic for detecting a mean shift in many cases for distributions in exponential family class. The performance of the proposed method is investigated and compared with control charts based on individual observations and T2 chart are provided and illustrated by a simulation study.

93 Multivariate Analysis: In Celebration of Ingram Olkin's 80th Birthday

Biometrics Section

Monday, August 9, 10:30 am–12:20 pm

A SINful Approach to Gaussian Graphical Model Selection

◆ Michael D. Perlman, University of Washington; Mathias Drton, University of Washington

University of Washington, Dept. of Statistics, Box 354322, Seattle, WA 98195-4322

michael@ms.washington.edu

Key Words: *graphical model selection, simultaneous tests, concentration graphs, covariance graphs, DAG, chain graphs*

Multivariate Gaussian graphical models are defined in terms of Markov properties, such as conditional independences associated with the underlying graph. Thus, model selection can be performed by testing these conditional independences, which are equivalent to specified zeroes among certain (partial) correlation coefficients. For concentration graphs, covariance graphs, acyclic directed graphs, and chain graphs (both LWF and AMP), we apply Fisher's z-transformation, Sidak's correlation inequality, and Holm's step-down procedure, to simultaneously test the multiple hypotheses obtained from the Markov properties. This leads to a simple method for model selection that controls the overall error rate for incorrect edge inclusion. In practice, we advocate partitioning the simultaneous p values into three disjoint sets, a significant set S, an indeterminate set I, and a nonsignificant set N. Then our SIN

model selection method selects two graphs, a graph whose edges correspond to the union of S and I, and a more conservative graph whose edges correspond to S only. Prior information about the presence or absence of particular edges can be incorporated readily.

Inequalities after 25 years: Schur-ly Generalizable

◆ Barry C. Arnold, University of California, Riverside

University of California, Riverside, Statistics Dept., Riverside, CA 92521-0002

barry.arnold@ucr.edu

Key Words: *majorization, cone order, Schur convex functions, inequality measurement*

In 1979, after a gestation period of nine years, Al Marshall and Ingram Olkin published their landmark volume, *Inequalities: Theory of Majorization and its Applications*. Prior to that date, only a limited number of statistics researchers utilized concepts related to majorization in their work. Subsequent to the publication of the red volume, and to a large degree because of its publication, it seems as though everyone has jumped on the majorization bandwagon. Any problem whose solution is a vector of the form (c,...,c) seems susceptible to rephrasing in terms of some cleverly chosen Schur convex function and its extreme value under the majorization ordering. Examples abound in areas such as: Experimental Design, Reliability, Multivariate Analysis, Inequality Measurement, Stochastic Ordering, Computer Algorithms,..., in short, almost everywhere! Moreover, a large number of cone orderings and other extensions of majorization have been found to be of interest and of wide applicability. A review will be presented of a necessarily selective sampling of statistical and probabilistic applications of majorization concepts that have appeared since the advent of the majorization "bible."

Multivariate Meta-analysis

◆ Betsy J. Becker, Michigan State University

Michigan State University, 462 Erickson Hall, East Lansing, MI 48864

bjbecker@msu.edu

Key Words: *meta-analysis, multivariate, correlation, regression*

Ingram Olkin is one of the key contributors to and founders of the field of statistical methods for research synthesis, also known as meta-analysis. In addition, Olkin and just a few others have done nearly all of the work to date on multivariate methods for meta-analysis. I will review some of the fundamental work done by Olkin, concerning multivariate effect-size data and multiple correlations (correlation matrices). Drawing on Olkin's work I will then present new analyses for multivariate synthesis of correlations and regression slopes. Tests of between-groups differences and "meta-regression" methods will be outlined. The indispensability of Olkin's prior work will be emphasized.

94 Statistics in Paleobiology and Paleoclimatology ◬

Section on Statistical Graphics, Section on Statistics and the Environment, Section on Physical and Engineering Sciences

Monday, August 9, 10:30 am–12:20 pm

Optimal Search for Undiscovered Fossil Species

◆ Woollcott Smith, Temple University

wksmith@unix.temple.edu

Key Words: *species diversity, cluster sampling, survey design*

Observed distributions of fossil finds and the observed diversity of fossil species within taxonomic groups can be used to design surveys aimed at increasing the expected number of new species found. This unusual design problem is addressed using classical stochastic models of species diversity and cluster sampling.

Bayesian Modeling of Mass Extinctions

◆ Steve C. Wang, Swarthmore College

Swarthmore College, Dept. of Mathematics and Statistics, 500 College Ave., Swarthmore, PA 19081

scwang@swarthmore.edu

Key Words: *paleontology, uniform distribution, Poisson process*

Mass extinctions have played a major role in the history of life. The end-Cretaceous extinction 65 million years ago, for instance, killed the dinosaurs and opened the way for many new mammal species. A much-debated question in paleontology is whether any particular mass extinction event was sudden or gradual. We address this question by presenting a Bayesian model for the mass extinction of a collection of species. We also calculate interval estimates of the duration of the extinction event. Such information provides critical evidence for inferring whether the extinction was sudden and caused by factors such as asteroid impact, or gradual and caused by factors such as climate change.

Statistics of Low-frequency Solar Variability in Paleoclimate and Model Data

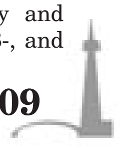
◆ Hee-Seok Oh, University of Alberta; Caspar Ammann, University Corporation for Atmospheric Research; Philippe Naveau, University of Colorado, Boulder

University of Alberta, Dept. of Math and Statistical Science, Edmonton, AB T6G 2G1 Canada

heeseok@ualberta.ca

Key Words: *solar activity, wavelet analysis, paleoclimate*

We present statistical analyses of proxies of different solar activity measures from the last millenium. Using nondecimated discrete wavelet transform, the spatial features of solar-modulated climate variations are extracted from proxy records and from two model experiments. We discuss important aspects of coherency and stationarity for the known solar variations at 11/22-, 80-88-, and



200-year cycles and quasi-cycles. Finally, we present a statistical framework which can help identify mechanisms and feedbacks involved in translating a solar-modulated perturbation into the Earth's climate system.

Statistical Issues in Reconstructing Past Climate Conditions

◆ Andrew R. Solow, Woods Hole Oceanographic Institution

Woods Hole Oceanographic Institution, Mailstop: 41, Woods Hole, MA 02543

asolow@whoi.edu

Key Words: *climate reconstruction, calibration, measurement error*

One of the central problems in paleoclimatology is reconstructing past climatic conditions. This paper discusses some statistical issues, focusing on the use of fossil beetle assemblages and isotopic measurements in corals.

95 Designed Missingness Strategies for Epidemiological Studies ▲ ⊞

General Methodology, Section on Survey Research Methods, Section on Statistics and the Environment, Section on Government Statistics, Section on Health Policy Statistics, Section on Statistics in Epidemiology, Social Statistics Section

Monday, August 9, 10:30 am–12:20 pm

Development of Cost-effective Statistical Sampling Strategies and Optimal Design Considerations for Exposure Assessment as Part of the National Children's Study

◆ Warren J. Strauss, Battelle Memorial Institute; Jeff Lehman, Statistics and Data Analysis Systems; Haluk Ozkaynak, U.S. Environmental Protection Agency

Battelle Memorial Institute, 505 King Ave., Columbus, OH 43201-2693

strauss@battelle.org

Key Words: *optimal design, measurement error, designed missingness, longitudinal, environmental exposure, replicate sampling*

This project focused on innovative statistical study design guidance for the acquisition of exposure data over time in a longitudinal study of children who participate in the National Children's Study (NCS). The NCS is intended to investigate environmental influences on children's health and development, including understanding any environmental exposures that may cause or exacerbate health impacts. Due to the large sample size and longitudinal nature of the NCS, unique statistical issues arise that must be addressed before a cost-effective sampling design can be developed to gather environmental and personal exposure data. A key issue for the NCS relates to obtaining enough samples to provide adequate statistical power to detect health effects attributable to exposure, while being cost-effective, minimizing participant burden, and staying within the study's budget. Relevant specific issues include identifying potential sources of bias and/or uncertainty in the exposure measures (nonresponse, subject burden, attrition, and measurement error); and strategies to address these issues, including sample weighting techniques, and replicate sampling to assess measurement error variance.

Outcome-dependent Study Design for Clustered Data

◆ Haibo Zhou, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Dept. of Biostatistics, Chapel Hill, NC NC 27599-7420

zhou@bios.unc.edu

Key Words: *outcome-dependent sampling, clustered data, random effects model, epidemiologic study*

We review recent development in efficient study design using outcome-dependent sampling techniques in epidemiologic studies. Such biased sampling schemes can enhance the cost-efficiency for a given study budget by using known outcome or auxiliary data information. Statistical issues and their applications will be discussed.

Split-questionnaire Designs for National Health Surveys

◆ Neal Thomas, Pfizer Inc.; Myron Katzoff, National Center for Health Statistics; Trivellore E. Raghunathan, University of Michigan; Nathaniel Schenker, National Center for Health Statistics

Pfizer Inc., 61 Dreamlake Dr., Madison, CT 06443

snthomas99@yahoo.com

Key Words: *split questionnaire design, matrix sampling, multiple imputation, NHANES*

The National Center for Health Statistics (NCHS) conducts nationally representative probability samples that include expensive and time-consuming clinical exams and labs. The amount of time and measurements these surveys require result in high costs and respondent burden. One potential solution to this problem is to split the questionnaire into several parts and administer a carefully selected subset of the questionnaire to each respondent. We develop a method for creating split-questionnaire forms that can be applied in complex settings which include the use of skip patterns. The method creates forms that include items that are good predictors of the excluded items so the subsequent analyses can recover much of the information about the excluded items that would have been collected in a complete survey. The form-design method is applied using pilot data from NHANES II to create forms for a simulation study based on the NHANES III sample. The simulation study demonstrates the feasibility of the approach applied to a real health status survey, and it produced practical advice about the appropriate items to include in split-questionnaire designs in future health surveys.

Optimal Design for Bivariate Longitudinal Data When Some of the Outcomes Will Be Missing by Design

◆ Cassandra Arroyo, Morehouse School of Medicine

Morehouse School of Medicine, Dept. of Medicine, Social Epidemiology Research Division, 720 Westview Dr., SW, NCPC Rm 313, Atlanta, GA 30310

carroyo@hsph.harvard.edu

Key Words: *latent variable, designed missingness*

Designed missingness arises often in many different biomedical applications, mostly in the form of ancillary validation substudies. While such designed missingness has been well studied for predictors, relatively little has been done when the designed missingness occurs in the outcome. This issue has been discussed in the sociology literature under the heading of matrix sampling, but has been restricted to the single group context. We propose a latent variable model for analyzing bivariate longitudinal data when one of the outcomes is missing by design. We then use the method to motivate a discussion of optimal design for studies where one of the outcomes is subject to designed missingness and offer practical advice for investigators to consider when faced with this issue.

96 Ensembles of Models in Data Mining

SSC, Section on Statistical Computing, Section on Statistical Graphics, Section on Physical and Engineering Sciences, Section on Quality and Productivity

Monday, August 9, 10:30 am–12:20 pm

An Out-of-bag Method for Regularizing Boosted Regression

◆ Greg Ridgeway, RAND Corporation

RAND Corporation, 1700 Main St., PO Box 2138, Santa Monica, CA 90407-2138

gregr@rand.org

Key Words: *nonparametric regression, boosting, out-of-bag estimation, regularization*

Selecting the optimal number of iterations is a key step when using any of the various flavors of boosting. The standard practice for estimating the optimal number of iterations is to leave out some fraction of the dataset as a test set and iterate the boosting algorithm until the predictive performance on the test set no longer improves. This practice, however, allocates a large part of the dataset for estimating the optimal number of iterations diluting the amount of information available for building the model structure. A variation on out-of-bag estimation can provide an approximately unbiased estimate of the improvement in generalization error attributable to the current iteration without decreasing the amount of data available for learning the model structure. When the out-of-bag estimate of improvement is zero, the iterations stop. The R package "gbm" implements this technique and I demonstrate on several datasets that this procedure offers a fully automated process for fitting boosting models, requiring fewer iterations and offering improved predictive performance.

Averaging Methods for High-throughput Screening Data

◆ Yuanyuan (Marcia) Wang, University of Waterloo

University of Waterloo, 200 Elm St., Apt. 706, Toronto, ON M5T 1K4 Canada

y32wang@uwaterloo.ca

Key Words: *HTS, averaging methods, KNN, CART, bias and variance trade-off*

In drug discovery, high-throughput screening (HTS) is used to assay large numbers of compounds against a biological target. A research pharmaceutical company might have of the order 2 million compounds available, and the human genome project is generating many new targets. Hence, there is a need for a more efficient strategy: smart or virtual screening. In smart screening, a representative sample (experimental design) of compounds is selected from a collection and assayed against a target. A model is built relating activity to explanatory variables describing compound structure. Our previous work shows that local methods, like K-nearest neighbors (KNN) and classification and regression trees (CART), perform very well. I will propose a method of averaging across multiple classifiers based on building classifiers on subspaces (subsets of variables). It improves the performance of KNN and CART for HTS data. Some interpretation of the method has also been considered.

97 Bayesian Methods In Genomics ▲ ⌘

WNAR, Section on Bayesian Statistical Science

Monday, August 9, 10:30 am–12:20 pm

Bayesian Variable Selection for Gene Expression Analysis

◆ Marina Vannucci, Texas A&M University

Texas A&M University, College Station, TX 77843-3143

mvannucci@stat.tamu.edu

Key Words: *Bayesian methods, microarrays, variable selection*

The analysis of the high-dimensional data generated by DNA microarrays poses challenge to standard statistical methods. This has revived a strong interest in classification and clustering algorithms. Currently, investigators first resort to data-filtering procedures or dimension reduction techniques before applying a statistical model. We propose the use of Bayesian variable selection techniques built directly into the statistical modeling. The flexibility of the priors allows us to incorporate biological information.

Multi-study Genomic Data Analysis

◆ Giovanni Parmigiani, Johns Hopkins University; Elizabeth Garrett, Johns Hopkins University; Xiaogang Zhong, Johns Hopkins University; Edward Gabrielson, Johns Hopkins University

Johns Hopkins University, 550 North Broadway, Suite 1103, Baltimore, MD 21205

gp@jhu.edu

Key Words: *genomics, meta-analysis, microarrays, breast cancer, lung cancer, validation*

Several recent studies sought to refine cancer classification using gene expression microarrays. Uncertainty remains regarding the extent to which these studies agree, and whether results can be integrated. We developed practical tools for cross-study comparison, validation, and integration of cancer molecular classification studies using public data. We show how to evaluate genes for cross-platform consistency of expression patterns, using "integrative correlations," which quantify cross-study reproducibility without relying on direct assimilation of expression measurements across platforms. We then compare associations of gene expression levels to differential diagnosis of squamous cell carcinoma versus adenocarcinoma, via reproducibility of the gene-specific t-ratios, and to survival, via reproducibility of Cox coefficients. We use this comparison as a testbed for developing simple approaches to the normalization of expression measurement across microarray platforms. Finally, we show preliminary progress on the cross-study validation of results of unsupervised cluster analyses.

Slippery Phenotypes: Survival Analysis Using Microarrays

◆ Marco Ramoni, Harvard Medical School; Paola Sebastiani, Boston University

Harvard Medical School, Harvard Partners Center for Genetics and Genomics, 77 Ave. Louis Pasteur, Boston, MA 02115

marco_ramoni@harvard.edu

Key Words: *microarrays, Bayesian statistics, survival analysis*

The development of microarray technology presents unprecedented opportunities to discover new aspects of human diseases and to deliver new drugs and new diagnostic methods. An emerging challenge for this endeavor is the characterization of clinical features—such as survival time—that do not easily fall into discrete categories. Traditional survival analysis methods have been sometimes successful in analyzing some survival genomic experiments, they are unable to account for biological and sample variability and rest on the unrealistic assumption that the expression of each gene is functionally independent of the expression of other genes. This talk describes a Bayesian approach to the joint analysis of gene expression data and survival time in microarray experiments that can overcome these limitations. The novelty of this approach is that it is able to integrate survival time and gene expression data while accounting for both the variability of gene expression and the functional dependencies among genes.

A Model of AFLP Evolution and Its Use in Bayesian Estimation of Phylogenetic Relationships

◆ Bret Larget, University of Wisconsin, Madison; Ruiyan Luo, University of Wisconsin, Madison

University of Wisconsin, Madison, Dept. of Statistics, 1210 West Dayton St., Madison, WI 53706

larget@stat.wisc.edu

Key Words: *phylogeny, MCMC, Bayesian statistics, evolution*

Amplified Fragment Length Polymorphism (AFLP) markers are a type of genetic information that are especially useful to plant biologists. We describe a Bayesian approach to modeling AFLP marker evolution that is informed by the underlying science. We describe an MCMC approach to the estimation of phylogeny from AFLP marker data and assessment of uncertainty. We apply the method to several datasets.

98 *JASA* Theory and Methods Invited Paper

JASA, Theory and Methods

Monday, August 9, 10:30 am–12:20 pm

Location-scale Depth

◆ Ivan Mizera, University of Alberta; Christine H. Mueller, Carl Von Ossietzky University of Oldenburg

University of Alberta, Dept. of Math & Statistical Sciences, Edmonton, AB T6G 2G1 Canada

mizera@stat.ualberta.ca

Key Words: *depth contours, exploratory data analysis, location-scale model, median, Mobius equivariance, robust estimation*

The paper introduces a half-space depth in the location-scale model, along the lines of the general theory given by Mizera on the basis of the idea by Rousseeuw and Hubert, complemented by a new likelihood-based principle for designing criterial functions. The most tractable version of the proposed depth, the Student depth, turns out to be nothing but the bivariate half-space depth interpreted in the Poincare plane model of the Lobachevski geometry. This fact implies many fortuitous theoretical and computational properties, in particular, equivariance with respect to the Mobius group and favorable time complexities of algorithms. It also opens a way to introduce some other depth notions in the location-scale context, for instance, location-scale simplicial depth. A maximum depth estimator of location and scale—the Student median—is introduced. Possible applications of the proposed concepts are investigated on data examples.

99 Causal Inference in Epidemiology ⌘

Section on Statistics in Epidemiology, Section on Health Policy Statistics

Monday, August 9, 10:30 am–12:20 pm

Causal Inference in Hybrid Intervention Trials Involving Treatment Choice

◆ Roderick J. Little, University of Michigan; Qi Long, University of Michigan; Xihong Lin, University of Michigan

University of Michigan, Dept. of Biostatistics, School of Public Health, 1420 Washington Heights, M4045, Ann Arbor, MI 48109

rlittle@umich.edu

Key Words: *clinical trials, doubly randomized preference trial, EM algorithm, partially randomized preference trial, randomization, selection bias*

Randomized allocation of treatments is a cornerstone of experimental design, but has drawbacks when a limited set of individuals are willing to be randomized, or the act of randomization undermines the success of the treatment. Choice-based experimental designs allow a subset of the participants to choose their treatments. We discuss causal inferences for experimental designs where some participants are randomly allocated to treatments and others receive their treatment preference. The paper was motivated by the "Women Take Pride" study, a doubly randomized preference trial to assess behavioral interventions for women with heart disease. We propose a model for inference about preference effects and develop an EM algorithm to compute maximum likelihood estimates of the model parameters, using data from the Women Take Pride study. We also expand these methods to handle a broader class of designs involving treatment preference, and discuss alternative designs from the perspective of the strength of assumptions needed to make causal inferences.

Causal Vaccine Effects on Binary Post-infection Outcomes

◆ Michael G. Hudgens, University of North Carolina, Chapel Hill; M. Elizabeth Halloran, Emory University

University of North Carolina, Chapel Hill, Dept. of Biostatistics, School of Public Health, 3107-E McGavran-Greenberg Hall, Chapel Hill, NC 27599

mhudgens@bios.unc.edu

Key Words: *causal inference, selection bias, vaccine*

The goal of prophylactic vaccination is to prevent infection or ameliorate disease or decrease infectiousness of individuals who become infected subsequent to vaccination. Thus, evaluation of several vaccine effects of interest condition on being infected. Standard approaches to assessing vaccine effects on the distribution of post-infection outcomes generally do not assess causal effects of vaccination, because comparison groups are selected post-treatment assignment. We consider estimation and testing of causal effects of vaccination on binary post-infection outcomes (denoted VEp) such as secondary transmission and severity of disease. We adopt the approach of Frangakis and Rubin (2002) and define causal effects within principal strata defined on the joint potential infection values under vaccine and placebo. While the causal VEp is generally not identifiable without unverifiable assumptions, using a maximum likelihood based approach we derive bounds on the causal estimands of interest. Sensitivity analysis and testing procedures are also developed. The methodology is demonstrated using data from field studies of a rotavirus vaccine candidate and a pertussis vaccine.

The Causal Effect of Postmenopausal Hormone Therapy on Coronary Heart Disease

◆ Miguel A. Hernan, Harvard School of Public Health; James Robins, Harvard School of Public Health

Harvard School of Public Health, Epidemiology Dept., 677 Huntington Ave., Boston, MA 02115

miguel_hernan@post.harvard.edu

Key Words: *causal inference, observational studies, survival analysis*

The Nurses' Health Study (NHS) found a lower risk of coronary heart disease among users of postmenopausal hormone therapy. The Women's Health Initiative (WHI), a randomized trial, reported the opposite finding. The following limitations of the NHS design may explain the discrepancy: (1) lack of comparability between women that did and did not initiate therapy, (2) lack of comparability between women who continued and discontinued therapy, and (3) inability to detect short-term effects of therapy. The NHS can be conceptualized as a randomized trial with unknown randomization probabilities. We describe a study protocol for the NHS trial and propose its re-analysis under the intention to treat principle, which is immune to bias due to explanation 2. We use different statistical models (Cox proportional hazards, accelerated failure time, risk ratio models) to reduce the possibility of bias due to idiosyncratic misspecification of any particular model. We propose a strategy to assess the influence of explanation 3. Finally, we describe analyses to assess the sensitivity of our causal estimates to bias from explanation 1.

100 Record Linkage and Unduplication Methodology and Results ◭ ⌘

Section on Survey Research Methods, Section on Government Statistics, Social Statistics Section

Monday, August 9, 10:30 am–12:20 pm

An Analysis of Person Duplication in Census 2000

◆ Robert E. Fay, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Rm. 3067-3, Washington, DC 20233-9001

Robert.E.Fay.III@census.gov

Key Words: *census duplication, erroneous enumeration, computer-matching*

Computer-matching on name and date of birth can be used to identify duplicate enumerations of persons in Census 2000. Computer-matching faces two important limitations. First, because name and date of birth are not always unique, computer-matching links some enumerations together that do not represent duplicated persons. Second, as a consequence of inaccurate or missing names or days of birth, computer-matching fails to identify some duplicates. Two previous papers presented probabilistic models to address the problem posed by coincidental sharing of date of birth by persons with the same name. The paper describes the application of the methods to Census 2000. Although estimates of duplication in Census 2000 are already available from a previous study, the earlier results are based on a sample whose size limits detailed analysis. Instead, the paper reports results from the full census. In addition to applying the probabilistic models described previously, the paper investigates the degree of improvement in matching resulting from a series of edits of the reported names developed by other U.S. Census Bureau researchers.

Administrative Records and Person Duplication in Census

◆ Donald M. Bauder, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 9200, Washington, DC 20233-9200

donald.m.bauder@census.gov

Key Words: *record linkage, administrative records*

This study uses administrative records to study person duplication in the census. For previous projects, Census Bureau staff computer-linked person records on the Hundred Percent Census Unedited File (HCUF) to records on an administrative records file, the Census Numident. Two HCUF records that are linked with one Numident record are evidence that a person is duplicated in the HCUF. We compare the results of this match with results from another duplication study, Further Study of Census Duplication (FSPD). We also use results from a clerical review of a sample of links from the HCUF to Census Numident match, and of FSPD links. We address these questions: ow well were the FSPD results confirmed by the HCUF to Census Numident match results? What is the potential for future uses of administrative records in duplication research and unduplication efforts? Our results include the following: Administrative records generally confirmed the FSPD processes; administrative records can be a valuable tool for confirming and disconfirming potential duplicates; and that administrative records can be useful for identifying duplicates missed by other processes.

A Record Linkage Strategy to Match New Source Lists

◆ Thomas M. Pordugal, National Agricultural Statistics Service

National Agricultural Statistics Service, 1400 Independence Ave., SW, Room 6344-A South Bldg., Washington, DC 20250-2007

tpordugal@nass.usda.gov

Key Words: *record linkage, List Frame, SuperStan, SuperMatch*

The National Agricultural Statistics Service (NASS) maintains a List Frame of all known and potential farm operators and agribusinesses. To keep the List Frame as complete and up-to-date as possible, the NASS continually receives new data lists from a variety of sources. These new lists are matched against the existing List Frame using probabilistic record linkage methodology. The new lists often contain different types and quality of information requiring different matching strategies. At a minimum, the lists contain name and address information. Different matching strategies, which include the blocking of variables within a pass, the matching variables used within a pass, and the determination of the number of passes, are required for different new lists. This paper presents an overview of the steps taken to match these new source lists to the NASS List Frame, guidelines for developing matching strategies, and a description of the record linkage software features used to process the data.

The Bigmatch Program for Record Linkage

◆ William E. Yancey, U.S. Census Bureau

U.S. Census Bureau, 4401 Suitland Rd., Room 3232-4, Suitland, MD 20746-4700

william.e.yancey@census.gov

Key Words: *record linkage, matching, blocking, unduplication*

The Bigmatch program is designed for large-scale record linkage projects. The program handles several blocking strategies at a time by performing key sorts in core memory. It can process large numbers of records pairs very quickly. It can be used for matching two files or for unduplicating one file. We discuss the program, its operating features, and some application results.

A Bayesian Record Linkage Methodology for Multiple Imputation of Missing Links

◆ Michael H. McGlincy, Strategic Matching, Inc.

Strategic Matching, Inc., PO Box 334, Morrisonville, NY 12962

mcglincym@strategicmatching.com

Key Words: *probabilistic, record, linkage, Bayesian, imputation*

Probabilistic record linkage can be an effective research technique even if available records lack strong personal identifiers or if identifying fields contain many errors or omissions. Traditional methodologies typically select a single set of linked record pairs for research based on a match weight test statistic and clerical review of marginal pairs. However, false positives links and false negative links can make such datasets unrepresentative of the total population of true linked pairs. The methodology described here addresses this problem. First, a full Bayesian model is developed for the posterior probability that a record pair is a true match given observed agreements and disagreements of comparison fields. Second, observed-data posterior distributions for model parameters and true match status are estimated simultaneously through MCMC data augmentation with multiple chains. This gives multiple complete and unbiased sets of imputed linked record pairs. Finally, population estimates are obtained from each imputation and consolidated using established techniques. Application of the methodology by a consortium of traffic safety researchers is described.

101 Statistical Methods for Assessing Disease Improvement in Chronic Disease

Section on Health Policy Statistics, WNAR

Monday, August 9, 10:30 am–12:20 pm

Outcome Measures in Pediatric Rheumatic Disease

◆ Brian M. Feldman, WNAR

WNAR, Hospital for Sick Children, 555 University Ave., Toronto, ON M5M 1M5 Canada

brian.feldman@sickkids.ca

Key Words: *health-related quality of life, pediatric, arthritis, physical function, outcome measurement*

Subjective, questionnaire-based outcome measures are considered important in rheumatology as they allow the patient's view to be incorporated. Several physical function and quality-of-life (HRQL) measures have been adapted, or developed for children with rheumatic illness. The current measures have several limitations: some require proxy respondents, some are poorly sensitive to change, and the clinical meaning of most is not really understood. Recent work has tried to remedy some of these limitations. We have recently determined the clinical meaning of scores on the Childhood Health Assessment Questionnaire (CHAQ) - e.g. a score of 0.125 means mild disability while 1.75 means moderate. Moreover, we have developed a modified version of the CHAQ that discriminates better between children with arthritis and healthy controls (relative sensitivity 2.32). We have also shown that children and their families differentiate HRQL, health status, and overall quality of life. Finally, we have shown that utility may be used to measure HRQL, but that the time trade-off, and multi-attribute methods may be more valid and understandable than the standard gamble.

Classifying Radiographic Progression Status in Early Rheumatoid Arthritis Patients

◆ Grace S. Park, University of California, Los Angeles; Weng Kee Wong, University of California, Los Angeles; Myungshin Oh, University of California, Los Angeles; John T. Sharp, University of Washington; Richard H. Gold, University of California, Los Angeles; Ken J. Bulpitt, Kaiser Permanente, West LA; Harold E. Paulus, University of California, Los Angeles

University of California, Los Angeles, Dept. of Biostatistics, School of Public Health, Box 951772, Los Angeles, CA 90095-1772

gspark@mednet.ucla.edu

Key Words: *joint damage, early rheumatoid arthritis, composite definition, progression vs. nonprogression*

Various methods are used to measure radiographic joint damage in rheumatoid arthritis (RA), but determining proportions of responsive individuals is difficult. We discuss different ways to classify radiographic damage in individuals with RA as progressive or nonprogressive. We assessed damage by total Sharp score (TSS), erosion score (ES), and joint space narrowing score (JSNS) for 751 serial hand, wrist, and foot films of 190 early RA patients from the Western Consortium study followed 6-60 months (mean 31 months). The damage progression rate was estimated for each patient. Within- and between-individual variance, measurement error, and different approaches to classifying radiographic progression status were examined. The selected index defines nonprogression as an increase of 0.1*trimmed SRM (standardized response mean for the central 95% of the sample) for five of six measures (ES and JSNS for hands, wrists, feet). 59% of 190 were categorized as nonprogressors, and had significantly lower disease activity and more clinical improvement than progressors. We conclude that joint damage in RA patients can be classified as progressive or nonprogressive using a composite definition.

Developing Measures of Approval for Myositis

◆ Peter A. Lachenbruch, U.S. Food and Drug Administration; Lisa Rider, National Institutes of Health; Frederick Miller, National Institutes of Health

U.S. Food and Drug Administration, 1401 Rockville Pike, HFM-215, Division of Biostatistics, Rockville, MD 20852

lachenbruch@cber.fda.gov

Key Words: *myositis, index development, CART*

The myositis syndromes are a group of autoimmune diseases characterized by chronic inflammation of muscle resulting in muscle weakness. We obtained consensus rating of improvement from a panel of clinicians. They obtained consensus in 204 of 222 patients. We then selected the variables to develop the index based on the clinician's evaluation of the variable's usefulness. The clinicians ranked the variables on their importance in defining improvement, the number of the variables required for improvement, the amount of improvement in each variable that was needed, and how much worsening could be permitted. Six candidate variables were selected. Third, we developed candidate rules. These were similar to indices used in rheumatoid arthritis and lupus or from exploratory statistical analyses of the data. Fourth, we examined the sensitivity, specificity, positive predictive value, negative predictive value and Area Under the Curve (AUC) of the rules. Fifth, a panel of experts in the field selected promising candidate rules for improvement from a set of candidates that excluded poorly performing rules.

Mean Changes vs. Dichotomous Definitions of Improvement

◆ Jennifer J. Anderson, Boston University

Boston University, 38 Harvard St., Arlington, MA 02476

jja@bu.edu

Key Words: *improvement criteria, dichotomy, rheumatoid arthritis, ankylosing spondylitis, simulation, multivariate normal*

In several areas within rheumatology dichotomous definitions of improvement are in use in the reporting of clinical trials, but it is to be expected that continuous definitions would offer improved discrimination between treatment groups. Nevertheless, a well-constructed dichotomous outcome (usually a composite) has advantages of clinical sense and specificity and may, under a variety of realistic conditions, have power that closely approximates that of standard continuous outcome measures. This has been seen for established dichotomous outcome definitions for two rheumatological conditions, rheumatoid arthritis (RA) and ankylosing spondylitis (AS). Simulation studies have been performed using multivariate normal generated data that approximates actual trial data for each of RA and AS patients. They demonstrate relative power of several dichotomous and continuous outcomes in realistic situations for each of RA and AS. These results will be presented and discussed.

Assessing Disease Progression Using a Composite Endpoint

◆ Weng Kee Wong, University of California, Los Angeles; Philip J. Clements, University of California, Los Angeles; Jim Streisand, Genzyme Corporation; Dan E. Furst, University of California, Los Angeles

University of California, Los Angeles, Dept. of Biostatistics, School of Public Health, 10833 Le Conte Ave., Los Angeles, CA 90095-1772

wkwong@ucla.edu

Key Words: *multiple outcomes, Kaplan Meier, Scleroderma, chronic disease*

We investigate the usefulness of using a composite "time-to-event" analysis to evaluate disease progression in 134 patients with diffuse cutaneous Scleroderma. All patients had 18 months or less of disease duration at study entry and followed for 24 months. The composite time-to-event analysis requires a composite endpoint with a Kaplan-Meier-type analysis. Serious medical events across several organ systems would make up the composite index and we consider the endpoint is attained if a patient reaches any one of the endpoints within the composite. For example, one may define the endpoint is attained if a patient experienced any one of the following in a given time period: (i) persistent high skin score, (ii) drop in predicted DLCO by 15%, or (iii) renal involvement or (iv) heart involvement. In addition, we present results on our sensitivity analysis where one or more of the organ involvement definition criteria are adjusted and the outcomes reanalyzed. One key finding is that the usefulness of such composite endpoint depends crucially on the choice of the individual endpoints and the sensitivity of that measure.

102 Latent Variable Model Applications in Health Sciences

Section on Health Policy Statistics, ENAR

Monday, August 9, 10:30 am–12:20 pm

Latent Class Pattern Mixture Models for Longitudinal Data with Dropout

◆ Jason Roy, University of Rochester

University of Rochester, Rochester, NY 14642

jason_roy@urmc.rochester.edu

Key Words: *incomplete data, latent variable*

Pattern mixture models are a popular approach for analyzing incomplete longitudinal data. The typical approach is to stratify the response model on dropout time. However, this approach may lead to sparse cells. An alternative is to assume the existence of a small number of dropout classes. Dropout classes are unobserved, but assumed to be related to dropout time. The response model is then stratified over classes. We propose a general model that is suitable for both continuous and discrete response. We consider both maximum likelihood and Bayesian methods for inference.

Bayesian Competing Factor Models for Bidirectional Latent Predictors

◆ Amy H. Herring, University of North Carolina, Chapel Hill; David B. Dunson, National Institute of Environmental Health Sciences; Nancy Dole, Carolina Population Center

University of North Carolina, Chapel Hill, Dept. of Biostatistics, CB #7420, Chapel Hill, NC 27599

aherring@bios.unc.edu

Key Words: *Bayes, categorical data, discrete choice model, joint modeling, latent variables, stress*

Researchers often rely on questionnaire data to study the health effects of stress. A subject indicates the occurrence of potentially stressful events and quantifies the strength of reaction to these events, ranging from strongly negative to strongly positive. These data are used to obtain measures of levels of negatively and positively perceived stress. Motivated by such data, we propose a latent variable model characterized by event-specific negative and positive reaction scores. If the positive reaction score dominates the negative reaction score for an event, the individual's reported response to that event will be positive, with an ordinal ranking determined by the value of the score. Measures of overall positive and negative stress are obtained by summing the reaction scores across the events reported by an individual. By incorporating these measures as predictors in a regression model and fitting the stress and outcome models jointly using Bayesian methods, inferences are conducted without assuming known weights for the different events. We propose an MCMC algorithm for posterior computation and apply the approach to study the effects of stress on pre-term delivery.

On the Study of Hypothetical Health Constructs

◆ Karen Bandeen-Roche, Johns Hopkins University

Johns Hopkins University, Dept. of Biostatistics, 615 N. Wolfe St., Baltimore, MD 21205

kbandeen@jhsph.edu

Key Words: *latent variable, measurement, pseudo-values, validity, aging, Monte Carlo*

In a number of fields, there exist important adverse health states that are viewed as widely recognizable by physicians but lack consensus conceptualization or modes of measurement. Two examples are "frailty" in geriatrics and "abnormal personality" in psychology. Advancing related research requires a combination of testing theories on the nature of constructs at issue and delineation of the operational characteristics of various proposals for their surrogate measurement. This talk discusses the formulation of such theories within latent variable models, proposes a strategy for evaluating internal and external validity, and applies the methodology to the frailty measurement problem. The validity evaluation strategy relies on randomized generation of underlying health state values a posteriori, increasingly termed "pseudo-values," and subsequent inference utilizing those values while accounting for the model-fitting process. The work ultimately aims to equip researchers with surrogate measures of hypothetical health constructs that more precisely operationalize clinical impression than do the ad hoc surrogates that predominate in extant research.

Latent Variable Modeling with Continuous and Classification Latent Variables

◆ Yasuo Amemiya, IBM T.J. Watson Research Center; Melanie M. Wall, University of Minnesota

IBM T.J. Watson Research Center, Rm. 33-222 1101 Kitchawan Rd., Route 134, Yorktown Heights, NY 10598

yasuo@us.ibm.com

Key Words: *latent class model, structural equation analysis, categorical response, latent health condition, psychological construct*

There are two types of latent variable modeling often used in health sciences research; structural equation modeling with continuous factors, and latent class analysis with unordered-categorical latent segments. This paper develops a statistical methodology for a more

general model with both continuous and categorical latent variables. Observed measurement types are allowed to include continuous and ordered categorical responses. Model-fitting methods and associated statistical inference procedures are discussed. This methodology can be useful in health science applications, where a health condition outcome variable is a latent classification, and some of possible predictor variables are psychological/behavioral constructs. An example relating a underlying eating disorder condition to a physical appearance satisfaction construct is presented.

Composite Measurement Summaries from Data of Mixed Types

◆ Mary D. Sammel, University of Pennsylvania; J. Richard Landis, University of Pennsylvania School of Medicine

University of Pennsylvania, Dept. of Biostatistics and Epidemiology, 432 Guardian Dr., 605 Blockley Hall, Philadelphia, PA 19104-6021

msammel@cceb.upenn.edu

Key Words: *factor analysis, item response theory, factor scores*

We utilize factor scores from a latent variable model to define a composite measure of a subject's symptoms evaluated at a particular point in time. Symptom measures can be a combination of data from different types (continuous, dichotomous, or likert/ordinal). The performance of the derived score is evaluated using classical test theory (CTT). Advantages of this method include: (1) appropriate modeling of data in the format in which it was collected, and (2) efficient estimation of the impact of subject covariates on the summary score. Item screening and reduction are discussed. Item characteristic curves, which quantify the amount of information contributed by each item, are defined for the various types of data. We illustrate our method using data from the NIH-funded national Interstitial Cystitis Data Base (ICDB) Study, where we will compare the subjects' overall assessment of their IC pain.

103 The Human Faces of Statistics ⊞

Social Statistics Section

Monday, August 9, 10:30 am–12:20 pm

Using the 2000 Census to Estimate the Size and Characteristics of Hispanic Subgroups in New York

◆ Arun P. Lobo, New York City Dept. of City Planning; Joseph J. Salvo, NYC Dept. of City Planning

New York City Dept. of City Planning, 22 Reade St. 4W, New York, NY 10007

plobo@planning.nyc.gov

Key Words: *Hispanic origin, Hispanic subgroups, nonsampling error, 2000 Census, census tracts, census questionnaire*

We begin by describing the enumeration of a large number of "unspecified Hispanics" and its impact on the count and characteristics of the major Hispanic subgroups in New York City. The large increase in the number of unspecified Hispanics was an artifact of question wording, rather than the result of a pan-ethnic Hispanic identity. We describe how the lower count of Hispanic subgroups affects subgroup characteristics and use of the data for City Planning purposes. Finally, we examine ways in which the count of Hispanic subgroups can be increased to better reflect their actual populations.

The Cross-cultural Legacy of Generation X: An Evaluation of Contemporary Civic and Government Engagement, Social Trust and the Decennial Census

◆ Melinda Crowley, U.S. Census Bureau

U.S. Census Bureau, 3311 Belleview Ave., Cheverly, MD 20785-1229

melinda.crowley@census.gov

Key Words: *decennial census, Generation X, civic engagement, government engagement, undercoverage issues*

The premise of this research is that response to surveys is motivated by a respondent's sense of civic responsibility. Past research shows that a primary reason respondents participate in government-sponsored survey requests is that they feel it is their civic duty to do so as good citizens. The main goal of this study is to investigate shared attitudes among Generation X (1968-1979) about civic engagement, government in general and decennial census participation in particular. According to past studies, GenX tends to be apathetic about community and political involvement and disillusioned with government. If GenX respondents in this study share such attitudes as their wider GenX counterparts do, then the Census Bureau will face obstacles reaching out to them in 2010. Disillusionment with the government will also compound existing enumeration barriers already identified and may have short and long-term implications for survey nonresponse issues, decennial undercoverage challenges, privacy and confidentiality concerns and decennial outreach campaigns. This presentation will also cover the beliefs GenXers have about the government.

Qui Bono? Winners and Losers in the California Public Higher Education System

◆ Walter R. Allen, University of California, Los Angeles

University of California, Los Angeles, Dept. of Sociology, 264 Haines Hall, Los Angeles, CA 90095-1551

wallen@ucla.edu

Key Words: *educational inequality, California high school students, race, class, gender disparities in educational achievement*

Why and how are statistics used so widely to explain educational inequities across racial groups? To what extent do conventional statistical approaches reproduce and validate race/ethnic/class/gender disparities in educational achievement? Given the pervasive "race blindness" in much of this literature, can statistics allow us to explore how deeply race affects educational processes and simultaneously lead us to do research that puts the humanity of individuals and groups at the forefront? Our approach is informed by empirical research on educational inequality in California that uses quantitative and qualitative data to illustrate the importance of codependent causal factors. After offering a critique of how statistics in key studies on educational inequality have been utilized in dehumanizing ways, we posit alternative theoretical explanations and challenge orthodox methods of examining educational experiences and outcomes. We then provide an in-depth case study of one anti-racist project that uses statistics as a central way of examining educational success and aspirations of high school students (CHOICES).

104 Statistics and Evolutionary Genetics ▲ ⊞

Biometrics Section

Monday, August 9, 10:30 am–12:20 pm

Maximum Likelihood Inference of Species Trees under the Coalescent Process

◆ James H. Degnan, University of New Mexico

University of New Mexico, Dept. of Mathematics and Statistics, Albuquerque, NM 87131

james@stat.unm.edu

Key Words: *phylogeny, gene tree, coalescence, maximum likelihood*

An evolutionary tree, or phylogeny, which reflects the evolutionary history of a set of species, might differ from the trees that reflect the evolutionary history of the genes from the same set of species. A tree reflecting the history of the species is called a species tree; a tree relfecting the history of a gene is called a gene tree. The goal of this talk is to implement maximum likelihood inference of the species tree from a set of (known) gene trees, assuming the coalescent process. This requires having both the likelihood of the gene tree and the probability of the gene tree given the species tree. Although methods for computing the likelihood of gene trees have been available for many years, until recently the probability of a gene tree for a given species tree has only been worked out for small trees—five or fewer species. This talk includes a method for computing this probability for an arbitrary number of species and the application of this method to the maximum likelihood inference of species trees.

Using Bayesian Networks for Paternity Calculations: Adding an Evolutionary Perspective

◆ Amanda Hepler, North Carolina State University

North Carolina State University, Bioinformatics Research Center, Campus Box 7566, Raleigh, NC 27695-7566

abhepler@stat.ncsu.edu

Key Words: *paternity index, Bayesian networks, forensic science, evolution, evidence interpretation, subpopulation effects*

This study expands the current use of Bayesian networks in the forensic setting by incorporating evolutionary effects into paternity calculations. Bayesian networks are gaining popularity as a graphical tool to communicate complex probabilistic reasoning required in the evaluation of DNA evidence. Features of HUGIN (a software package used to create Bayesian networks) are demonstrated that have not, as yet, been explored. These features greatly simplify the process of building Bayesian networks, allowing researchers to use these networks to solve new, more complex problems. Due to the increasing use of DNA evidence in courtrooms, and in light of recent studies on the potential impacts of ignoring evolution, this study is a natural extension to the body of research that already exists on Bayesian networks. We explore three paternity examples, a simple case with two alleles, a simple case with multiple alleles, and a missing father case. Networks are built for each example which incorporate the effects of evolutionary relatedness. We then compare these new networks to previous networks.

Revealing Uncertainty in Phylogenetic Trees

◆ Edward M. Johnston, Consultant

Consultant, 274 Clarendon St. #5, Boston, MA 02116

e.johnston@verizon.net

Key Words: *phylogenetic trees, bioinformatics, graphics, evolution*

Algorithms are widely available for creating phylogenetic trees. In many cases the available data are compatible with more than one tree, and there is no method of presenting this range of solutions that meets with complete satisfaction. Authors may start from the same data but arrive at different trees, so the question of objectivity may arise. Tree development is a hard problem with parallels in other areas of statistics. For example, when a statistician fits a linear model to a set of data, there is usually an opportunity to measure the success of the fit, and present it to the reader either numerically or by showing confidence limits in a diagram. Edward Tufte has proposed that diagrams should be information-rich and give readers the chance to judge the phenomena for themselves. He has observed that phylogenetic trees do not clearly separate evidence and assumptions. This paper attempts to satisfy Tufte with an improved tree diagram that allows the reader to assess the fit between hypothesis and reality.

Model Selection in Mixture Tree Analysis with Application to Genome Sequence Analysis

◆ Shu-Chuan Chen, Arizona State University; Bruce G. Lindsay, Pennsylvania State University

Arizona State University, Dept. of Mathematics and Statistics, Tempe, AZ 85287

shu@math.asu.edu

Key Words: *ancestral mixture model, mixture tree, single nucleotide polymorphisms*

Clustering methods have been broadly investigated in the last decade. Since the rapid progress of human genome sequencing, more efficient clustering methods are highly demanded. We will first show how an ancestral mixture model can be used to build up a hierarchical tree from binary sequence data, using as an example genetic single nucleotide polymorphisms (SNP) data. Some properties of the ancestral mixture model, such as its nested structure and the relationship to the coalescent process of population genetics, will be presented. A model selection method based on an easy-to-calculate quadratic distance will then be proposed. This distance arises by first applying kernel smoothing to both the data and the fitted model to get densities e* and a* on the sequence space. Then one uses the L2 distance between these to assess the fit of the data to the model. Some asymptotic properties of the quadratic distance will be discussed. An example of SNP data will be presented to demonstrate how our method works.

105 Analyses of Call History to Improve Telephone Surveys—Survival Analysis Applications ▲ ⌘

Section on Survey Research Methods, Section on Health Policy Statistics

Monday, August 9, 10:30 am–12:20 pm

Evaluating the Survival Analysis Method for Estimating Residency Rates

◆ Barbara L. Carlson, Mathematica Policy Research, Inc.; Daniel Kasprzyk, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., PO Box 2393, Princeton, NJ 08543-2393

bcarlson@mathematica-mpr.com

Key Words: *RDD, CASRO, response rate, AAPOR, Community Tracking Study*

To calculate a response rate for a survey, each sample member must first be classified according to its eligibility status. For telephone surveys with random-digit-dial samples, we try to determine for each sampled telephone number whether it is in service and belongs to a residence. The residency status of a certain portion of the sample, however, is never resolved even after making many call attempts. Several methods have been proposed to estimate residency among unresolved telephone numbers. The most conservative of these assumes the same residency rate among the unresolved cases as was found for the resolved cases. This inevitably overestimates the number of eligible telephone numbers and unduly reduces the response rate. A newer approach proposed in the literature tries to estimate a residency rate by modeling the "time to resolution" using survival analysis. The numbers designated as residential, nonresidential, and unresolved at each attempt are incorporated into the model. We apply a variation of this methodology to the HSC Community Tracking Study Household Survey and report on how it compared to the more conservative method of estimating residency.

Toward a Better Estimation of Working Residential Number Rate among the Undetermined: The Application of Survival Analysis in WRN Estimation

◆ Lidan Luo, NORC, University of Chicago; Hiroaki Minato, NORC, University of Chicago

NORC, University of Chicago, 55 East Monroe St., Suite 4822, Chicago, IL 60603

luo-lidan@norc.net

Key Words: *survival analysis, unit response rate, working residential number, REACH 2010 Risk Factor Survey, call history, CASRO method*

The AAPOR standard on response rate calculations permits some latitude in the assumptions made regarding the eligibility rate among cases for whom eligibility could not be determined. The CASRO method, for example, assumes the cases of unknown eligibility are eligible (WRNs) in the same proportion as cases for which WRN status can be determined. However, if the unknown cases are WRNs at a different rate, such a method might yield too inaccurate a response rate. The survival analysis approach described by Brick et al. (2002) takes advantage of the relationship between the level of difficulty (the number of calls) in reaching a household and the WRN rate. We demonstrate the method of survival analysis in estimating the WRN rate among the undetermined by simulations and a post-survey application to REACH Risk Factor Survey. Our results show that when more determined numbers are needed after the planned number of calls have been made, the survival analysis method seems always to give at least as good an estimate of the WRN rate among the undetermined as the traditional CASRO method does.

Estimating the Working Number Rate for a Cellular Telephone Survey

◆ Charlotte Steeh, Georgia State University; Zhenhua Hu, Georgia State University

Georgia State University, 140 Decatur St., 1042 Urban Life, Atlanta, GA 30303

cgsteeh@gsu.edu

Key Words: *cellular, RDD, working number rate, response rate, survival analysis*

One major difficulty of using a cellular telephone number sampling frame for a survey of the general population is the reduced ability to identify nonworking numbers. Since no comprehensive list of cellular telephone numbers exists to serve as a source of outside information and since cellular telephones are used in quite different ways than fixed-line telephones—turned off over long periods of time, e.g., the usual procedures for determining working numbers in conventional telephone surveys do not apply to surveys using a cellular frame. We describe the use of survival analysis to estimate the working number rate, revise the method proposed by Brick et al. (2002) to take account of conditional probabilities, compare our proposed method with other alternatives, and apply the procedure to the call histories of two national surveys conducted in 2003, one using cellular numbers and the other using fixed-line numbers. We believe that our procedure provides the best estimate of the working number rate in both types of surveys.

106 Bayesian Computation and Modeling

Section on Bayesian Statistical Science

Monday, August 9, 10:30 am–12:20 pm

Nonparametric Bayesian Density Estimation with Adaptive Autoregressive Priors

◆ Suhwon Lee, University of Missouri, Columbia; Paul L. Speckman, University of Missouri, Columbia; Dongchu Sun, University of Missouri, Columbia

University of Missouri, Columbia, 146 Middlebush Hall, Columbia, MO 65201

swlee2@stat.missouri.edu

Key Words: *Bayes, nonparametric density estimation, intrinsic autoregressive process, adaptive prior*

A simple model for Bayesian density estimation is introduced. Data are first binned, and a hierarchical model is used on the log of the bin probabilities. To implement a Bayesian locally adaptive method, we introduce a new class of spatially adaptive priors. we allow the variance components in the IAR priors to be nonconstant by taking an IAR prior for the log variance terms. Bayesian computations are implemented with MCMC. Applications are shown for real and simulated datasets.

Benchmark Estimation—Theoretical Results and Applications

◆ Subharup Guha, The Ohio State University; Steven MacEachern, The Ohio State University; Mario Peruggia, The Ohio State University

The Ohio State University, 1958 Neil Ave., Cockins Hall, Room 404, Columbus, OH 43210

subha@stat.ohio-state.edu

Key Words: *benchmark estimation, post-stratification, variance reduction*

We illustrate the technique of benchmark estimation using an example from the literature. For a subsampling rate of 1-in-50, the observed reductions in MSE often exceed 30% for a number of posterior features. Much larger gains are expected for certain complex estimation methods and for the commonly used thinner subsampling rates. Benchmark estimation can be used wherever other fast or efficient estimation strategies already exist. We discuss some asymptotic properties of benchmark estimators that provide an insight into the gains associated with the technique. The observed gains are found to closely match the theoretical values predicted by the asymptotic, even for k as small as 10.

Benchmark Estimation—Motivation and Basics

◆ Steven MacEachern, The Ohio State University; Subharup Guha, The Ohio State University; Mario Peruggia, The Ohio State University

The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210

snm@stat.ohio-state.edu

Key Words: *MCMC, variance reduction*

While studying various features of the posterior distribution of a vector-valued parameter using an MCMC sample, systematically subsampling the MCMC output can only lead to poorer estimation. evertheless, a 1-in-k subsample is often all that is retained in investigations where intensive computations are involved or where speed is essential. In these computationally intensive settings, we seek to create a discrete representation of the posterior distribution which is superior to the empirical distribution of a random sample of points drawn from the posterior. Benchmark estimation is one technique which yields this better representation. It relies on a number of estimates that are based on the best available information (the entire MCMC sample), and uses these to improve other estimates made on the basis of the subsample. The approach is implemented by creating weights for the subsample which are quickly obtained as the solution to a system of linear equations. Benchmark estimation leads to substantial gains at a negligible computational cost.

Intuitive, Easily Programmed Steps that Reduce MCMC Autocorrelation

◆ Todd L. Graves, Los Alamos National Laboratory; Paul L. Speckman, University of Missouri, Columbia; Dongchu Sun, University of Missouri, Columbia

Los Alamos National Laboratory, MS F600, PO Box 1663, Los Alamos, NM 87545

tgraves@lanl.gov

Key Words: *Metropolis algorithm, YADAS, Bayesian computation*

Often, an analyst writes software to analyze a Bayesian model using Markov chain Monte Carlo, only to find that high correlations between parameters cause mixing to be so poor as to render the algorithm useless. We will demonstrate how augmenting existing algorithms with Metropolis(-Hastings) steps in well-chosen directions can dramatically improve mixing. These directions can usually be found through easy analysis of individual terms in the posterior distribution. The MCMC software package YADAS makes it particularly convenient to create algorithms featuring these steps.

Characterizing Autocorrelation and Decorrelating Steps in MCMC Algorithms for Linear Models

◆ Paul L. Speckman, University of Missouri, Columbia; Todd L. Graves, Los Alamos National Laboratory; Dongchu Sun, University of Missouri, Columbia

University of Missouri, Columbia, Dept. of Statistics, Columbia, MO 65211

speckman@stat.missouri.edu

Key Words: *MCMC, mixing*

Markov chain Monte Carlo algorithms for Bayesian analysis often suffer from poor mixing. This poor mixing is generally associated with high posterior correlations between parameters. In the case of Gaussian linear models with known variances, we characterize which functions of parameters mix poorly and demonstrate a practical way of removing all autocorrelations from the algorithms by adding extra Gibbs steps in suitable directions.

107 Small-area Estimation II ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Monday, August 9, 10:30 am–12:20 pm

A Bayesian Approach for Combining Information from Multiple Surveys in Small-area Estimation Using Public-use Data

◆ Dawei Xie, University of Michigan; Trivellore E. Raghunathan, University of Michigan

University of Michigan, 2151 Hubbard, #11, Ann Arbor, MI 48105

xiedawei@umich.edu

Key Words: *posterior predictive distribution, constrained estimates*

Cancer surveillance research requires accurate county level prevalence rates of screening and risk factors. Raghunathan et al. (2003) developed a hierarchical Bayesian approach for obtaining such estimates by combining information from in-house data from the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS). Due to confidentiality concerns, however, it is not clear whether these model-based estimates obtained using in-house data can be released to researchers for use in their research. We develop estimates at the MSA level using publicly available data and then use the estimated model, the county level covariates and the direct estimates from BRFSS to obtain county level estimates. A Markov chain Monte Carlo (MCMC) method is used to generate the joint posterior predictive distribution of the county level unknown quantities. Yearly county level estimates for 49 states, District of Columbia and the whole state of Alaska in 1997-2000 are developed and compared to the HB model estimates from the in-house NHIS and BRFSS data.

Two-stage Nonparametric Approach for Small-area Estimation

◆ Pushpal Mukhopadhyay, Iowa State University; Tapabrata Maiti, Iowa State University

Iowa State University, 204, Snedecor Hall, Ames, IA 50010

pushpal@iastate.edu

Key Words: *nonparametric mixed model, Nadaraya-Watson kernel estimate, mean square prediction error*

Small-area estimators commonly borrow strength from other related areas. These indirect estimators use models (explicit or implicit) that relate the small areas through supplementary data. Various unit-level and area-level small-area models are proposed in the literature, but all these models assume the small-area mean is linearly related with supplementary information. We propose an area-level nonparametric regression estimator based on Nadaraya-Watson kernel on small-area mean. We adopt a two-stage estimation approach proposed by Prasad and Rao (1990). The asymptotic properties of the proposed estimator are studied and a second order approximation to the mean squared prediction error (MSPE) of the two-stage estimator and the estimator of MSPE approximation are obtained under normality. Finally, we perform a simulation study to show the superiority of the proposed estimator.

Empirical Bayes Estimation Smoothing of Relative Risks in Disease-mapping

◆ Jane L. Meza, University of Nebraska Medical Center

University of Nebraska Medical Center, 984350 Nebraska Medical Center, Omaha, NE 68198-4350

jmeza@unmc.edu

Key Words: *empirical Bayes, standardized mortality ratio, disease-mapping*

A common index of disease incidence and mortality is the standardized mortality ratio (SMR). The SMR is a reliable measure of relative risk for large geographical regions such as countries or states, but may be unreliable for small areas such as counties. A multilevel Poisson model with covariates will be used to produce smoothed estimates of the small-area-level SMR. A hybrid of parametric bootstrap and delta methods is used to estimate the MSE. The proposed measure captures all sources of uncertainty in approximating the MSE of the proposed empirical Bayes estimator of the SMR.

Comparing Different Small-area Estimators of Income Parameters in Panel Surveys

◆ Maria R. Ferrante, University of Bologna; Silvia Pacei, University of Bologna; Enrico Fabrizi, University of Bologna

University of Bologna, via Belle Arti 41, Statistical Dept., Bologna, 40126 Italy

ferrante@stat.unibo.it

Key Words: *average equivalized household income, European Household Community Panel, borrowing strength over time, linear mixed models, EBLUP estimators*

In Italy, estimates of equivalized household income parameters can be obtained from the Bank of Italy Survey on Income or by the European Households Community Survey (EHCS), both designed to provide estimates for large areas within the country (NUTS1). In order to obtain reliable estimates on a smaller scale, we have considered the EHCS, that is an annual panel survey, for which several successive waves are now available. The EHCS conveys a lot of information covering various characteristics of the household, allows cross-country comparison, and shows a large overlapping over successive waves. Focusing on the estimation of income parameters, the aim of this work is to experiment different small-area linear mixed models and related EBLUP estimators, where individual autocorrelation over time is taken into account. Estimators considered are compared to "standard" small-area ones, such as those based on cross-sectional borrowing strength. Their performance is evaluated through a simulation study in which samples are drawn from a pseudo-population generated replicating the EHCS sample observations by sample weights.

108 Models in Seasonal Adjustment ▲ ⌘

Business and Economics Statistics Section

Monday, August 9, 10:30 am–12:20 pm

Seasonal Adjustment of Weekly Time Series with Application to Unemployment Insurance Claims and Steel Production

◆ William P. Cleveland, Federal Reserve Board

Federal Reserve Board, Washington, DC

william.p.cleveland@frb.gov

New seasonal patterns in unemployment insurance claims have prompted a change in the method of seasonal adjustment. Seasonal adjustment of weekly series at the Federal Reserve and of initial unemployment claims at the Bureau of Labor Statistics was accomplished until recently by fixed coefficient regression. The seasonal variables were trigonometic expansions with periods 365 (366 in leap years) and harmonics to model the effect of a week ending on a particular day of the year. These variables capture the changing position of weekdays from one year to the next. While we continue to estimate holiday effects as stable over extended periods, estimation of the rest of the seasonal pattern permits an evolution of the pattern from year to year beyond what method suggested takes advantage of the periodicity of the seasonal desgin matrix to reduce the amount of computation required. Weighting schemes to achieve parameter estimates are determined by signal extraction formulas set to achieve a rate of adaptation dictated by appropriate objective functions.

Time-varying Trading-day Effects in Seasonal Adjustment of Time Series

◆ Donald E.K. Martin, U.S. Census Bureau; William R. Bell, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233-9100

donald.e.martin@census.gov

Key Words: *trading day, time-varying, RegComponent*

Trading-day effects reflect variations in a monthly time series due to the changing composition of months with respect to the numbers of times each day of the week occurs in the month. A relevant question regarding trading-day effects is whether they remain constant over time. This is especially pertinent for retail sales time series in which trading-day effects presumably depend on consumers' shopping patterns and on hours that retail stores are open, two things that have changed over time in the U.S. Seasonal adjustment practitioners sometimes deal with this issue by restricting the length of the series to which the trading-day model is fit. However, in series where the trading day varies through time, information is lost in so doing. We investigate possible time variation in trading-day effects in a large set of Census Bureau time series. We fit a model to the data that allows for stochastic trading-day coefficients that follow random walk models, and with residuals that follow an ARIMA model. As the models are a special case of the general RegComponent model, they are fit using the REGCMPNT program.

Robust Estimation of Trend and Seasonal Components in the Presence of Outliers and Level Shifts

◆ Thomas Trimbur, U.S. Census Bureau

U.S. Census Bureau, Statistical Research Division, Time Series, 4700 Silver Hill Rd. - FOB 4-3228, Washington, DC 20233-9100

Thomas.M.Trimbur@census.gov

Key Words: *level shifts, unobserved components, state-space, heavy-tailed density, non-Gaussian model, robust filter*

This paper develops methods for estimating trends and seasonal components in time series subject simultaneously to both level shifts and additive outliers. The approach is based on using heavy-tailed densities for the innovations in unobserved components models. In this framework, the timing and magnitude of large changes in the level of a process, or of outlying observations, is stochastic. Such flexibility may be helpful in modeling the range of behavior found in real time series; more informative and robust estimation becomes feasible, and this avoids the risk of major change in assessment of outlier/break points, that may have major impact on results in procedures based on binary categorization. After setting out the econometric methodology, the paper provides empirical illustrations. It is shown how the model-based assessment of the sample observations is updated optimally with the arrival of new data. Outliers and level shifts are rated on a continuous scale and distinction is made between them on the basis of available information. Trend and seasonal estimates are obtained which are less subject to revision with future data.

Span Diagnostics for Model-based Seasonal Adjustment

◆ Roxanne Feldpausch, U.S. Census Bureau; Kellie Wills, U.S. Census Bureau; Catherine Hood, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233

roxanne.feldpausch@census.gov

Key Words: *signal extraction, spectrum diagnostics, ARIMA models*

We investigate diagnostics for determining whether a series is suitable for ARIMA model based (AMB) seasonal adjustment. Since model inadequacy can lead to inadequacies in the AMB decomposition, evaluating model fit is an essential part of AMB seasonal adjustment. We investigate how effectively two model-fit diagnostics—ACF and Ljung-Box Q—identify series for which AMB seasonal adjustment will be problematic. If the process generating the data changes over time, the ARIMA model chosen for the full data may be inappropriate for some time period, leading to instability of the adjustment. We evaluate model fit in overlapping spans of data to detect situations in which adding or subtracting small amounts of data results in poor fit. We also consider the use of spectrum peaks to identify data spans that do not have detectable seasonality, as well as changes in the trading day pattern. We present results from data simulated from known ARIMA models, as well as real data examples.

109 In Search of Power and Significance 10 Years Later: Statistics as a Unified Discipline for the Analysis of Stochastic Cost-Effectiveness Studies in Health Care

Section on Health Policy Statistics

Monday, August 9, 10:30 am–12:20 pm

The ICER Age: An Uncertain History

◆ Ahmed Bayoumi, St. Michael's Hospital

St. Michael's Hospital, Inner City Health Research Unit, Toronto, ON M5B 1W8 Canada

ahmed.bayoumi@utoronto.ca

Key Words: *cost-effectiveness analysis, confidence intervals, Fieller's theorem, ICER, health economics, graphical representations*

The most common outcome measure used in cost-effectiveness analysis is the Incremental Cost-Effectiveness Ratio (ICER), which divides the difference in cost of two programs by the difference in effectiveness. Although popular, the ICER has two significant limitations. First, a sample ICER is a biased but consistent estimator of the true population ICER. Second, quantifying the uncertainty associated with a ratio statistic like the ICER can be difficult. When individual-level data are available, the most promising parametric method to calculate confidence bounds uses Fieller's theorem. Nonparametric bootstrap methods have also been advanced, along with methods for bias correction. Calculating confidence intervals for samples where the mean effectiveness is near zero presents particular challenges. Graphical representation of costs and outcomes on a cost-effectiveness plane are often helpful. The bivariate density of costs and effectiveness outcomes can be demonstrated with confidence ellipses or contour plots. Rays emanating from the origin can often be drawn to represent calculated confidence bounds.

Statistical Cost-effectiveness Analysis: Incremental Net Benefit

◆ Andrew R. Willan, Research Institute, HSC

Research Institute, HSC, Population Health Sciences, 555 University Ave., Toronto, ON M5G 1X8 Canada

willana@mcmaster.ca

Key Words: *clinical trials, cost-effectiveness, incremental net benefit*

The collection of patient level cost data in clinical trials initially led investigators to conduct statistical inference on the incremental cost-effectiveness ratio (ICER). The problems with ratio statistics, however, quickly led to the proposal for using the incremental net benefit (INB) as an alternative. We illustrate how INB solves the each of the problems associated with ICERs. A brief discussion on censored data and covariate adjustment will also be given.

The Cost-effectiveness Acceptability Curve—Leaving the Planes, Heading for the Frontier

◆ Elisabeth Fenwick, University of York

University of York, Dept. of Economics and Related Studies, York, YO10 5DD UK

ealf100@york.ac.uk

Key Words: *cost-effectiveness plane, statistical cost-effectiveness analysis, cost-effectiveness acceptability curve, decision-making*

Within the last 10 years, cost-effectiveness acceptability curves (CEACs) have been introduced and widely adopted within economic evaluation of health care technologies. They provide a visual impression of the joint uncertainty in the single dimension of the decision space. In particular, they overcome the particular problems associated with estimating confidence intervals for cost-effectiveness ratios. However, despite appearing in all of the major medical journals, some issues still surround the use of CEACs within health care decision-making. Firstly, there are common fallacies regarding the nature and shape of CEACs, which largely result from the "textbook" illustrations of CEACs. The reality is that the CEAC, as a graphical transformation of the cost-effectiveness plane, can take many shapes and turns depending upon the location of the joint density of incremental cost and effects within the plane. Secondly, there is confusion regarding the role that CEACs fulfil within the context of expected value decision-making. The assessment of decision uncertainty, as provided by CEACs, is a vital first step towards valuing the worth of collecting further information.

In Search of Power and Significance 10 Years Later: Statistics as a Unified Discipline for the Analysis of Stochastic Cost-effectiveness Studies in Health Care

◆ Jeffrey S. Hoch, University of Western Ontario

University of Western Ontario, Dept. of Epi & Biostats, Kresge Bldg., Room K117, London, ON N6A 5C1 Canada

jeffhoch@biostats.uwo.ca

Key Words: *net-benefit regression, cost-effectiveness, stochastic CEA, economic evaluation, econometrics, health economics*

Economic evaluation is often seen as a branch of health economics divorced from mainstream econometric techniques. Instead, it is perceived as relying on statistical methods for clinical trials. Furthermore, the statistic of interest in cost-effectiveness analysis, the incremental cost-effectiveness ratio is not amenable to regression-based methods, hence the traditional reliance on comparing aggregate measures across the arms of a clinical trial. We explore the potential for researchers undertaking cost-effectiveness analysis to exploit the plethora of established statistical techniques through the use of the net-benefit framework—a recently suggested reformulation of the cost-effectiveness problem that avoids the reliance on cost-effectiveness ratios and their associated statistical problems. This allows the formulation of the cost-effectiveness problem within a standard regression type framework. Practical advantages of the net-benefit regression approach include being able to use established statistical techniques, adjust for imperfect randomization, and identify important subgroups.

The Future for Statistical Analysis of Net-benefit in Cost-effectiveness Analysis

◆ Andrew H. Briggs, University of Oxford

University of Oxford, HERC, Dept. of Public Health, Old Rd. Campus, Headington, Oxford, OX3 7LF UK

andrew.briggs@dphpc.ox.ac.uk

Key Words: *cost-effectiveness, net-benefit, clinical trials, decision models*

The use of the net-benefit framework has greatly aided the statistical analysis of cost-effectiveness studies. In particular, the increasing use of the clinical trial as a vehicle for economic analysis has generated many statistical issues that have been made easier through the use of the net-benefit framework. However, decision models are also used for economic analysis most commonly where the analyst does not have access to the primary data and the net-benefit framework has much to offer this type of analysis also. For most trial-based analyses, some form of modeling is inevitable in order to capture the full lifetime cost and effects of treatments under evaluation. In decision models, it is increasingly recognized that there must be a firm statistical basis for parameter estimation. I will speculate that future developments in this area will revolve around an increasing convergence between trial-based and decision modeling-based evaluations. To illustrate the potential, examples of such hybrid analyses will be presented and the extent to which the use of the net-benefit aids the implementation and presentation of such studies will be emphasized.

110 Robustness of Nonparametric and Multiple Testing Methods

Section on Nonparametric Statistics

Monday, August 9, 10:30 am–12:20 pm

The Effect of Dependence on Confidence Intervals for a Population Proportion

◆ Weiwen Miao, Macalester College; Joseph L. Gastwirth, George Washington University

Macalester College, 1600 Grand Ave, St. Paul, MN 55105

miao@macalester.edu

Key Words: *confidence interval, dependent observation, coverage probability, jury discrimination, expected length*

The binomial model is widely used in statistical applications. Usually, the success probability, p, and its associated confidence interval are estimated from a random sample. Thus, the observations are independent and identically distributed. Motivated by a legal case where grand jurors could serve on consecutive years, this paper shows that when the observations are dependent, even slightly, the coverage probabilities of the usual confidence intervals can deviate noticeably from their nominal level. Several modified confidence intervals that incorporate the dependence structure are proposed and examined. Our results show that the modified Wilson, Agresti-Coull, and Jeffreys confidence intervals perform well and can be recommended for general use.

Data-driven Percentile Modified Wilcoxon Tests and Their Recomposition

◆ Olivier Thas, Ghent University; Jean-Pierre Ottoy, Ghent University

Ghent University, Coupure Links 653, Gent, 9000 Belgium

olivier.thas@UGent.be

Key Words: *changepoint, symmetry, two-sample problem, Wilcoxon rank sum, rank test, Pearson statistic*

The two-sample Wilcoxon rank sum statistic can be derived as the first component of the Pearson chi-squared statistic in a particularly constructed contingency table. For this test a “percentile modification” has been proposed, which is equivalent to splitting the contingency table into two independent subtables, and computing the Wilcoxon statistic on one of the subtables. Although this procedure does not use all data in the sample, it often results in a power increase. The splitting position is determined by an arbitrarily chosen trimming proportion p. To circumvent this problem, we propose a new test statistic by using a data-dependent choice for p. We show that its asymptotic null distribution is the supremum of a time-transformed Brownian Motion. We consider two applications of the Wilcoxon statistic: testing for symmetry and the two-sample problem. In a simulation study it is shown that our solution often results in a power advantage. Also, instead of using only one subtable, we suggest to compute the Wilcoxon statistic on both subtables, and to consider their sum as a new test statistic, which we consider as a recomposition of statistics, rather than a decomposition.

Evaluating Methods of Symmetry

◆ Jayawant N. Mandrekar, Mayo Clinic; Sumithra J. Mandrekar, Mayo Clinic; Stephen S. Cha, Mayo Clinic; Kent R. Bailey, Mayo Clinic

Mayo Clinic, 200 First St. SW, Harwick 7, Rochester, MN 55905

mandj@mayo.edu

Key Words: *distribution, skewness, symmetry*

Skewness refers to the asymmetry in a distribution. The fundamental property of a symmetric distribution is that the corresponding upper and lower percentiles are equidistant from its median. Based on this underlying assumption, we propose an approach to identify the asymmetry in a continuous distribution using the absolute deviations below and above the median. We compare the performance characteristics of this new approach with several skewness measures and a non parametric test of symmetry using simulations. In conclusion, a graphical power law transformation by Hines and Hines (1987) to make a distribution symmetric will also be presented.

Spherical Depth, Multivariate Median, and the Sign Test

WITHDRAWN

◆ Fengjuan Xuan, Pennsylvania State University; Tom Hettmansperger, Pennsylvania State University; Ryan Elmore, Australian National University

Pennsylvania State University, 331B Thomas Bldg., University Park, PA 16802

fxuan@stat.psu.edu

Key Words: *data depth, multivariate median, sign statistic, u-statistic*

The notion of spherical data depth, based on random hyperspheres in R^d, can be used as an exploratory data-analytic technique in multivariate data analysis. In addition, a new computationally efficient, affine equivariant, sample multivariate median and an affine invariant multivariate sign test will be introduced based on the spherical data depth. The large sample efficiency and robustness properties for both the median and the test are discussed.

Robust Testing in Unbalanced Heteroscedastic One-way Random Effects Models Using an R-estimator

◆ Inkyung Jung, University of North Carolina, Chapel Hill; Pranab K. Sen, University of North Carolina, Chapel Hill

ijung@bios.unc.edu

For testing the significance of random effect in the one-way random effect model, F-tests can be conducted under normality. However, if the normality assumption is violated, the F-tests may be inefficient and even inconsistent in some cases. Moreover, for heteroscedastic data, an exact F-test statistic cannot be constructed even under normality and test statistics having only approximately F distributions have been suggested. A robust rank-based test, assuming only symmetric but otherwise arbitrary continuous distributions for the random effect and random errors, is proposed here. The proposed testing procedure eliminates the need to estimate other nuisance parameters, which is required for parametric F-tests. This testing procedure is based on normal distributional approximation. Simulation studies suggest that for significance level, the proposed test is much more robust than F-tests against heteroscedasticity, unbalancedness, and departures from normality.

Improved Control and Estimation of False Discovery Rate When Substantial Dependence among the Tests Exists

◆ Cheng Cheng, St. Jude Children's Research Hospital; Stanley B. Pounds, St. Jude Children's Research Hospital

St. Jude Children's Research Hospital, Dept. of Biostatistics, 332 N. Lauderdale St., Memphis, TN 38015-2794

cheng.cheng@stjude.org

Key Words: *false discovery rate, multiple comparison, dependent tests, smoothing, microarray, data-mining*

The fast growth of data collection capabilities in science has stimulated statistical research in analysis of large datasets and massive multiple tests. It seems to be a consensus now that the false discovery rate (FDR) is an error assessment quantity more preferred than the familywise Type I error rate. Following the pioneering work of Benjamini and Hochberg, there has been much theoretical and methodological development for FDR control and estimation, for the applications in which the tests are "weakly dependent." In certain applications however, more substantial dependency may exists among the tests. We compare the control and estimation performance of several procedures, including the adaptive FDR control and the q value methods, when substantial dependency exists among the tests. We demonstrate that an accurate but slightly up-biased estimate of the proportion of the null hypotheses is important and propose a smoothing-based method to improve the existing procedures.

Practical Consideration on Optimal Combination of Multiple Tests

◆ Hua Jin, University of California, San Francisco; Ying Lu, University of California, San Francisco

University of California, San Francisco, 185 Berry St., Dept. of Radiology, San Francisco, CA 94107

hua.jin@radiology.ucsf.edu

Key Words: *multiple diagnostic predictors, receiver operating characteristic (ROC) curve, likelihood ratio score, Mann-Whitney statistic, permutation test*

Multiple alternative diagnostic tests for one disease are commonly available to clinicians. It's important to use all the good diagnostic predictors simultaneously to establish a new predictor with higher statistical utility. The likelihood ratio score leads to the uniformly most powerful test achieving the largest area under the receiver operating characteristic (ROC) curve on the basis of the famous Neyman-Pearson. However, it seems complex for practical uses. One may ask when it is possible to replace use of the complicated likelihood ratio score with a relatively simple forms such as the most commonly used linear combinations. We propose a formal statistical framework to deal with this problem. We used the Mann-Whitney statistic to estimate the area under the ROC curve and a permutation reference distribution to test the null hypothesis that the simple linear or polynomial combinations of markers have similar area under ROC curves as those of the maximum likelihood ratio scores. Monte Carlo simulations were conducted to evaluate the performance of the proposed test. The algorithm applied to data from the Study of Osteoporotic Fractures (SOF) for illustration.

111 Statistical Graphics I

Section on Statistical Graphics

Monday, August 9, 10:30 am–12:20 pm

Parallel Coordinate Plots of Gene Expression Data Differentiated by Phenotype

◆ Kellie J. Archer, Virginia Commonwealth University; Viswanathan Ramakrishnan, Virginia Commonwealth University

Virginia Commonwealth University, Dept. of Biostatistics, 1101 East Marshall St. B1-066, Richmond, VA 23298-0032

kjarcher@vcu.edu

Key Words: *microarray, gene expression, parallel coordinates*

While methods for visualizing microarray data associated with an analytical results exist, such as displaying a heatmap with associated dendrograms from hierarchical cluster analysis, in general, visualization methods for raw microarray data are not readily available due to the inherent multidimensionality. Recently, plots of gene expression data that incorporate biological information have become available. One example is the cumulative gene expression plot for a selected chromosome suggested by Gentleman (2003). What is ultimately of interest in the clinical use of microarrays is the measurement of differences in signal intensities between different phenotypes. A modification of Gentlemans cumulative expression plot by chromosome, where absolute expression for each gene is plotted via parallel coordinates, is presented. In addition, color is added to permit visualization of difference due to phenotype. These plots may aid in determining regions of deleted chromatoid material characteristic of a phenotype and may be more sensitive than traditional cytogenetic analyses. The parallel coordinate plot will be demonstrated using 31 breast cancer samples hybridized to HG-133A chips.

Large Datasets in Drug Discovery: How to Collect, Summarize, and Visualize Data from Image-based Cellular Assays in Meaningful Ways

◆ Vadim Kutsyy, Cytokinetics, Inc.

Cytokinetics, Inc., 280 E. Grand Ave., South San Francisco, CA 94022

vadim@kutsyy.com

Key Words: *drug discovery, large data, visualize data, cellular assays*

Automated high-content, cell-based tools are among the most powerful nonclinical methods for understanding the broad biological effects of uncharacterized compounds. High-content screening platforms include fluidics for plating and treating cells, automated microscopes for imaging, and software for image and data analysis. I will talk about image and data analysis for such screens. A typical image-based cellular assay generates more than 100,000 images, leading to multivariate data for tens of millions of individual cells. Thess data need to be presented to biologists and chemists in meaningful ways. A number of different methods and algorithms are employed to summarize data at different levels. This process creates an interesting statistical problem related to large datasets. Different tools are used to present these data to scientists and link it to images, and chemical or biological properties. This talk will describe methods and tools used by cytokinetics to facilitate analysis, visualization, and drilling down of discovery data across the enterprise.

Comparison of Data-driven vs. Theory-driven Method: Decision-tree Models as Data Discovery Tools in Social Science Research

◆ Bonita J. Soley, U.S. Army Research Institute for the Behavioral and Social Sciences; Krista L. Langkamer, George Mason University

U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: DAPE-ARI-PS, 2531 Jefferson Davis Hwy., Arlington, VA 22202-3926

bonita.j.soley@us.army.mil

Key Words: *comparative methods, data-mining, decision tree analysis, multivariate data visualization, classification tree, predictive analysis*

Social science research often takes a purely replicative, theory-driven approach. However, with a large database, it may be possible to utilize a more exploratory, or data-mining, approach. This research examines the use of one type of data-mining, decision-tree analysis, on an Army survey database. The objective is to determine if informed, data-driven exploratory research can replicate as well as build upon previous research by identifying trends not discovered by using strictly theory-driven methods. SPSS AnswerTree software, which has the capability of creating CHAID, Exhaustive CHAID, C&RT, and QUEST models, is used to test a model of factors related to soldier retention in the military. This model was chosen because there are numerous studies available on the topic for our comparisons. We will build multiple decision-tree models using an identical variable set; compare our models to each other; and compare our models to existing strictly theory-driven models to examine how ours are similar to, or build upon, strictly theory-driven models. Finally, the implications of using a combination data-driven, theoretical approach when analyzing large databases are discussed.

Social Network Blockmodels and Allegiance

◆ John T. Rigsby, Naval Surface Warfare Center Dahlgren Division; Jeffrey L. Solka, Naval Surface Warfare Center Dahlgren Division

Naval Surface Warfare Center Dahlgren Division, 17320 Dahlgren Rd. B10, Dahlgren, VA 22448

rigsbyjt@nswc.navy.mil

Key Words: *social network analysis, graph theory, computer networks, blockmodel*

Social Network Analysis is the study and analysis of groups as social entities to attempt to mathematically understand the interactions of people and groups in society. This analysis includes concepts such as nodes and ties, groups, subgroups, cliques, social prominence and rank, centrality, and prestige. Computers have forms of social interactions; they connect to each other and have client/server and peer to peer relationships. Computers also have trust relationships. Grouping computers into social cliques or block structures is difficult because of the different roles of computers. This paper covers a new way to estimate the number of blockmodel partitions and help with measuring change in the network model over time. The measure is called allegiance and deals with measuring actor strength as the blockmodel changes. The change in an actor's allegiance over time can be used to measure changes in the network and explain blockmodel differences.

Exploratory Analysis of Graphical Summaries of Scenarios

◆ Paul Whitney, Pacific Northwest National Laboratory; Olga A. Kuchar, Pacific Northwest National Laboratory; George Chin, Pacific Northwest National Laboratory; Mary Powers, Pacific Northwest National Laboratory; Katherine E. Johnson, Pacific Northwest National Laboratory; James M. Sloughter, University of Washington

Pacific Northwest National Laboratory, 2400 Stevens, Richland, WA

paul.whitney@pnl.gov

Key Words: *social networks, graphs, exploratory data analysis*

Events, organizations, and scenarios are often represented as highly labeled, free-form graphs; these graphs are composed of free-form text, structured text, and icons to create a visual summary. Such graphical depictions are increasingly common in the intelligence and law enforcement fields. These graphs often serve as an index or detailed summary for analysts who construct them, and also function as presentation. While software exists to assist in the construction, manipulation, and searching of these graphs, there is an increasing need for data analytic tools associated with graph representations. Methodology was developed to represent the graph summaries and to enable the exploratory analysis of large individual and/or collection of such summaries. Our methodology is based on the following strategy: (1) constructing a numeric vector representation of the objects of interest; (2) apply data analytic methodology on the numeric vectors; and (3) map the results of the data analysis back onto the data and problem space. The numeric representation is based on both graph-theory and text-analysis concepts. Examples of scenario retrieval and summary are presented.

Structured Multicategory Support Vector Machine with ANOVA Decomposition

◆ Yoonkyung Lee, The Ohio State University; Ja-Yong Koo, Inha University; Yuwon Kim, Seoul National University; Sangjun Lee, Seoul National University

The Ohio State University, Dept. of Statistics, 1958 Neil Ave., Columbus, OH 43210

yklee@stat.ohio-state.edu

Key Words: *Multicategory Support Vector Machine, functional ANOVA decomposition, feature selection, classification, 1-norm penalty*

The Support Vector Machine (SVM) has been a popular choice of classification method for many applications in machine learning. While it often outperforms other methods in terms of classification accuracy, the implicit nature of its solution renders the method less attractive in providing insights into the relationship between covariates and classes. Using structured kernels can remedy the drawback. Borrowing flexible model building idea of functional ANOVA decomposition, Multicategory Support Vector Machines with ANOVA kernels are considered in this paper. An additional penalty is imposed on the sum of weights of functional subspaces, which encourages a sparse representation of the solution. Incorporation of the additional penalty enhances the interpretability of a resulting classifier with often improved accuracy. The proposed method is demonstrated through simulation studies and applications to real data.

112 Survival Analysis ⌘

Biopharmaceutical Section

Monday, August 9, 10:30 am–12:20 pm

A Joint Longitudinal and Survival Model that Flexibly Incorporates Nonproportional Hazards Treatment Effects

◆ Daniel L. Gillen, University of Chicago; Sebastien Haneuse, University of Washington

University of Chicago, 5841 South Maryland Ave., MC2007, Chicago, IL 60637

dgillen@health.uchicago.edu

Key Words: *survival, nonproportional hazards, longitudinal, Bayesian*

Time-to-event is a common outcome in many prospective studies. In addition, researchers may also monitor additional biologic endpoints that are hypothesized to be associated with survival as patients are continuing to be followed. Recently, Brown and Ibrahim (2003) considered a Bayesian semiparametric hierarchical model for simultaneously modeling survival and longitudinal data in a proportional hazards setting. We consider an extension of the Brown and Ibrahim model which flexibly accounts for nonproportional hazards covariate effects on survival using a conditional hazards survival model similar to that introduced by McKeague and Tighiouart (2000). In the conditional hazards survival model, the baseline log-hazard and covariate effects are estimated via step functions, forming a first-order autoregressive process, while the grid of jump times that define these step functions form a time-homogenous Poisson process. The model is fit using a Metropolis-Hastings-Green algorithm. We demonstrate the method using a clinical example in which time to mortality is modeled as a function of longitudinally measured serum albumin in end-stage renal disease patients.

Tests for Time to a Favorable Event

◆ David R. Bristol, Purdue Pharma LP; Nicole O. Stouffer, Purdue Pharma LP

Purdue Pharma LP, 6 Cedar Brook Dr., Cranbury, NJ 08512

david.bristol@pharma.com

Key Words: *censoring, analgesia*

When assessing the time to success (favorable event), the times to failure (unfavorable event) can be handled in various ways. Small times to success are consistent with a response to an effective treatment; also, large times to failure may contribute to the evidence of an effective treatment. The approaches considered here are to treat the unfavorable events as (1) censoring events at the time of occurrence, (2) censoring events at the end of the study (EOS), or (3) observed events at EOS. The analyses for between-treatment comparisons are the log-rank and the (generalized) Wilcoxon tests when the unfavorable events are considered censoring events and the Wilcoxon rank sum test when the unfavorable events are considered observed events at EOS. The methods of data handling and analysis may affect the significance of the treatment differences. Results of simulations, conducted using various exponential distributions for times to success and failure, are presented to examine the impact. If the two treatments have the same distributions of the time to failure, the log-rank test is preferable.

A New Testing Approach for Comparing the Overall Homogeneity of Survival Curves

◆ Hongkun Wang, University of Rochester; Xun Lin, Pfizer Global Research and Development

University of Rochester, 60 Crittenden Blvd. Apt. 732, Dept. of Biostatistics, Rochester, NY 14620

wang@bst.rochester.edu

Key Words: *survival analysis, censoring, log-rank test, Wilcoxon test, statistical power*

A new statistical testing approach is developed for comparing the overall homogeneity of survival curves. The proposed new method has greater power than the commonly used tests to detect overall differences between crossing survival curves. The small-sample performance of the new test is investigated under a variety of situations by means of Monte Carlo simulations. Furthermore, the applicability of the proposed testing approach is illustrated by a real data example from a kidney dialysis trial.

A Simple Hybrid Variance Estimator for the Kaplan-Meier Survival Function

◆ Craig B. Borkowf, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, CDC/NCID/DVRD/ Influenza Branch, MS A32, 1600 Clifton Rd. NE, Atlanta, GA 30333

CBorkowf@cdc.gov

Key Words: *clinical trial, Kaplan-Meier survival function, Greenwood variance estimator, Peto variance estimator, simulation study, survival analysis*

We propose a new hybrid variance estimator for the Kaplan-Meier survival function. This new estimator approximates the true variance by a binomial variance formula, where the proportion parameter is a piecewise nonincreasing function of the KM survival function and its upper bound. Also, the effective sample size equals the number of subjects not censored prior to that time. In addition, we consider an adjusted hybrid variance estimator that modifies the regular estimator for small sample sizes. We present a simulation study to compare the performance of the regular and adjusted hybrid variance estimators to the Greenwood and Peto variance estimators for small sample sizes. We show that the new variance estimators give better variance estimates and confidence intervals with more nominal coverage rates than the traditional variance estimators. Indeed, the Greenwood and Peto variance estimators can substantially underestimate the true variance in the left and right tails of the survival distribution, even with moderately censored data. Finally, we illustrate the use of these new and traditional variance estimators on an example from a leukemia clinical trial.

A Novel Approach of Estimation for Additive Rate Regression Models with Parametric Underlying Failure-time Distributions

◆ M. Brent McHenry, University of Pittsburgh; Stuart R. Lipsitz, Medical University of South Carolina; Debajyoti Sinha, Medical University of South Carolina

University of Pittsburgh, 135 Cannon St. Suite #303, Charleston, SC 29425

mbmst29@pitt.edu

Key Words: *exponential distribution, piecewise exponential distribution, hazard rate difference, MLE, Poisson regression, additive hazard rate regression model*

For failure time outcomes, modeling the hazard rate as an exponential function of covariates is by far the most popular. However, in the last few decades, additive hazard rate regression models have received some attention, in which the hazard rate is modeled as a linear function of the covariates. Popular fully parametric distributions include the exponential and piecewise exponential. For an additive rate regression model in which the distribution of the failure time is exponential or piecewise exponential, we show that the maximum likelihood estimates (MLE) can be obtained using a Poisson linear model, without any additional programming or iteration loops. As a result, the MLEs can be obtained in any generalized linear models program. We apply the method to datasets in which the additive hazard rate regression model appears more appropriate.

Estimation of Lag Time between Onset of and Death from an Occult Tumor

◆ Hojin Moon, U.S. Food and Drug Administration; Hongshik Ahn, SUNY, Stony Brook; Ralph L. Kodell, U.S. Food and Drug Administration

U.S. Food and Drug Administration, National Center for Toxicological Researc, Div of Biometry and Risk Assessment, 3900 NCTR Dr., Jefferson, AR 72079

hmoon@nctr.fda.gov

Key Words: *bioassay, cause of death, lag time, likelihood, tumor lethality*

A new statistical procedure for estimating the lag time between onset of and death from an occult tumor is proposed for data without cause-of-death information for each animal or context of observation for each tumor. The proposed method is based on a nonparametric model. Tumor lethality information required for the lag-time estimation is obtained using statistical methods. In the proposed method, numbers of fatal and incidental tumors are statistically imputed by using the estimates of the tumor onset function and tumor-specific survival function obtained from closed form solutions for nonparametric maximum-likelihood estimators under explicit and implicit inequality constraints. The proposed method is illustrated using a real dataset. The accuracy of the estimation of lag time is evaluated via a Monte Carlo simulation study.

Alternative Approach for Estimating Tumor Response Rate

◆ Xiaolong Luo, Johnson & Johnson Pharmaceutical R&D, LLC

Johnson & Johnson Pharmaceutical R&D, LLC, 920 Rte. 202 South, Raritan, NJ 08869

xluo@prdus.jnj.com

Key Words: *clinical trial, cancer, binomial, Markov, interim analysis, consistency*

In clinical oncology trials, patients receive treatments of various modalities. During a study, a patient's tumor volume is periodically measured and a confirmed shrinkage, i.e., tumor response, is often used as a study endpoint. The assessment of such tumor response rate is typically based on binomial distribution modeling. This approach is acceptable when the trial is finished and all clinical assessments have been completed. However, it can lead to many practical issues for a trial interim analysis. One dilemma is whether to include ongoing patients. We cannot include them since their endpoints have not been reached and we do not want to exclude them either since they would represent a significant portion of the data. In this paper, we propose a Markov process model for the longitudinal tumor volume measurements and develop an algorithm to estimate tumor response rate. The method would allow us to utilize all ongoing measurements independent of whether final endpoint has been reached. We will use simulation analyses to compare both approaches. In addition, we will derive large sample theory for the new estimator's asymptotic consistency and limiting distribution.

113 Nonparametric and Varying Coefficients Models for Longitudinal and Clustered Data

Section on Nonparametric Statistics, Section on Statistics in Epidemiology, Biometrics Section

Monday, August 9, 10:30 am–12:20 pm

Asymptotic Properties and Inferences for Varying Coefficient Regression with Longitudinal Variables

◆ Weishi Yuan, Duke University; Colin O. Wu, National Heart, Lung, and Blood Institute

Duke University, Institute of Statistics & Decision Sciences, Box 90251, Durham, NC 27708-0251

wsyuan@yahoo.com

Key Words: *regression, varying coefficient, longitudinal data, cross-validation, simulations*

Longitudinal data usually involve repeated measurements from a set of randomly chosen subjects, and this type of data frequently appears in biomedical and epidemiological studies. An important objective of statistical analyses with this type of data is to evaluate the effects of the covariates, which may or may not depend on time, on the time-varying response variable of interest. A well-developed regression methodology could have important practical impacts in evaluating new medical treatment, identifying influential risk factors, verifying existing biological models, etc. The varying-coefficient model is a structural nonparametric model that is particularly useful in exploring the time trend and associations between longitudinal outcomes and covariates. We investigate here a class of two-step smoothing methods based on covariate centering for estimating the coefficient curves in a linear time-varying coefficient model with time-dependent covariates. A cross-validation criterion is used to select the smoothing parameters. Methods for statistical inferences are suggested based on the asymptotic distributions of the estimators or a bootstrap procedure.

Jackknifing Bias of Box-type Approximation under Dependence: A Nonparametric Approach for Longitudinal Data

◆ Xiaofeng Wang, Case Western Reserve University ; Zhaozhi Fan, University of New Hampshire; Jiayang Sun, Case Western Reserve University

Case Western Reserve University, Dept. of Statistics, CWRU, 10900 Euclid Ave., Cleveland, OH 44106-7054

xiaofeng@cwru.edu

Key Words: *Box-type approximation, longitudinal data, jackknife, heteroscedastic errors, robust analysis of variance*

Asymptotically distribution-free rank tests have been developed for testing nonparametric hypotheses in factorial experiments when data are independent. The asymptotic null distribution developed by Brunner, Dette, and Munk (1997) for these tests is conservative when data are dependent, because the estimator of d, the degrees of freedom used in the approximation is biased. We provide a bias-corrected estimate of d using a Box-type correction. This is related to the jackknife delete-k procedure. The choice of k is studied. The new approximation to the null distribution of these test statistics is shown to be more accurate than BDM's approximation, via analysis and simulation studies. Our methodology is applied to a longitudinal case study in epidemiology.

Profile-kernel vs. Backfitting in the Estimation of Partially Linear Models

◆ Zonghui Hu, Texas A&M University; Naisyin Wang, Texas A&M University; Raymond J. Carroll, Texas A&M University

Texas A&M University, 1 Hensel Dr., #W3K, College Station, TX 77840

zhu@stat.tamu.edu

Key Words: *local linear smoothing, backfitting, bandwidth, consistency, asymptotics*

We study the profile-kernel and backfitting method in partially linear models for clustered/longitudinal data. For independent data, despite the inconsistency of backfitting estimator noted by Rice (1986), the two estimators have the same asymptotic variance as shown by Opsomer and Ruppert (1999). When an undersmoothing nonparametric procedure is adopted, the two methods are considered as equivalent. Theoretical comparisons of the two estimators for multivariate responses are investigated. We find out that for correlated data, backfitting often produces a larger asymptotic variance than the profile-kernel, in addition to its bias problem. Our simulation study clearly shows that performance of profile-kernel is superior to that of backfitting for finite samples. The application of both methods to an opthalmology dataset is provided.

Rank-based Analyses of Repeated Measures Models

◆ John D. Kloke, Western Michigan University; Joseph McKean, Western Michigan University

kloke@stat.wmich.edu

Key Words: *rank, repeated measures, Wilcoxon*

Rank-based analyses of repeated measures models are developed. In these models the subjects are independent, while a compound symmetry variance-covariance structure is assumed for within a subject. We briefly present the theory of these analyses and then discuss with examples several of the commonly used models including the means model and analysis of covariance models. Emphasis will be placed on the Wilcoxon analysis. Besides examples, some simulation results will be provided.

The Wilcoxon Signed-rank Test for Cluster Correlated Data

◆ Denis Larocque, HEC Montréal

HEC Montréal, Dept. of Management Sciences, 3000, chemin de la Cote-Sainte-Catherine, Montréal, PQ H3T 2A7 Canada

denis.larocque@hec.ca

Key Words: *sign test, one-sample location problem, Wilcoxon signed-rank test, intraclass correlation, one-way random effect, clustered observations*

We adapt the Wilcoxon signed-rank test to the case of cluster correlated data. A simple modification of the estimator of the asymptotic variance is sufficient to obtain a valid asymptotic procedure. However, the resulting test is no longer distribution-free. We derive the asymptotic null distribution of the statistic. We also study the Pitman asymptotic efficiency of the test and of the sign test. It is shown that for the normal and t distributions, the sign and Wilcoxon signed-rank tests always get better compared to an average based test as soon as intracluster correlation is present. Interestingly, the sign test can even be more efficient than the Wilcoxon signed-rank test in some situations. A simulation study is performed in order to investigate the finite sample performance of the tests. The results show that the performance of the Wilcoxon signed-rank test is very good for all designs and distributions considered when compared to competitors based on signs and on the overall average. In fact, the test is as powerful as the one based on the overall average for normal data as soon as intracluster correlation is present.

Functional Regression for Clustered Data

◆ Yolanda Munoz Maldonado, Texas A&M University

Texas A&M University, Dept. of Statistics, 3143 TAMU, College Station, TX 78743-3143

Y_Munoz70@hotmail.com

Key Words: *functional data, regression curves, clustered data, smoothing splines, Kalman filtering*

Functional regression analysis is studied in a clustered data setting. We propose a general approach for estimating regression curves using spline smoothing type methods that are suitable for such an environment. Efficient algorithms to compute these estimators are developed via use of the Kalman Filter. The algorithms are implemented using the SAS PROC STATESPACE procedure. Run-time comparisons are performed between this approach and an alternative mixed models implementation using SAS PROC MIXED.

114 Statistical Issues in Microarrays and Proteomics

ENAR, WNAR

Monday, August 9, 10:30 am–12:20 pm

Generalized Models and Design Issues in Genomic and Proteomic Experiments

◆ Mark Carpenter, Auburn University

Auburn University, Dept. of Mathematics and Statistics, Auburn, AL 36849

carpdem@auburn.edu

Key Words: *generalized models, generalized gamma, missing at random, microarray, 2D-gels*

Because modern high-throughput genomic and proteomic experiments can potentially generate tens of thousands of response variables, research in statistical methodology for these experiments have understandably focused on issues of data screening, data reduction, adjustments for multiple testing, and other high-end data analyses. We develop methods for separately modeling each response variable using generalized models. This approach increases the sensitivity for discovering biologically significant results while minimizing false discovery due to model choice. Since issues of experimental design depend heavily on the statistical methods and models to be applied, we also address issues of sample size estimation for simple designs. In particular, we discuss the appropriateness of technical replication (same-sample repeated measures) when there is a fixed experiment budget where increase in technical replications would reduce the number of actual biological replicates (independent observations). The methods are applied to data from microarray and 2D-gel proteomic experiments.

Analytic Strategies for Pre-processing in Proteomics Experiments

◆ Laura L. Johnson, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, NCI/CPSB, 6116 Executive Blvd., Room 705, MSC 8314, Bethesda, MD 20892-8314

johnslau@mail.nih.gov

Key Words: *proteomics, mass spectrometry, SELDI-TOF, measurement error*

Using Surface Enhanced Laser Desorption/Ionization-Time of Flight (SELDI-TOF) mass spectrometry (MS) data as an example, we discuss current statistical methods concerning how to align peaks and how to calibrate mass/charge measurements across multiple samples. Output from a specimen yields a series of (x,y) points with x the protein mass divided by the number of charges introduced by ionization (M/Z) and y the protein intensity of the M/Z value in the specimen. Two measurement error issues, measurement error in the absolute values of the intensity for any given M/Z point and drift experiment to experiment by the M/Z axis (x-axis) need to be addressed before analysis can begin. These methods apply to data from other proteomic platforms and are vital when constructing and evaluating disease classifiers.

MS-based Statistical Analysis for Proteomic Marker Discovery

◆ Xuena Wang, Stony Brook University; Wei Zhu, SUNY, Stony Brook

Stony Brook University, Dept. of Applied Math and Statistics, Stony Brook, NY 11794

wangxn@ams.sunysb.edu

Key Words: *Gaussian random field, stepwise discriminant analysis, K-nearest neighbor method, cross-validation, resampling, multiple comparison*

With the maturation of the new Mass Spectrometry (MS) technologies, as well as the availability of a well-defined clinical database, the MS-based statistical analysis of volume-produced serum or plasma mass spectra retrieved from healthy and disease-affected individuals plays an increasingly important role in the field of proteomic research, especially the diagnosis of disease and the follow-up of treatment effect. To decrypt a disease-specific proteomic pattern is extremely challenging considering the complexity of the sample individuality and the multitude of disease stages and combinations. For a large scale screening of diseases with relative low prevalence such as the ovarian cancer, Positive Predictive Value (PPV) of 100% becomes critical. We propose a novel statistical procedure identifying the best model of biomarkers for the optimal classification between two different groups of subjects. In comparison with other approaches reported to date, this procedure is shown to be more robust and have higher discriminatory power when applying to the ovarian cancer study in the analytical platforms of both low-resolution PBS-II TOF MS and high-resolution QqTOF MS.

Spectral Clustering of Microarray Data

◆ David L. Tritchler, Ontario Cancer Institute; Fallah Shafagh, University of Toronto; Joseph Beyene, Hospital for Sick Kids

Ontario Cancer Institute, 610 University Ave., Toronto, ON M5G 2M9 Canada

tritchle@uhnres.utoronto.ca

Key Words: *clustering, eigenanalysis, DNA microarray*

This paper introduces a clustering method motivated by a multivariate analysis of variance model and computationally based on eigenanalysis (thus the term “spectral” in the title). Our focus is on large problems, and we present the method in the context of clustering genes using microarray expression data. We provide a computational algorithm and discuss its time complexity. The method is extended to enable supervision by clinical information and the problem of specifying the number of clusters is discussed. A leukemia dataset is analyzed to demonstrate the use of the method.

Effects of Pooling RNA in Microarray Class Comparisons

◆ Joanna H. Shih, National Cancer Institute; Aleksandra M. Michalowska, National Cancer Institute; Kevin K. Dobbin, National Cancer Institute; Yumei Ye, National Cancer Institute; Tinghu Qiu, National Cancer Institute

National Cancer Institute, 6130 Executive Blvd., EPN, Room 8132, Bethesda, MD 20852-7434

jshih@mail.nih.gov

Key Words: *microarray, pooling, variation, sample size, power*

In microarray experiments investigators sometimes wish to pool RNA samples before labeling and hybridization due to insufficient RNA from each individual sample or to reduce the number of arrays for the purpose of saving cost. The basic assumption of pooling is that the expression of an mRNA molecule in the pool is close to the average expression from individual samples. Recently, a method for studying the effect of pooling mRNA on statistical power in detecting differentially expressed genes between classes has been proposed, but the different sources of variation arising in microarray experiments were not distinguished. We study the implication of pooling in detecting differential gene expression taking into account different sources of variation. We present a formula for the required number of subjects and arrays to achieve the same power as in a nonpooled experiment. We show that due to the loss of degrees of freedom for a pooled design, a large increase of the number of subjects may be required to achieve a comparable power. We also check the adequacy of the model assumption of pooling using real data.

Sample Size Determination in Microarray Experiments for Class Comparison and Prognostic Classification

◆ Kevin K. Dobbin, National Cancer Institute; Richard Simon, National Institutes of Health

National Cancer Institute, 6130 Executive Blvd., NIH, Bethesda, MD 20892-7434

dobbinke@mail.nih.gov

Key Words: *microarrays, experimental design, sample size*

Determining sample sizes for microarray experiments is important but the complexity of these experiments, and the large amounts of data they produce, can make the sample size issue seem daunting, and tempt researchers to use rules of thumb in place of formal calculations based on the goals of the experiment. Here we present formulas for determining sample sizes to achieve a variety of experimental goals, including class comparison and the development of prognostic markers. We derive results describing the impact of pooling, technical replicates and dye-swap arrays on sample size requirements. We show how results depend on the relative sizes of different sources of variability. We consider a variety of common types of experimental situations and designs that are used with single-label and dual-label microarrays. We discuss procedures for controlling the false discovery rate. Our calculations are based on relatively simple yet realistic statistical models for the data, and provide straightforward sample size calculation formulas.

Variability Sources in Gene Expression Data

◆ Johanna Hardin, Pomona College

Pomona College, Dept. of Mathematics, 610 N. College Way, Claremont, CA 91711

jo.hardin@pomona.edu

Key Words: *microarray, t-distribution, simulation, robust, simulation*

One of the difficulties in analyzing microarray data is in knowing what the underlying distribution is. We know how to transform microarray data to deal with symmetry and nonconstant variance across different expression levels, but the resulting transformed data have much heavier tails than a normal distribution. We have provided a method of fitting a t-distribution to a dataset using a robust estimate of scale. T-distributions arise from scale mixtures of normals as well as from inherently long-tailed distributions. Using multiple microarray datasets of different species and technologies, we investigate the source of the variability for genes within a sample and for genes across replicates.

115 Networks

Section on Physical and Engineering Sciences

Monday, August 9, 10:30 am–12:20 pm

Active Network Monitoring through Tomography Techniques

◆ Bowei Xi, University of Michigan; George Michailidis, University of Michigan; Vijayan N. Nair, University of Michigan

University of Michigan, 439 West Hall, 550 East University, Dept. of Statistics, Ann Arbor, MI 48109

xbw@umich.edu

Key Words: *network tomography, Exponentially Weighted Moving Average control chart, large scale monitoring*

Estimation of quality of service parameters, such as packet loss rates and delay distributions, is of considerable importance to network administrators and service providers for monitoring purposes. Active network tomography techniques enable the estimation of packet loss rates and delay distributions of all internal links in a network from measurements obtained from nodes located on its periphery. This is an example of a large-scale statistical inverse problem. Least squares algorithms are introduced for the fast estimation of the link loss rates. The link monitoring problem is then formulated as a change detection problem and the Exponentially Weighted Moving Average (EWMA) control chart is adapted to the problem. The properties of these charts are discussed and the results illustrated both with synthetic and real network traces.

Estimation of Network Link Characteristics Based on Time Series Data

◆ Earl Lawrence, University of Michigan; George Michailidis, University of Michigan; Vijayan N. Nair, University of Michigan

earl@umich.edu

Key Words: *internet, network, link loss, link delay, time series*

As the internet grows in both size and importance, accurate methods of measuring its performance are increasingly important for both planning and monitoring purposes. There are several techniques for measuring network link performance characteristics such as delay and loss, but all of them ignore the traffic's time-based correlation—consecutive packets along the same link are likely to experience similar delays and losses. Ignoring this information can lead to biased estimates and remove vital information that would be useful in real-time monitoring. We discuss the estimation of network link characteristics taking into account the time series nature of the data. The focus will be on efficient inference methods from end-to-end measurements (i.e., the traffic will cross several links).

Origin-Destination Matrix Estimation from Partially Observed Packet/Vehicle Trajectories

◆ Jaimyoung Kwon, University of California, Berkeley; Pravin Varaiya, University of California, Berkeley

University of California, Berkeley, 367 Evans Hall #3860, Berkeley, CA 94720

kwon@stat.berkeley.edu

Key Words: *Origin-Destination matrix, method of moment estimator, network, partially observed data*

Accurate information on Origin-Destination (OD) traffic matrices is critical in engineering tasks for both computer and transportation networks. Since observing OD traffic matrices directly is difficult for large networks, they have been estimated either from survey sampling data or from link counts data using network equilibrium, gravity, or distribution-assignment models. Another data type is partially observed packet or vehicle trajectories. We develop Method of Moment (MM) algorithm for estimating time-varying OD matrices from such data. Estimator errors are calculated either from asymptotic approximation or bootstrap. Simulation study shows that the proposed estimator and standard errors estimates perform well under ideal settings. Recently, such partially observed vehicle trajectories data became available from vehicles equipped with ETC transponders and "readers" installed at various locations in San Francisco Bay Area. The algorithm is also applied to these data to derive hourly OD matrices.

Wavelet Methods for the Detection of Anomalies and Their Application to Network Traffic Analysis

◆ Deukwoo Kwon, Texas A&M University; Marina Vannucci, Texas A&M University

Texas A&M University, 3143 TAMU, Dept. of Statistics, College Station, TX 77843-3143

dwkwon@stat.tamu.edu

Key Words: *network traffic anomalies, online detection, wavelet packet transform, iterated cumulative sum of squares, Schwarz information, changepoint detection*

We develop an integrated tool for the online detection of network anomalies. In order for the new approach we consider two kinds of changepoint detection algorithms. One is for the variance change detrection the other the jump detection. For the former detection we modify iterative cumulative sum of squares (ICSS) and the test based on the Schwarz information criterion (SIC). For the latter we also modify the jump detection suggested by Wang (1995). Those are all for the offline detection methods. We modify those algorithms with moving window technique for the online detection purpose. We make simulated network traffic data with several sophisticated types of attacks against the network. We mainly use the wavelet transform since wavelet transform allows us to be able to use those algorithms. We also examine the performance of the ICSS and the SIC with simulated data and compare two algorithms using the mean delay concept.

The Class of Beta-exponential Distributions: Properties, Estimation, and Applications

◆ Jacinth A. Maynard, University of Pittsburgh, Bradford

University of Pittsburgh, Bradford, 234E Swarts Hall, 300 Campus Dr., Bradford, PA 16701

jam87@exchange.upb.pitt.edu

Key Words: *class of beta-exponential distributions, moments, parameter estimation, goodness of fit, hazard function*

A new general model, a class of beta-exponential distributions, generated from the distribution of the beta random variable, is developed. The beta-exponential distribution is a three parameter probability model. Properties such as moments and limiting properties are established. The method of moments and the maximum likelihood method are used to estimate the parameters. Some real-life data are fitted and the goodness of fit is compared to that of the Weibull, the gamma, the exponentiated-exponential, the Lagrange gamma, and the beta-normal family. The hazard function of the beta-exponential distribution is investigated and compared to the hazard functions of the gamma, the Weibull, and the exponentiated-exponential distributions. In five of the seven datasets considered, the beta-exponential distribution provided a better fit than one or more of the other distributions used for comparison. Furthermore, the hazard function of the beta-exponential distribution behaves similarly to, but more general than, the hazard function of the Weibull, gamma, and exponentiated-exponential distributions.

A Nonparametric Test

◆ Kaisheng Song, Florida State University

Florida State University, Dept. of Statistics, Tallahassee, FL 32306

kssong@stat.fsu.edu

Key Words: *nonparametric, test*

In this paper, a new nonparametric test is proposed and its applications are discussed.

Generalized Prediction Intervals

◆ Chih-Ming Wang, National Institute of Standards and Technology; Hari K. Iyer, Colorado State University

National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305

jwang@boulder.nist.gov

Key Words: *generalized inference, random effects models, variance components*

We develop the method of the generalized prediction pivotal quantities (GPPQ) and use the method to derive prediction intervals for various models. We illustrate the method of GPPQ by constructing prediction intervals for future observations when sampling from the balanced one-way random effects models.

116 Regression Models ⌘

General Methodology

Monday, August 9, 10:30 am–12:20 pm

One-sample Regression Procedure for Testing a General Linear Hypothesis under Heterogeneous Error Variances

◆ Hubert J. Chen, National Cheng Kung University; Miin-jye Wen, National Cheng Kung University

National Cheng Kung University, 1 University Rd., Tainan, Taiwan, ROC

jpchen@email.stat.ncku.edu.tw

Key Words: *linear models, heterocedasticity, hypothesis testing, confidence region, Student's t distribution*

Assuming a general linear model with unknown and possibly unequal normal error variances, the goal is to propose a one-sample testing procedure for testing a general linear hypothesis and constructing a confidence region concerning a set of estimable functions of regression parameters. A new test statistic is constructed based on a weighted sample mean at each of predictor's data points; it turns out to be a quadratic function of several independent Student's t random variables under the null hypothesis. As a result, the distribution of the proposed test statistic is completely independent of the unknown error variances. Hence, the p value and/or the critical values of such test can be obtained from computer simulation using SAS language for small samples, or approximated by a chi-squared distribution for large samples.

Multiobjective Modeling and Two-parameter Ridge Regression

◆ Stan Lipovetsky, GfK Custom Research Inc.; W. Michael Conklin, GfK Custom Research Inc.

GfK Custom Research Inc., 8401 Golden Valley Rd., PO Box 27900, Minneapolis, MN 55427

lipovetsky@Gfkcustomresearch.com

Key Words: *multiple regression, ridge regression, multicollinearity, net effects, multiobjective modeling, robustness*

We consider simultaneous minimization of the model errors, deviations from orthogonality between regressors and errors, and deviations from other desired properties of the solution. This approach corresponds to a regularized objective that produces a consistent solution not prone to multicollinearity. We obtain a generalization of the ridge regression to two-parameter model that always overperforms a regular one-parameter ridge by better approximation, and has good properties of orthogonality between residuals and predicted values of the dependent variable. The results are very robust and convenient for the analysis of regression. Numerical applications in marketing research demonstrate that this technique works well and is useful for interpretation of the results of regression modeling.

Practical Confidence Intervals for Regression Quantiles

◆ Yunming Mu, University of Illinois; Xuming He, University of Illinois, Urbana-Champaign; Masha Kocherginsky, University of Chicago

University of Illinois, Dept. of Statistics, 101 Illini Hall, 725 S. Wright St., Champaign, IL 61820

ymu@uiuc.edu

Key Words: *confidence interval, Markov chain marginal bootstrap, regression quantile, standard error*

Routine applications of quantile regression analysis require availability of reliable and practical algorithms for estimating standard errors, variance-covariance matrices as well as confidence intervals. Since the asymptotic variance of a quantile estimator depends on error densities, the standard large sample approximations have been found to be highly sensitive to minor deviations from the i.i.d. error assumption. We propose a time-saving resampling method based on a simple but useful modification of the Markov chain marginal bootstrap to construct confidence intervals in quantile regression. This method is compared to several existing methods with favorable performance in speed, accuracy and reliability. The talk is based on joint work with Masha Kocherginsky and Xuming He.

Orthogonal Regression by Unitary Transformations

◆ Jonathon Schuler, Naval Research Laboratory; James E. Gentle, George Mason University

Naval Research Laboratory, Optical Sciences Division/Code 5661, 4555 Overlook Ave. SW, Washington, DC 20375

jonathon.schuler@nrl.navy.mil

Key Words: *robust, orthogonal, regression*

This paper revisits and generalizes a published technique (1988), a routine for converting regression algorithms into corresponding orthogonal regression algorithms, that implemented orthogonal regression of two variables by determining a unitary rotation of variables such that the regression slope of the transformed variables became zero; subsequent derotation of this zero-slope regression axis corresponded to the orthogonal regression of the original variables. This paper extends the technique to the case of hyperplane surfaces resulting from the regression of three or more variables, particularly when some variables are assumed to admit measurement errors while others are known to not. This generalized approach leverages the use of any robust ordinary regression algorithm under any error metric, offering specific advantages in performance over the Total Least Squares approaches of orthogonal regression when there are deviations from the assumption of error independence and homoscedasticity.

Simultaneous Confidence Bands in Regression with Constrained Predictors

◆ Mortaza Jamshidian, California State University, Fullerton; Wei Liu, University of Southampton; Ying Zhang, University of Central Florida; Johnathan Donnelly, University of Southampton

California State University, Fullerton, Mathematics Dept., 800 N. State College Blvd., Fullerton, CA 92834

mori@fullerton.edu

Key Words: *linear regression, simultaneous confidence bands, statistical simulation*

This article presents a method for the construction of a simultaneous confidence band for the normal-error multiple linear regression model. The confidence bands considered have their width proportional to the standard error of the estimated regression function, and the predictor variables are allowed to be constrained in intervals. Published papers in this area give exact bands only for the simple regression model. When there is more than one predictor variable, only conservative bands are proposed in the statistics literature. This paper advances this methodology by providing exact confidence bands for regression models with any number of predictor variables. Additionally, a criterion is proposed to assess the sensitivity of a simultaneous confidence band. This criterion is defined to be the probability that a false linear regression model is excluded from the band at least at one point and hence this false linear regression model is correctly declared as a false model by the band. Finally, the paper discusses computational algorithms for obtaining the confidence band.

Roles of Shrinkage Concepts in "Black Box" Linear Models Analysis

◆ Dan Spitzner, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics (0439), 413B Hutcheson Hall, Blacksburg, VA 24061

dan.spitzner@vt.edu

Key Words: *ridge regression, model selection, Bayesian model averaging, continuum regression, cross-validation, information criteria*

The roles of various shrinkage concepts are examined within a broad "black box" framework, in which optimality of an objective function serves as the sole criterion for choice of estimator. Shrinkage concepts studied are: "shrinkage to a point," e.g., James-Stein and ridge regression; "shrinkage by dimension reduction," e.g., variable subset selection and Bayesian model averaging; and the "nonlinear shrinkage" of continuum regression. These are studied under a broad range of both frequentist and empirical Bayes optimality criteria, in which the conventional framework is expanded to include weighted variations designed to match inferential goals such as prediction or parameter estimation. The study delineates a type of singularity by which dimension reduction contributes only slightly, if at all, to estimator performance, leaving "shrinkage to a point" to act the main vehicle for optimality.

Maximum Canonical Correlation Regression

◆ Mu Zhu, University of Waterloo

University of Waterloo, 200 University Ave. W, Waterloo, ON N2L 3G1 Canada

mzhu@post.harvard.edu

Key Words: *Lagrange multipliers, marketing, multivariate analysis, penalized least squares, projection*

Suppose we want to make predictions using variables that are different from the ones we have in the training data. What makes this possible is the availability of a small pilot study from which some information can be obtained on the joint behavior of these different sets of predictors. Using both the training data and data from the pilot study, we can then solve a simultaneous optimization problem that combines least-squares regression and canonical correlation analysis. We show what the solution looks like in a simple case, look at some examples and briefly discuss directions for further research.

117 Topics in Statistical Consulting

Section on Statistical Consulting

Monday, August 9, 10:30 am–12:20 pm

A Matrix-based Genetic Algorithm for Computing the Minimum Volume Ellipsoid

◆ Eric B. Howington, University of Alabama; J. Brian Gray, University of Alabama

University of Alabama, 300 Alston Hall, Tuscaloosa, AL 35487-0226

ehowingt@cba.ua.edu

Key Words: *masking, MVE, outlier, robust covariance estimation, robust diagnostics, swamping*

The minimum volume ellipsoid (MVE), introduced by Rousseeuw (1985), is a useful construct for computing robust multivariate outlier diagnostics and robust covariance matrix estimates. Exact computation of the MVE is impractical for all but the smallest datasets. Many algorithms based on the use of elemental subsets have been put forth for approximating the MVE. We review these approaches briefly and propose a new direct-search, genetic algorithm for finding the MVE. Empirical evidence is provided to show the improved performance of this method over existing techniques.

The Onset, Cessation, and Rate of Growth of Loblolly Pines in the FACE Experiment

◆ Susanne Aref, Virginia Polytechnic Institute and State University; David Moore, University of Illinois, Urbana-Champaign; Evan H. DeLucia, University of Illinois, Urbana-Champaign

Virginia Polytechnic Institute and State University, Dept. of Statistics, 401C Hutcheson Hall, Blacksburg, VA 24060

saref@vt.edu

Key Words: *nonlinear regression, segmented model*

The Duke Forest FACE (Free Air Carbon Dioxide Enrichment) experiment was set up to study what happens in a forest when the CO_2 level is elevated as compared to a forest in an ambient environment. A part of the experiment concerns the impact of the elevated CO_2 levels on the onset and cessation of growth of loblolly pine trees (Pinus taeda L.). The response variable was the basal area of about 200 trees in six experimental rings measured approximately monthly from 1996 to 2002. The question was how to determine the time of these events each year. Using nonlinear segmented regression models the rate and the duration of growth, and actual growth were determined from model estimates of onset and cessation of growth. Both the rate of growth and the yearly growth were significantly greater for elevated CO_2 levels (at a 10% level). The magnitude and onset of growth, and the growth period did not appear to differ for the two CO_2 levels, but did differ between categories or sizes of trees as well as between years. Since the growth of a tree is also based on the size of the tree, the relative basal area index was analyzed along with the basal area.

Confidence Intervals for Duan et al.'s Two-step Expenditure Estimation Method in the Presence of Weights

◆ Anna T.R. Legedza, Beth Israel Deaconess Medical Center; Christina Wee, Beth Israel Deaconess Medical Center; Roger B. Davis, Beth Israel Deaconess Medical Center

Beth Israel Deaconess Medical Center, Division of General Medicine and Primary Care, 330 Brookline Ave., Boston, MA 02215

alegedza@bidmc.harvard.edu

Key Words: *Duan et al. two-step method, 1998 MEPS survey, weighted observations, bootstrap confidence intervals*

The policy objective of this work is to examine health care expenditures associated with overweight and obesity using the 1998 Medical Expenditure Panel Survey (n=11,212). We estimated the influence of body mass index on annual health care expenditures using a two-step approach proposed by Duan et al. (1982). Because of the complex sampling frame in which the data was collected, we propose modifications of this approach suited to analysis in SAS-callable SUDAAN. Our chief contributions are bootstrap confidence intervals in the presence of weighting and code to perform this operation. The analyses, interpretations, and conclusions in this work are those of the authors and do not reflect those of the National Center for Health Statistics or the Agency for Healthcare Research and Quality.

Evaluation of the Use of Molecular Markers to Identify and Select Genotypes

◆ Ted B. Bailey, Iowa State University

Iowa State University, Dept. Statistics, 1212 Snedecor Hall, Ames, IA 50010

tbbailey@iastate.edu

Key Words: *molecular markers, simple sequence repeats, Marker-Assisted Selection, Iron-Deficiency Chlorosis, selection efficiency*

An experiment designed to evaluate the efficacy of Marker-Assisted Selection (MAS) for Iron-Deficiency Chlorosis (IDC) in a soybean population was carried out in four environments. Identification of desirable genotypes was based on three simple sequence repeat (SSR) molecular markers that had previously shown to be linked to genes conferring IDC resistance. The effects of three SSR markers was determined using means, a measure of selection efficiency and genotype frequencies. A 2.6-fold selection efficiency relative to selection based on phenotype and up to 73% of lines with superior IDC resistance possessed genotypes favorable for IDC resistance. The effects of specific markers, and marker combinations, were shown to be dependent on environment.

Comparing Treatments in In-vivo Drug-screening Experiments

◆ Leah Martell, Novartis

Novartis, One Health Plaza, 438/3432, East Hanover, NJ 07963

leah.martell@pharma.novartis.com

Key Words: *drug-screening, multiple comparisons, parametric, nonparametric tests*

The in-vivo drug-screening experiments impose a particular set of issues like small sample size, deviation from normality and equal variance, large number of treatments, requiring a tailored approach when comparing different groups. The large body of literature addresses the subject in a more general context which is sometimes not pertinent in this circumstance. Based on a survey of relevant literature I provide a list of applicable strategies, as well as a list of recommendations reflecting the current understanding of the problems. The focus is on all pairwise comparisons for continuous data and independent samples. Examples are taken from different drug treatments with a single endpoint measurement.

Proposal of CT Texture Factor Analysis with Validation of 950Hu Density Mask

◆ Hyun J. Kim, University of California, Los Angeles; David Gjertson, University of California, Los Angeles; James Sayre, University of California, Los Angeles; Jonathan Goldin, University of California, Los Angeles

University of California, Los Angeles, 10833 Le Conte Ave., B3-227R, CHS, Biostatisitics, Public Health, Los Angeles, CA 90095-1721

gracekim@mednet.ucla.edu

Key Words: *texture, factor analysis, density mask*

Density mask is known to be a good CT quantitative structural measurement of chronic disease emphysema. It results from threshold of attenuation present in emphysematous lung. This study proposes the texture factors, which can be more informative in subtle changes than density mask. From eight regions of interest (ROI) (2 arch,2 hemidia of Right and Left), two donuts (axial, peripheral), and three thirds (anterior, middle, posterior) per patient, local texture information was extracted and represented by a 18-dimensional vector that contained statistical moments of the CT attenuation distribution, acquisition-length parameters, and co-occurrence descriptors. Principal component analyses (PCA) and were used to reduce four components such as mean, dispersion, correlation, and percentile components. Factor analyses (FA) were used to reduce to two factors. Multivariate analysis of variance test and Hotelling t test with Bonferroni adjustment were used to validate the two factors using categories of density masks. 140 cases of each ROI resulted as same two factors from PCA and FA. The two texture factors can be used to distinguish severity in emphysema.

Practical Considerations for the Independent Statistical Consultant

◆ Nestor Rohowsky, IDCS, Inc.

IDCS, Inc., 5718 Laurel Ave, La Grange, IL 60525

misterdata@prodigy.net

WITHDRAWN

Key Words: *independent consulting, nonstatistical, flexibility*

Independent statistical consulting can be a very rewarding career. However, in order to maximize your satisfaction from the experience, you, the consultant, must be aware of, understand, and accept the myriad of nonstatistical tasks and influences which directly or indirectly impact on your job performance. Do you have a basic knowledge of accounting and marketing? Are you able to simply express, both verbally and in writing, complex ideas to a nonstatistical audience? Are you willing to take on tasks which you may consider to be “beneath” your level of education and experience? Do you have a realistic idea of the pay rate you should ask for? Many statisticians new to or considering independent consulting have never given these kinds of questions the consideration they are due. This presentation will identify some of the more common nonstatistical aspects of successful independent consulting and offer some practical ideas for dealing with them.

118 Time Series and Point Processes ⌘

IMS

Monday, August 9, 10:30 am–12:20 pm

Estimation in Stationary Markov Renewal Processes with Application to Earthquake Forecasting in Turkey

◆ Enrique E. Alvarez, University of Connecticut

University of Connecticut, Dept. of Statistics, 215 Glenbrook Rd., U-4120, Storrs, CT 06269-4120

ealvarez@merlot.stat.uconn.edu

Key Words: *earthquakes, marked point process, Markov renewal process, stationarity, window-censoring*

Consider a process in which different events occur, with random inter-occurrence times. In Markov renewal processes, the sequence of events is a Markov chain and the waiting distributions depend only on the types of the last and the next event. Suppose that the state-space is finite and that the process started far in the past, achieving stationary. Weibull distributions are proposed for the waiting times and their parameters are estimated jointly with the transition probabilities through maximum likelihood, when one or

several realizations of the process are observed over finite windows. The model is illustrated with data of earthquakes of three types of severity that occurred in Turkey during the 20th century.

Curve Aggregation

Florentina Bunea, Florida State University; Hernando Ombao, University of Illinois; ◆ Anna Auguste, Florida State University

Florida State University, Dept. of Statistics, OSB #214, Corner of Call St. and Woodward Ave., Tallahassee, FL 32306-4330

auguste@stat.fsu.edu

Key Words: *aggregation, Oracle inequalities, time series, nonparametric regression, log spectrum*

We study the problem of estimating the population log spectrum when the data consists in a collection of curves assumed to follow a stationary time series. We show that this can be re-casted as an aggregation problem in nonparametric regression models and we suggest a penalized least squares approach to the problem. The resulting estimator achieves optimal rates of aggregation, in the minimax sense. Our method is then applied to an EEG dataset obtained from patients suffering from depression.

Localized Kullback-Leibler Method for Spectral Estimation

◆ Jianzhong Liu, University of Illinois; Hernando Ombao, University of Illinois

University of Illinois, Dept. of Statistics, 725 S. Wright St., Champaign, IL 61820

jliu7@uiuc.edu

Key Words: *bandwidth selection, localized Kullback-Leibler distance, periodograms, spectrum, generalized additive models*

A consistent estimator for the spectral density of a stationary random process can be obtained by smoothing the periodograms across frequency. The degree of smoothness of many spectra may change over frequency. Thus, it is reasonable to choose a span that also varies across frequency. We will propose a variable span selection method that is based on the Kullback-Leibler distance between the raw and smoothed periodograms. Our approach is to form a localized neighborhood around each frequency; compute the Kullback-Leibler distance and finally select a frequency-specific optimal bandwidth. This criterion, originally developed for use in fitting generalized additive models, utilizes the approximate full local likelihood of periodograms, which asymptotically behave like independently distributed gamma random variables. The resulting span selector is simple and easy to implement. We will present simulation results and analysis of time series dataset.

Semiparametric Regression for the Rate Functions of Recurrent Failure Times

◆ Rajeshwari Sundaram, University of North Carolina, Charlotte

University of North Carolina, Charlotte, Dept. of Math and Statistics, 350-A Fretwell, 9201, University City Blvd., Charlotte, NC 28223

rsundara@uncc.edu

Key Words: *recurrent event, censoring, semiparametric regression, survival analysis, minimum distance, time-dependent covariates*

We discuss the semiparametric regression of the rate function of censored point process associated with the recurrent event times, in the presence of time-dependent covariates. Such models, which do not make any Poisson-type assumption on the point process, are appropriate in modeling recurrent times as they leave the dependence structure among recurrent events unspecified. Inference of such proportional rates and proportional means model have been studied by Pepe and Cai (1993), Lawless and Nadeau (1995) and Lin, Wei, Yang, and Ying (2000). We define a weighted version of the Aalen-Breslow type estimator for the unknown baseline cumulative rate function. By varying the weights, we propose a class of minimum distance estimators for the regression coefficients. These estimators are robust and are easy to compute. The proposed estimators are strongly consistent and asymptotically normal. Moreover, the asymptotic covariance of the regression estimators can be consistently estimated. We make numerical comparison of the proposed estimators with the existing ones.

Bootstrap for Periodic Time Series

◆ Soumendra Lahiri, Iowa State University

Iowa State University, Statistics Dept., Ames, IA 50011

snlahiri@iastate.edu

Key Words: *block bootstrap*

A variant of the block bootstrap method for periodic time series is proposed for situations where the period is large. Consistency of the proposed method for variance and distribution function estimation is discussed. A method for choosing the optimal block size is also presented.

On the Linear Combination of Spatio-temporal Covariances and Variograms

◆ Chunsheng Ma, Wichita State University

Wichita State University, Dept. of Mathematics and Statistics, 319 JB, Wichita, KS 67260-0033

cma@math.twsu.edu

Key Words: *spatio-temporal, covariance, variogram*

The covariance function and the variogram are commonly used to describe the space-time interaction for a process that evolves in both space and time. Although the sum of two covariance functions (or variograms) is also a covariance function (or variogram), this is not necessarily true for the difference or for a linear combination of two covariance functions (or variograms). We discuss the conditions for a linear combination of two covariance functions (or variograms) to be a valid covariance (or variogram), and propose some new spatio-temporal covariance models and variograms.

119 Noninferiority, Superiority, Equivalence, Placebo

Biopharmaceutical Section

Monday, August 9, 10:30 am–12:20 pm

Testing Superiority and Noninferiority Hypotheses in Active Controlled Clinical Trails E

◆ Yi Tsong, U.S. Food and Drug Administration; Juan (Joanne) Zhang, U.S. Food and Drug Administration

U.S. Food and Drug Administration, HFD-705, 5600 Fishers Lane, QMRS, CDER, Rockville, MD 20857

tsong@cder.fda.gov

Key Words: *active control, cross-trial comparison, noninferiority test, superiority test, simultaneous test*

Switching between testing for the hypothesis of superiority and the hypothesis of noninferiority has been an important statistical issue in the design and analysis of active controlled clinical trials. In practice, it is often conducted with a two-stage testing procedure. It has been shown that there is no Type I error rate adjustment required to switch to testing for the hypothesis of noninferiority once the data fail to support the superiority claim. Neither is the adjustment required to switch to testing the hypothesis of superiority once the null hypothesis of noninferiority is rejected. It has been shown to be true in trials with properly pre-specified noninferiority margin in a generalized historical control approach. However, when using a cross-trial comparison approach for noninferiority testing, controlling the Type I error rate may become an issue with this conventional two-stage procedure. We proposed to adopt a single-stage simultaneous testing concept as proposed by Ng (2003) for testing both hypotheses of noninferiority and superiority simultaneously. The Fieller's confidence interval procedure as proposed by Hauschke et al. (1999) is used in the proposed approach.

Statistical Analysis of Noninferiority Trials with a Rate Ratio in Small-sample Matched-pair Designs

◆ Ivan S.F. Chan, Merck & Co., Inc.; Nian-Sheng Taug, Center for Applied Statistics; Man-Lai Taug, Harvard Medical School; Ping-Shing Chan, Chinese University of Hong Kong

Merck & Co., Inc., Clinical Biostatistics, UN-A102, 785 Jolly Rd., Building C, Blue Bell, PA 19422

ivan_chan@merck.com

Asymptotic methods have recently been developed for analyzing noninferiority trials with rate ratios under the matched-pair design. In small samples, however, the performance of these asymptotic methods may not be reliable. We investigate alternative methods that are desirable for assessing noninferiority trials under small-sample matched-pair designs. We propose an exact and an approximate exact unconditional test and the corresponding confidence intervals based on the score statistic. The exact unconditional method guarantees the Type I error rate not to exceed the nominal level, and it is recommended if strict control of Type I error (protection against any inflated risk of accepting inferior treatments) is required. However, the exact method tends to be overly conservative (thus less powerful) and computationally demanding. Through empirical studies, we demonstrate that the approximate exact score method, which is computationally simple to implement, controls the Type I error rate reasonably well and has high power for hypothesis testing. On balance, the approximate exact method offers a very good alternative for analyzing correlated binary data from matched-pair designs with small sample sizes. We illustrate these methods using real examples taken from a crossover study of soft lenses and a Pneumocystis carinii pneumonia study.

Sample-size Calculations for Noninferiority Clinical Trials Based on Poisson Regression Models

◆ Patrick Darken, ALTANA Pharma U.S.; Peter Lane, GlaxoSmithKline

ALTANA Pharma U.S., 210 Park Ave., Florham Park, NJ 07932

patrick.darken@altanapharma-us.com

Key Words: *equivalence trial, noninferiority trial, Poisson regression, sample-size calculation, generalized linear model, power*

Poisson regression is commonly used for the analysis of count data, especially when mainly low counts and zeroes are anticipated. In clinical trials to be analyzed by Poisson regression, a complication for determining sample-size requirements is that when comparing the rates of an event between treatments, differences are usually considered in relative rather than absolute terms. An additional complication is that the occurrence of events from the same subject may be correlated: there is then a need to allow for overdispersion in the calculation of variance estimates, as the data can no longer be treated as arising from a true Poisson process. This can occur, for example, when counting the number of exacerbations of a disease-like asthma experienced by patients during a trial. The generalized linear model provides a convenient framework for these analyses, but care is needed in choosing between Pearson or deviance estimates of overdispersion or using GEE. We present formulae for the calculation of sample sizes for noninferiority clinical trials whose primary outcome will be analyzed using Poisson regression, along with details of the derivation and validation using simulation.

Choosing the Analysis Population in Noninferiority Studies: Per Protocol or Intent-to-treat

◆ M. Matilde Sanchez, Merck & Co., Inc.; Xun Chen, Merck & Co., Inc.

Merck & Co., Inc., RY34-A316, PO Box 2000, Rahway, NJ 07065

matilde_sanchez@merck.com

Key Words: *dropout due to lack of efficacy, hybrid ITT/PP analysis, nontrivial missing, LVCF method, MLE method, noncompliance*

For superiority trials, the intent-to-treat population (ITT) is considered the primary analysis population because it tends to avoid the over-optimistic estimates of efficacy that results from a per-protocol population (PP). However, the roles of the ITT and PP population in noninferiority studies are not clearly defined. A simulation study is conducted to systematically investigate the

impact of different types of missingness and protocol violation on the conservatism or anti-conservatism of ITT and PP analyses in noninferiority trials. The requirement that noninferiority be shown for both population does not necessarily guarantee the validity of a noninferiority conclusion and a sufficiently powered PP analysis is not necessarily powered for ITT analysis. It is important to assess the potential types and rates of protocol deviation and missingness and to obtain some prior knowledge regarding the treatment trajectory of the test treatment versus the active control at the design stage such that proper analysis plan and careful power estimation can be carried out.

Designing an Optimal Futility Analysis for an Equivalence Trial Using Simulation and Historical Data

◆ Jeff Maca, Novartis Pharmaceuticals

Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07869

jeff.maca@pharma.novartis.com

Key Words: *futility analysis, simulation, equivalence, clinical trial*

When designing a clinical trial to study a novel medication or a new formulation of a registered medication, the sponsor might consider including a futility analysis to minimize the risks involved in the study. These include both the unnecessary health risk for patients using an ineffective drug, as well as the risk to the company of continuing to run a trial which will not be able to support the claim desired by the sponsor upon completion of the study. In this case study, the design of a futility analysis was further complicated by the study objective of showing equivalence, and the lack of any strong knowledge of the true nature of the true treatment effect. Other unknown parameters that were expected to affect the futility analysis were a projected high dropout rate and an unknown enrollment distribution. Through the use of historical data and resampling techniques, an estimate of the efficacy profile of the new treatment was developed, and then used to determine the optimal time, endpoint, and criteria used in performing the futility analysis.

Efficacy without a Placebo Group in Mixed Factorial Designs

◆ Charles H. Goldsmith, McMaster University

McMaster University, Centre for Evaluation of Medicines, 105 Main St. East, Level P1, Hamilton, ON L8N 1G6 Canada

Goldsmit@Mcmaster.ca

Key Words: *placebo, efficacy, mixed factorial designs, estimation, osteoporosis, bone loss*

The new Helsinki guidelines for conducting experiments in humans suggest that if there is an effective therapy for a clinical condition, it is unethical to give a placebo to estimate the efficacy of a new therapy. Goldsmith showed that for two- or three-level factorial designs with two or more factors, efficacy can be demonstrated without a placebo group provided the model for the data does not contain any interactions with the new therapy. The two-level and three-level results are extended to designs with at least two mixed level factors, where no interactions with the new therapy permit the estimation of contrasts amongst the levels of the new therapy factor without having a placebo group. Estimable contrasts include efficacy of the new therapy. Suppose the new therapy has two levels: placebo and active, and the effective therapy has three levels: placebo, low dose, and high dose. From the five groups excluding the double placebo, two estimates of new therapy efficacy can be obtained from the nonplacebo levels of the known therapy and pooled to estimate new therapy efficacy. The results will be illustrated using a study with a 2x3 factorial design.

Bias in a Placebo-controlled Study Due to Mismeasurement of Disease Status and Regression Effect

◆ Hung-Mo Lin, Pennsylvania State University; Robert H. Lyles, Emory University; John M. Williamson, Centers for Disease Control and Prevention

Pennsylvania State University, College of Medicine, A210, Health Eval Sci., 600 Centerview Dr., Hershey, PA 17033

hlin@hes.hmc.psu.edu

Key Words: *bias, misclassification, placebo-controlled trial, regression to the mean, intervention*

We raise the concern of whether the use of a placebo group in a randomized clinical trial is sufficient to eliminate bias in the assessment of the effectiveness of a drug when enrollment into the trial prior to intervention requires diagnosis of a dichotomous disease, and the diagnostic test is subject to uncertainty. Due to misclassification and the regression effect, the observed difference in the proportions of diseased individuals between the treatment and placebo groups at follow-up will be equal to the true difference multiplied by the positive predictive value at screening and the difference between the sensitivity and the false-positive value at follow-up. Thus, measurement error of disease status before and after administering the intervention attenuates the intervention effect. We provide the statistical methods required for variance estimation and inference about the intervention effect after correcting for RTM. Validation data corresponding to both the screening and follow-up conditions are necessary to provide additional information on the validity of the diagnostic test.

120 Teaching for a Particular Discipline or Clientele

Section on Statistical Education

Monday, August 9, 10:30 am–12:20 pm

Group Learning, Contextual Projects, Simulation Models, and Student Presentations for Enticing Engineering Statistics Students

◆ Jorge L. Romeu, Syracuse University

Syracuse University, Dpt. Mech., Aeronaut. & Manufac. Eng., 151 Link Hall, Syracuse, NY 13244-1240

jlromeu@syr.edu

Key Words: *statistics education, modeling, data analysis, industrial applications*

Most engineering students take few, if any, statistics courses in their curriculum. And these are often theoretically oriented and packed with material, which does not help raise student interest in the subject. ECS526, “Engineering Statistics” is the main statistics course in our Manufacturing Engineering and Engineering Management MS curricula. ECS526 provides the statistics methods and practice that serve as a basis for all other engineering program courses that use statistics. We discuss our teaching approach using learning groups, contextual projects, discrete event simulation models of real systems, and modern technology, to help bring out student motivation. We discuss the use of student group projects and simulation models to generate, model, and analyze data, within practical engineering applications. We discuss how class topics are divided and assigned to student groups for study, resolution, and presentation to their peers. We discuss course objectives and classroom strategies, testing, grading schemes, software tools used, and results obtained. Finally, we present several examples of student final projects.

Training Environmental Statisticians—Tomorrow’s Problem-solvers

◆ William F. Hunt, Jr., North Carolina State University; Kimberly S. Weems, North Carolina State University; Nagambal Shah, Spelman College; Monica Stephens, Spelman College; Michael T. Crotty, North Carolina State University

North Carolina State University, Dept. of Statistics, Campus Box 8203, Raleigh, NC 27695-8203

whunt@stat.ncsu.edu

How could a win-win strategy be used to train young people in environmental statistics and at the same time analyze environmental data for federal, state and local agencies, that have not been analyzed until now? This presentation will discuss two courses that have been developed to train undergraduate students in environmental statistics and the impact the courses have had. This training comes in support of a National Science Foundation Grant, Collaborative Research: Training Environmental Statisticians Using Complicated Data Sets to Make More Informed Environmental Decisions. A collaborative effort is being undertaken with Spelman College, a historically black college for women in Atlanta, Georgia. This collaborative effort will demonstrate that this approach is portable to other universities and colleges with an undergraduate statistics program and at those without, as long as there are some courses in statistics and a statistician with an interest in environmental statistics. The intent of this collaborative effort is to adapt, modify, and enhance the Environmental Statistics Practicum Program, which was developed at NC State University. The collaborators will demonstrate that the environmental statistics program can be modified, adapted, and enhanced at Spelman College, which represents those colleges without a formal undergraduate statistics program.

A Course in Randomization, Clinical Study Designs, Allocation Schedules, and Interactive Voice Response Systems for Clinical Research Personnel

◆ Gregory T. Golm, Merck & Co., Inc.; Thomas Bradstreet, Merck & Co., Inc.; Laura A. Coffey, Merck & Co., Inc.

Merck & Co., Inc., 785 Jolly Rd., UN-A102, Blue Bell, PA 19422

gregory_golm@merck.com

Key Words: *allocation schedule, statistical education, Interactive Voice Response System*

In most clinical trials, randomized treatment assignment is one of the most important features that enable valid statistical inference. Randomization schedules (allocation schedules for patients and component schedules for Interactive Voice Response Systems) provide the blueprints for carrying out the randomization scheme according to the study design and thus form a nexus between statistical theory and clinical practice. Designing and executing a randomization schedule involves people across a wide variety of job roles, many of whom are not formally trained in all aspects of the randomization process. We have developed a highly interactive course for clinical research personnel that explains both the theory behind the randomization schedule and the practical aspects of its implementation. The course includes group workshops based on a wide variety of real clinical trial examples.

Challenges of Teaching Design and Analysis of Questionnaires to Business Students

◆ Bodapati V.R. Gandhi, University of Puerto Rico, Mayaguez

University of Puerto Rico, Mayaguez, College of Business Administration, Mayaguez, PR 00681

bvrgandhi@hotmail.com

Key Words: *basic statistics, applications, business research methodology, teaching, questionnaire*

This paper contrasts the design and curricula of introductory statistics courses with those of more advanced business research seminars and methodology courses that use statistics as a tool. The gaps between the training provided in basic statistics courses and the required practical applications encountered in survey data analysis are presented. The results of a pilot study of student opinion are discussed. Modifications of curriculum and teaching strategies are suggested.

Creating a Knowledge-centered Learning Environment for the Instruction of Stochastic Processes

◆ Timothy I. Matis, New Mexico State University

New Mexico State University, Dept. of Industrial Engineering, PO Box 30001 - MSC4230, Las Cruces, NM 88003

tmatis@nmsu.edu

Key Words: *stochastic processes, educational materials, visual technology, pedagogy*

This paper describes an innovative approach to the instruction of introductory stochastic processes through the creation of modules that present real-world applications of this material to the students. The modules create a contextual framework for understanding that the students progressively build upon. This work is part of an ongoing project sponsored by the National Science Foundation in which the latest in visual technology and is utlized to produce the modules through collaboration with industrial, academic, and government partners. Prelimnary evaluations are presented that assess the efficacy of this instructional approach.

Professional Masters of Applied Statistics Degree at Penn State

◆ James L. Rosenberger, Pennsylvania State University; Mosuk Chow, Pennsylvania State University

Pennsylvania State University, Dept. of Statistics, 326 Thomas Bldg., University Park, PA 16802-2111

JLR@stat.psu.edu

Key Words: *applied statistics, degree, curriculum, continuing education*

In Fall 2001 Penn State initiated a new graduate degree, the Master's of Applied Statistics. Approximately 15 students per year enter the program and are required to earn 30 credits, including a practicum involving statistical consulting with a capstone project. The program has been attractive to a diverse audience, both as a joint degree program and as a stand alone graduate program. This talk will describe the curriculum and analyze the niche that this program fits into regarding the training of the statistical workforce. A number of students have entered this program as a means to change or enhance their current employment focus. Others have completed this degree along with the PhD in another discipline which utilises statistics. The immediate success of this program suggests that there is an important unmet need in the workforce for appropriate training in applied statistics.

Adjusting for Nested/Hierarchical Measures

◆ Ralph Carlson, University of Texas Pan American; Hilda Medrano, University of Texas Pan American

University of Texas Pan American, 1201 W. University Dr., Edinburg, TX 78541

rcarlson@panam.edu

Key Words: *nested/hierarchical measures, transitive variance*

Psychological and educational measures are often organized by subtests, scales, or factors into nested/hierarchical structures. Measures are nested within a hierarchical structure if each level of one measure is within one and only one level of another measure. On the Wechsler Intelligence Scales a discrepancy of one-half or more standard deviations between the Performance Type of scale and the Verbal Type of scale would yield a Full Scale, which is the third level of hierarchy that is not interpretable due to a lack of "cohesion." When there are significant differences within a lower level of a nested hierarchical measure, all successive levels are not interpretable due to a lack of "cohesion"; therefore, there is a problem with loss of information. The study presents a model that adjusts for nested effects in nested/hierarchical measures. Thus, this model provides a solution by alleviating the loss of information due to a lack of "cohesion" across successive levels of nested/hierarchical measures. Nested variance is transitive and flows from bottom of nested/hierarchical structure to the apex; therefore, adjustment and interpretation of nested effects must be from the bottom up. The current study presents an example of the decomposition and adjustement of nested/hierarchical effects. This study presents a refinement in thinking, analysis, and interpretation of nested/hierarchical measures.

121 Regular Contributed Posters

General Methodology, Biometrics Section, Biopharmaceutical Section, Section on Statistical Computing, Section on Statistics and the Environment, Section on Bayesian Statistical Science, Section on Physical and Engineering Sciences

Monday, August 9, 12:00 pm–1:50 pm

A Comparison of a Blinded Procedure to Estimate Treatment Difference and Standard Deviation with an Unblinded Procedure in Clinical Trials

◆ Farid Kianifard, Novartis Pharmaceuticals; Michael Chen, Novartis Pharmaceuticals

Novartis Pharmaceuticals, Biometrics, U.S. Clinical Development and Medical Affairs, East Hanover, NJ 07936

farid.kianifard@pharma.novartis.com

Key Words: *drug development, folded normal distribution, sample size re-estimation*

In drug development, one often uses the estimated treatment difference and standard deviation of a normally distributed response variable from previous clinical trials to determine the sample size for future trials. Chen and Kianifard introduced a procedure, based on properties of the folded normal distribution, to estimate these parameters without unblinding for an ongoing trial, thereby avoiding bias due to unblinding and potentially shortening drug development time. One may decide to abandon plans for future trials if estimated treatment difference is much smaller than the minimal clinically meaningful treatment difference, or choose to reduce the sample size for future trials if the difference is much larger than expected. In another scenario, most of the planned number of patients may have completed the trial, but patient enrollment is somewhat short of target. This approach can be used to estimate treatment difference and standard deviation for available patients, with the possibility of stopping enrollment. In a simulation study, we compare the power of this procedure to an unblinded procedure for sample size re-estimation in an ongoing clinical trial.

Data-mining From Different Perspectives

◆ Morteza Marzjarani, Saginaw Valley State University

Saginaw Valley State University, 7400 Bay Rd., University Center, MI 48710

marzjara@svsu.edu

Key Words: *databases, statistics, machine learning*

Data-mining is a very powerful tool today in data analysis. Fast and easy access to large datasets allows researchers to use large datasets looking for hidden patterns, and use these patterns for making more accurate decisions or predictions. Data-mining can be viewed from several perspectives. Here, we would like to consider three of such perspectives: the machine learning perspective, the database perspective, and the statistical perspective. In each perspective the emphasis is different. In case of database, since mining is applied to large datasets, efficiency is a major concern. In case of machine learning, datasets are not necessarily large, but

effectiveness is an important issue. Finally, in case of data mining from statistical perspective, validity or accuracy of the mathematical models is of great concern.

A Bayesian Two-stage Phase II Design with Historical Controls

◆ Michael W. Sill, Roswell Park Cancer Institute; Roger L. Priore, Roswell Park Cancer Institute

Roswell Park Cancer Institute, GOG Statistical and Data Center, Elm and Carlton Sts., Buffalo, NY 14263

msill@gogstats.org

Key Words: *Bayesian, design, phase II, clinical trials, sequential, two-stage*

Heitjan (1997) proposed a useful concept called “Bayesian Persuasion Probabilities” for use in phase II clinical trials. Essentially, it involves two competing views of a study agent’s effectiveness. Prior to observing the data, the “skeptic” has a tendency to believe in the null hypothesis that a new drug is ineffective while the “enthusiast” has a tendency to believe in the alternative hypothesis (the drug is effective). According to the author, data from a good study should be able to lead both views to the same conclusion. In practice, however, it would take large sample sizes to reach this consensus with high probability. In the proposed design, we will require the data to persuade the enthusiast to abandon the alternative hypothesis and accept the null hypothesis as plausible before a study can close early. On the other hand, the data at the end of the trial must persuade the skeptic to abandon the null hypothesis and accept the alternative as plausible before the design will recommend further study in a phase III trial. This approach seems to yield designs with better operating characteristics for typical phase II sample sizes.

Exact Analysis of Correlated Binary Response Data—Test of Treatment Effect for Parallel Group Design

Dar Shong Hwang, BRSI; ◆ James S. Lee, Sankyo Pharma; Chyi-Hung Hsu, Novartis Pharmaceuticals

Sankyo Pharma, 399 Thornall St., Edison, NJ 08837

jlee@sankyopharma.com

Key Words: *correlated binary response, parallel group design, multinomial distribution, exact conditional distribution, treatment effect*

For large clinical trials with repeated categorical measurements, current statistical methods tend to apply large sample approximation. The adequacy of the approximation could be questionable in certain situations. An exact analysis for parallel group design having repeated measurements with binary response is proposed. When the binomial samples are independent, exact and asymptotic tests for treatment effect assuming no treatment by stratification interaction are well known. This paper attempts to test the same hypothesis but with correlated binary response data over time. We express again the constant odds ratio in terms of the parameters of the multinomial distribution. Existence and optimality of conditional distributions for the inference on these parameters are investigated. Subsequently, exact conditional distributions are derived. Statistical inference to evaluate if there is a constant odds ratio across time may then be conducted. When there is no treatment by time interaction, the same conditional distributions are further modified to test the treatment effect. The above methodology may be extended from two time points to three or more time points.

A Bayesian Hierarchical Method for Fitting Multiple Health Endpoints in Toxicity Studies

◆ Taeryon Choi, Carnegie Mellon University; Mark J. Schervish, Carnegie Mellon University; Ketra Schmitt, Carnegie Mellon University; Mitchell Small, Carnegie Mellon University

Carnegie Mellon University, Dept. of Statistics, 5629 Hempstead St. Apt. 5, Pittsburgh, PA 15217

tchoi@stat.cmu.edu

Key Words: *Bayesian hierarchical model, perchlorate, dose-response curve, posterior predictive distribution*

Bayesian hierarchical models are built to fit multiple health endpoints from dose-response studies of a toxic chemical, perchlorate. In particular, several models are developed to fit data from the Springborn 90-day study of Springborn laboratory Inc. (1998). We propose empirical models to fit the data based upon a mechanistic model derived from the assumed toxicological relationships between dose and the various endpoints. The model building is compatible with the tentative mode-of-action for perchlorate toxicity proposed by the EPA. We use logistic regression models to estimate the probabilities of histopathology endpoints and multivariate regression models for hormone data. All of the models are estimated in a fully Bayesian framework, and predictions about each endpoint in response to dose are simulated based on the posterior predictive distribution. A hierarchical model that tries to exploit possible similarities between different combinations of sex and sacrifice date allows us to produce more stable estimates of dose-response curves. Finally, we investigate whether or not we can make use of any information from the Caldwell et al. (1995) study using alternative approaches.

Toward Statistically Rigorous Biometric Authentication

◆ Sinjini Mitra, Carnegie Mellon University

Carnegie Mellon University, Dept. of Statistics, 5000 Forbes Ave., Pittsburgh, PA 15213

smitra@stat.cmu.edu

Key Words: *authentication, biometrics, statistical model, spectrum, statistical properties*

The modern world has seen a rapid evolution of the technology of biometric authentication, prompted by an increasing urgency to ensure a system’s security. The need for efficient authentication systems has sky-rocketed since 9/11, and the proposed inclusion of digitized photos in passports shows the importance of biometrics in homeland security today. Based on a person’s essentially unique biological traits, these methods are potentially more reliable and less prone to fraud than traditional methods like PINs and ID cards. Our goal is to build statistically rigorous authentication tools for handling diverse databases where accuracy is imperative. We first examine statistical properties of an existing system and show how they can make it more attractive to users. We then demonstrate that simple statistical models based on the image spectrum can be used to build authentication schemes by using model coefficients as features. An initial simple model is presented as a natural framework and although it yields low error rates, it ultimately needs to be refined for practical authentication of individuals within a larger population. Several of the associated challenges are discussed.

A Correlated Logistic Regression Model for Population Pharmacodynamic Data

◆ John Kwagyan, Howard University College of Medicine; Nnenna Kalu, Howard University College of Medicine; Robert E. Taylor, Howard University College of Medicine

Howard University College of Medicine, Collaborative Alcohol Research Center, Dept. of Pharmacology, Washington, DC 20059

jkwagyan@howard.edu

Key Words: *pharmacodynamics, correlated models, alcohol sensitivity, likelihood estimation*

Dichotomous type population pharmacodynamic data often arise in practice, especially in a clinical pharmacological setting. Data from subjects may consist of a multiple yes or no response to a defined stimulus at a given concentration and/or time. For example, in toxicity studies, the response will be whether the adverse effect occurs or not at a given concentration. In alcohol studies, the response will be whether the individual becomes intoxicated or stimulated at a given concentration or time. There is usually the tendency of certain individuals not responding to the stimulus at every concentration or time. In the study of acute tolerance to alcohol, greater than 50% of subjects reported no subjective intoxication effects at every assessment time. Consequently, in studying the effect of certain stimulus, data that is randomly sampled will lead to individuals that are largely devoid of the outcome of interest. The overall distribution of responsiveness in such data should appropriately be a mixture. We present an adaptation of the disposition model for correlated outcomes to dichotomous population pharmacodynamic data.

Parameter Estimation for a Proposed Joint Distribution of Correlated Bernoulli Trials

◆ Yufeng Li, University of Alabama, Birmingham; Charles R. Katholi, University of Alabama, Birmingham

University of Alabama, Birmingham, 153 Wallace Tumor Institute, 1824 6th Ave. South, Birmingham, AL 35294

yufengli@uab.edu

Key Words: *correlated Bernoulli trial, longitudinal, maximum likelihood estimation, Grey code*

We have proposed a joint distribution for multivariate correlated Bernoulli trial. It combines the advantages of both Generalized Estimate Equation method and the Bahadur's multivariate distribution. This distribution can be applied to repeated measurements of binary outcome on study subjects and studies of family disease. It has potential application with higher dimensions of longitudinal or repeated data analysis. The parameters have been estimated numerically with quasi-Newton method with line search technique both under no restricted maximum likelihood estimation and restricted maximum likelihood estimation. We adopted the Grey Code to calculate scale parameter from log likelihood function of the distribution. Simulation study shows that the estimation for parameters, marginal probabilities and correlation among response are robustness from initial value selected. We have also developed statistical inference by using likelihood ratio test. We can test null hypothesis that all successive probability are equal. We can also test the hypothesis whether the correlation among response have certain type of structure. We have applied this distribution to a real data analysis.

Analyzing Actual and Percentage Change as an Endpoint with or without Baseline as a Covariate in Clinical Trials

◆ Pandurang Kulkarni, Eli Lilly and Company; Yongming Qu, Eli Lilly and Company

Eli Lilly and Company, 11246 Knightsbridge Lane, Fishers, IN 46038-9101

kulkarni@lilly.com

Key Words: *change, percent change, endpoint, clinical trials, sample size*

In many clinical trials from various therapeutic areas, the efficacy of a treatment is assessed using percentage change from baseline as the response variable as opposed to the actual change. Percentage change may be more intuitive to interpret than the actual change. However, choice of endpoint can have important implications. We reported the findings about using percentage change vs. actual change in several clinical trials at the 2002 JSM. Using the actual change from baseline is not always optimal if the within-subject variation is greater than the between-subject variation. We explored the testing aspects of comparing therapies using (a) the actual value, (b) the actual value with baseline as a covariate, (c) change from baseline. Choice of endpoint influences variation, power, sample size, and interpretations, leading to important differences in understanding of treatment effects and the underlying disease mechanisms. These factors will be discussed and guidance will be provided, in the context of a regulatory environment, in order to better understand the issues and to help optimize the design of clinical trials.

Multiple Imputation Techniques in Small Sample Clinical Trials

◆ Sunni A. Barnes, Mayo Clinic; Stacy R. Lindborg, Eli Lilly and Company; John W. Seaman, Baylor University

Mayo Clinic, 200 First St. SW, Rochester, MN 55905

sbarnes@mayo.edu

Key Words: *multiple imputation, LOCF, missing data, MAR, Bayesian least squares*

Clinical trials allow researchers to draw conclusions about the effectiveness of a treatment. However, the statistical analysis used to draw these conclusions will inevitably be complicated by the common problem of attrition. Resorting to ad hoc methods such as case-deletion or mean imputation can lead to biased results, especially if the amount of missing data is high. Multiple imputation, on the other hand, provides the researcher with an approximate solution that can be generalized to a number of different datasets and statistical problems. Multiple imputation is known to be statistically valid when n is large. However, questions still remain about the validity of multiple imputation for small samples in clinical trials. We investigate the small-sample performance of several multiple imputation methods, as well as the last observation carried forward method. We also introduce a nonparametric multiple imputation procedure that incorporates more information than those currently in use and does so for less computational expense.

Analysis and Characterization of In-vitro Action Potential Duration Assays

◆ Chao-Min L. Hoe, Merck & Co., Inc.; Lori Mixson, Merck & Co., Inc.; Spencer Dech, Merck & Co., Inc.; Joseph Salata, Merck & Co., Inc.; Peter Siegl, Merck & Co., Inc.

Merck & Co., Inc., WP 37C-305, PO Box 4, West Point, PA 19486-0004

hoe@merck.com

Key Words: *repolarization, action, potential, duration*

The ventricular repolarization is determined by the physiological process—duration of the cardiac action potential. Delayed repolarization could cause ventricular tachycardia, torsade de pointes, and lethal arrhythmic. Any non-antiarrhythmic pharmaceutical that increases the cardiac action potential duration should be considered to pose a risk to humans. In-vitro electrophysiology studies with one negative control, one positive control and two compounds on multicellular guinea pig papillary muscle were conducted in order to investigate the effect of test substances on action potential duration. In these studies, the action potential of a single cell was measured by six different transmembrane voltage frequencies. Values of action potential duration (APD) at specific degrees of repolarization (e.g., APD90 means APD at 90% repolarization), resting membrane potential (RMP), action potential amplitude (APA) and maximum rate of depolarization (Vmax) were recorded. Muscle tissues from six guinea pigs were exposed to the test articles in a dose-escalating study, followed by a wash-out period and exposure to the positive control. A mixed effects model was fit to the data.

Tree-based Models for Predicting Future Domestic Violence Incidents

◆ Yan He, University of California, Los Angeles

University of California, Los Angeles, 461 N. Euclid Ave. #203, Dept. of Statistics, Pasadena, CA 91101

yanhe@stat.ucla.edu

Key Words: *classification and regression trees (CART), domestic violence (DV), missing imputation, RandomForest, variable importance*

The overall goal of this paper is to develop a short screening instrument that will help predict future domestic violence incidents and their seriousness. A key rationale is that a dangerous situation currently exists in the household and is likely to continue in the future. The main tools to be used for inspecting future incidents are classification and regression trees, bagging, and RandomForest. The effects of changing priors and costs are studied. Variable importance is measured. One important feature of this data is that the data is based on designed questionnaire and follow-up investigation, thus some predictor variables have significant portion of missing values. Various missing imputation methods are carefully studied and the classification results are compared.

An Application of Cluster Analysis to a Clinical Trial

◆ Kao-Tai Tsai, Aventis Pharmaceuticals

Aventis Pharmaceuticals, 300 Somerset Corp. Blvd., Bridgewater, NJ 08807

kao-tai.tsai@aventis.com

Key Words: *projection pursuit clustering, K-mean clustering, classification trees*

In typical clinical trials, patients often have different specific profiles and therefore have different responses to the medical treatments. It would be easier to provide the proper medical treatments if one can relate the patient profiles with the efficacy of various treatment regimens. Hence, grouping or segmentation of patients has gaining increasing attention in medical research. Using a recent clinical trial data, we try to identify subgroups of the patient population using various clustering procedures including the classical K-mean analysis and more recent projection pursuit clustering as well as the classification trees.

Exact Analysis of Correlated Binary Data Repeated Measures/RBD Design

◆ Dar Shong Hwang, BRSI; Chyi-Hung Hsu, Novartis Pharmaceuticals

BRSI, PO Box 562, Florham Park, NJ 07932

dshwang@optonline.net

Key Words: *correlated binary data, repeated measurement design, randomized block design, multinomial distribution, exact conditional distribution, optimality*

A parametric multinomial distribution was used to describe each of the patients having repeated binary response data. The hypothesis of equal response rates is re-expressed in terms of a subset of multinomial parameters. Existence and optimality of conditional distributions which depend on these parameters were presented. The exact distributions were then derived for the case of three or more time points which appears to be easier to follow. Reduction to the familiar Freeman-Halton (F-H) exact test for the comparison of k independent binomials was proven. Under Ho, the null distribution reduces to that of F-H test when applied to a RBD design. Hence, computation of Type I error and approximation by Mantel-Haenszel (M-H) procedure using existing software is shown. Application to D.R. Cox's exact test of dose-response relationship is shown. Generalization to multicategories of nominal response data per time point resulted in "Generalized" Cochran's Q test, with relationship to existing tests, software, and M-H approximation as in the binary case. The approach is of the population model type, not the randomization model type, having the advantage of inference.

Power and Sample Size Calculation for Comparing Early Viral Decline Using Bootstrap Approach

◆ Xiaoping S. Hu, Roche Laboratories, Inc.

sylvia.hu@roche.com

Key Words: *power, sample size, bootstrap*

Early post treatment HCV viral decline is a good predictor for the treatment success. Comparing treatment difference in mean HCV RNA change from baseline or slop of decline in HCV RNA are of interest. However, the post-treatment HCV RNA is not normally distributed. There are at least two main different treatment response patterns, one group of patients respond immediately, whose viral load immediately go down to undetectable or near undetectable. Another group of patients response to the treatment poorly. These patients' HCV RNA decline slightly or even do not change. Therefore, the post-treatment viral load has a double mode distribution. The usual power and sample size calculation formulae do not hold. Bootstrap approach was used to calculate the power and sample size for testing mean difference and slope difference between two treatment. An example from real clinical trial data was given to illustrate the method and the outcome.

Regression Neural Nets and Trees for Modeling Factors Affecting Cotton Yield

◆ Andy Mauromoustakos, University of Arkansas

University of Arkansas, 101 AGRX, AGRI STAT LAB, Fayetteville, AR 72701

amauro@uark.edu

Key Words: *regression, neural nets, trees*

Meloidogyne incognita and Thielaviopsis basicola population densities and influential soil factors were evaluated in 2001 and 2002 in a natural infested field to determine their effects on 2002 yield using various statistical models. A 15-acre field was subdivided into 512 plots (100ft x 4 rows) and sampled for M. incognita in April, May, July, and October; T. basicola in April and October; and soil fertility and soil texture in April. Telone II was applied at 0, 1.5, 3, and 4.5 gal/acre for nematode control comparisons for both years. Plots were aggregated and averaged (from 512 to 40) based on four soil textural ranges (0-30% sand, 31-45% sand, 46-60% sand, and >60% sand) and the aforementioned Telone II applications. Statistical analyses were conducted using JMP SAS software. Statistical models fitted to the data included multiple regression, neural networks, and partitions trees. The data variables used by all three models included percent sand, Telone II, M. incognita (April 2002), and T. basicola (October 2001). Results of all three models explained at least 89% of the variability of the 2002 yield.

Comparison of Sequential and Multistage Designs for Selecting Promising Therapeutic Modalities for Stroke in Preclinical Experiments

◆ Paula K. Roberson, University of Arkansas for Medical Sciences; William C. Culp, University of Arkansas for Medical Sciences

University of Arkansas for Medical Sciences, Biostatistics; Slot 781, 4301 W. Markham, Little Rock, AR 72205

robersonpaulak@uams.edu

Key Words: *sequential methods, multistage designs, bias, preclinical experiments*

During preclinical development, it may be important to screen a number of competing alternative therapies simultaneously, identifying a small subset of sufficient promise for further intensive and costly development. In the case of binary outcomes, the goal is to maximize the likelihood of identifying procedures with high success probabilities while minimizing the chances of continuing investment in procedures with low success rates. For cost efficiencies, it is desirable to minimize the number of observations required to identify inferior therapies, while concurrently obtaining reasonable estimates of therapeutic success in study arms deemed adequate, to aid planning of future experiments. This poster will detail our development of such a plan as part of a developmental program of stroke therapy. Expanding on randomized phase II clinical trials designs, the experiments were designed as randomized concurrent single arm studies, rather than a single comparative study. Results of a simulation study to assess Type I/II errors, as well as estimation bias and variability, of sequential and multistage designs will be presented, including rationale for the design ultimately chosen.

Checking Model Assumptions—Getting the BUGS Out

◆ Steven A. Gilbert, Rho, Inc.

Rho, Inc., 199 Wells Ave., Suite 302, Newton, MA 02459

sgilbert@rhoworld.com

Key Words: *BUGS, clinical trial*

Clinical trials destined for FDA submission are generally analyzed in a frequentist framework. It is not uncommon to find that modeling assumptions are violated, as indicated by the presence of skewed or heavy-tailed residuals, nonlinear relationships, or outlying values. Unfortunately, it is often difficult to explore the extent to which the violations of the assumptions adversely affect the inferences drawn from the frequentist models. However, a Bayesian analysis that accounts for some of the unmet modeling assumptions can be performed using WinBUGs, and the results can be compared to the inferences generated using standard analyses. We examine several datasets based on actual clinical trials, and explore the practical aspects of combining the frequentist and Bayesian approaches.

122 Section on Statistics in Sports Speaker Luncheon (Fee Event)

Section on Statistics in Sports

Monday, August 9, 12:30 pm–1:50 pm

Simeon Poisson Meets the Fastest Game on Ice

◆ Robin Lock, St. Lawrence University

St. Lawrence University, Dept. of Mathematics, Computer Science, and Statistics, Canton, NY 13617

rlock@stlawu.edu

Key Words: *sports ratings, ice hockey, Poisson model*

Recent controversy over computer models in the BCS system of ranking NCAA football teams has focused interest on statistical methods for assessing and predicting team performance. In recognition of meeting in Toronto (home of the Hockey Hall of Fame) we consider a multiplicative Poisson model for scoring in the game of ice hockey. The model is used to rate and rank teams while accounting for varying schedule strengths, predict future game outcomes, and calculate probabilities of game events. We assume that the expected scoring rate in a game depends on a team's offensive ability, the defensive ability of its opponent, and an adjustment for home ice advantage. Relevant parameters are estimated from previous game scores. We examine the performance of this model using college hockey data on men's and women's teams playing at the NCAA Division I level and compare results to other ranking systems such as the ratings percentage index (RPI), pairwise comparisons, and media polls.

123 Biopharmaceutical Section Roundtable Luncheons (Fee Event)

Biopharmaceutical Section

Monday, August 9, 12:30 pm–1:50 pm

Statistical Issues in Drug-Interaction Studies and the Use of These Studies in the Labels

◆ Yonghua Wang, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, PO Box 4000, Princeton, NJ 08543-4000

yonghua.wang@bms.com

There are different ways to design drug-interaction studies. There are different ways to analyze data from such studies even when the designs are the same. There are questions on whether or not to power or how to power such studies. There are questions on how the above mentioned issues will affect the labels of the drugs in such studies. Our discussion will be around the above issues in drug-interaction studies.

Diagnostic Devices

◆ Lakshmi Vishnuvajjala, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1350 Piccard Dr., HFZ-542, CDRH, Rockville, MD 20850

rlv@cdrh.fda.gov

Key Words: *diagnostic, CDRH, monitors, test kits*

Diagnostic devices comprise over a third of all device submissions to the center for Devices and Radiological Health (CDRH) at the Food and Drug Administration (FDA). These include in vitro diagnostic tests like tests for cholesterol, PSA, hepatitis, and glucose and tests for genetic markers and cancer markers, and home test kits for various conditions. Also included are radiological devices like mammography and x-ray, and other diagnostics devices that use ultrasound and autofluorescence. There are diagnostic devices in every area from fetal monitors to cardiac monitors, and fluorescence to detect cervical cancers and throat infections. The methodologies are different from those used for therapeutic devices, and emerging technologies present particular challenges to this area of devices.

Issues in Megatrials

◆ David H. Henry, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, 1095 Cornell Dr., Yardley, PA 19067

david.henry@bms.com

Key Words: *clinical trials, multicenter, megatrial*

Very large multicenter clinical trials (megatrials) pose many interesting and challenging issues for statisticians. These include not only statistical concerns, but also operational, data, and regulatory problems. We will identify pitfalls associated with each of these areas and discuss how statisticians have addressed these issues.

Biomarkers and Their Validation in Pharmacogenomics

◆ Sue-Jane Wang, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857

wangs@cder.fda.gov

Key Words: *biomarker, pharmacogenomics, pharmacogenetics*

The recent advances in genomics, such as gene expression and single neucleotide polymorphisms, and proteomics, such as, protein expression, opened the door for drug trials to investigate adaptively a potential single biomarker or composite biomarkers measured using noninvasive procedures for pharmacologic response and/or toxicity classifications. We shall discuss exploration of potential, probable biomakers and statistical/biological validations in the pharmacogenomics/pharmacogenetics studies when they are used as a scientific finding and/or for future drug trial design and evaluation of safety concern.

Statistical Issues in Review of Carcinogenicity Using Transgenic Mice

◆ Karl K. Lin, U.S. Food and Drug Administration

U.S. Food and Drug Administration, DB II (HFD-715), CDER, FDA, 5600 Fishers Lane, Rockville, MD 20857

link@cder.fda.gov

Key Words: *ICH Guideline S1B, group size, latency and multiplicity effects*

The high cost and long time needed to conduct a standard long-term in-vivo carcinogenicity study, and the increased insight into the mechanisms of carcinogenicity due to the advances made in molecular biology have led to the use of transgenic mice as alternative in-vivo approach to the assessment of carcinogenicity. People also feel that genetically altered mice can be better animal surrogates for human cancer because they carry some specifically activated oncogenes that are known to function in both human and animal cancers. The ICH Guideline S1B allows sponsors to conduct one long-term rodent carcinogenicity study together with a short-term test using transgenic rodents. More and more drug companies are following the ICH guideline and conducting carcinogenicity studies using transgenic mice to replace the traditional two-year mouse studies. There are statistical issues such as group size, methods of test of carcinogenicity of new drugs, adjustment for multiplicity etc., in this new preclinical field of drug safety assessment. A recent development in this area of statistical review of carcinogenicity studies within CDER/FDA will be discussed.

124 Section on Bayesian Statistical Science Roundtable Luncheons (Fee Event)

Section on Bayesian Statistical Science

Monday, August 9, 12:30 pm–1:50 pm

Teaching Bayesian Thought to Nonstatisticians

◆ Dalene Stangl, Duke University
Duke University, Box 90251, ISDS, Durham, NC 27708
dalene@stat.duke.edu

Key Words: *pedagogy, Bayes theorem*

Bayesian methods are now commonly used in many fields including medicine, epidemiology, and social and natural sciences, When explaining Bayesian concepts to nonstatisticians in these fields, where should you begin? How much foundational material do they need to know? Is subjective probability on the list of concepts to be explained? How about conditional probability? How about parameters as random variables? What concepts are crucial and which are not? What resources are available to help explain the key concepts? What works and what does not work when teaching nonstatisticians with little mathematical training? These are a few of the questions that will be addressed at this roundtable discussion.

Bayesian Spatial Statistics

◆ Alan E. Gelfand, Duke University
Duke University, Institute of Statistics and Decision Science, Old Chemistry Bldg., Durham, NC 27708-0251
alan@stat.duke.edu

Key Words: *areal unit data, point-referenced data, hierarchical modeling, Bayesian computation*

With the dramatic increase in collection of spatial and spatiotemporal data, the need for more sophisticated modeling to analyze such data arises. Hierarchical models, implemented within the Bayesian framework, become attractive. Topics for discussion will include: modeling strategies to incorporate spatial features, spatial process models vs. Markov random field models, dynamic modeling for spatio-temporal data, multivariate spatial-process modeling, and computational issues.

125 Section on Government Statistics Roundtable Luncheon (Fee Event)

Section on Government Statistics

Monday, August 9, 12:30 pm–1:50 pm

Protecting Confidentiality by Releasing Simulated Microdata: A Discussion of Benefits, Limitations, and Practical Experiences

◆ Jerome Reiter, Duke University
Duke University, ISDS, Box 90251, Durham, NC 27708
jerry@stat.duke.edu

Key Words: *confidentiality, disclosure, multiple imputation, synthetic data*

Given the proliferation of information on individuals available in external databases, and the improvements in record linkage technologies, it is not inconceivable that statistical agencies and data collectors will be unable or unwilling to release genuine microdata in the near future. One solution is to release simulated data constructed so as to preserve many of the relationships in the real data but not reveal any confidential values. We will discuss at a nontechnical level the foundations of this approach, focusing on its benefits and limitations as compared to other disclosure limitation methods. We also will discuss practical experiences and issues with releasing simulated microdata. The goal of the discussions is to introduce statisticians in federal agencies or organizations that release public use data to this promising new approach to disclosure limitation. The lunch organizer, Jerry Reiter, is an assistant professor at Duke University. He has consulted with the U.S. Bureau of the Census on methods for generating and analyzing simulated microdata. He has published several articles on the topic in *Journal of Official Statistics*.

126 Section on Health Policy Statistics Roundtable Luncheons (Fee Event)

Section on Health Policy Statistics

Monday, August 9, 12:30 pm–1:50 pm

Data on Patient Safety

◆ Richard R. Carlson, HealthPartners
HealthPartners, Care Systems Improvement MS 21103G, PO Box 1309, Minneapolis, MN 55440-1309
rick.r.carlson@healthpartners.com

Key Words: *patient safety, health care*

Recent Institute of Medicine publications, “To Err is Human” (2000) and “Crossing the Quality Chasm” (2001), highlighted patient safety deficiencies in the health care system. In the course of elaborating on the issue of patient safety, the need for effective and efficient measurement is highlighted. “Crossing the Quality Chasm” states “existing literature does not allow a comprehensive estimate of the burden of harm due to poor quality. The literature on health care quality covers only a portion of the full range of quality concerns.” This roundtable will discuss ideas for measuring patient safety in health care, including both inpatient and outpatient settings.

Generalizability Theory

◆ Joseph C. Cappelleri, Pfizer Inc.

Pfizer Inc., Eastern Point Rd. (MS8260-2222), Groton, CT 06340

joseph_c_cappelleri@groton.pfizer.com

Key Words: *variance components, reliability, measurement error, rating scales, psychometrics, analysis of variance*

Generalizability theory is a statistical theory about the dependability of behavioral measurements. Dependability refers to the accuracy of generalizing from a person’s observed score on a measure (e.g., behavioral observation, opinion survey, rating scale) to an estimate of her ideal score that would have been received over a defined universe. Generalizability theory liberalizes classical theory by employing analysis-of-variance methods that allow a researcher to untangle multiple sources of error variance, rather than keeping error variance undifferentiated as in classical theory. Although the theory is now over 30 years old, with roots in psychological and educational research, it is only in the last several years that generalizability theory has sprouted in the health sciences. This roundtable will provide an introduction to generalizability theory, discuss its concepts and usefulness, and highlight its practical application in health measurement scales.

Propensity Scores: Helping Nonstatisticians Get the Message

◆ Thomas E. Love, Case Western Reserve University

Case Western Reserve University, Center for Health Care Res. and Policy, 2500 MetroHealth Dr., Cleveland, OH 44122

thomaslove@case.edu

Key Words: *health policy, consulting, propensity score, health services research, statistical education, observational study*

There is an increasing call for statisticians to help justify causal inferences from observational or quasi-experimental data, especially in health policy studies. As a result, propensity score (and related) methods are increasingly part of the statistician’s daily lives. Yet these techniques are new to many of our “customers.” At this roundtable, we will discuss effective ways to describe an observational study’s results, assumptions, caveats and conclusions accurately and usefully for a clinical or policy-oriented audience, focusing on propensity score matching and regression-adjustment methods.

127 Section on Physical and Engineering Sciences Roundtable Luncheons (Fee Event)

Section on Physical and Engineering Sciences

Monday, August 9, 12:30 pm–1:50 pm

Designs on the Web

◆ John P. Morgan, Virginia Polytechnic Institute and State University; Leonard Soicher, Queen Mary, University of London

Virginia Polytechnic Institute and State University, Dept. of Statistics, Blacksburg, VA 24061

jpmorgan@stat.vt.edu

Key Words: *block design, optimal design, web resource*

Where do you go to find a good design for a comparitive experiment? Textbooks contain limited tables of standard designs, most commonly BIBDs and less frequently lists of lattices, PBIBDs, and members of generalized cyclic families. While the last 30 years have seen tremendous progress in finding optimal designs, corresponding tables, when they exist, are scattered throughout the literature. Design search software is likewise limited, with inability to allow a multicriteria approach to design selection and no guarantee that an optimal design will be produced. In 2001 the United Kingdom Engineering and Physical Sciences Research Council awarded a project titled, “A Web-based Resource for Design Theory.” From this award is growing the Design Theory Resource Server (DTRS) for combinatorial and statistical designs at www.designtheory.org. DTRS is developing (i) a database of designs, including serving as a repository of designs as they continue to be discovered, (ii) software for constructing and manipulating designs, and investigating their statistical and combinatorial properties, (iii) documentation which will explain the types of design, their uses, and more.

128 Section on Quality and Productivity Roundtable Luncheon (Fee Event)

Section on Quality and Productivity

Monday, August 9, 12:30 pm–1:50 pm

Sharing Secrets for Effective Use of Six Sigma in Development and Market Growth

◆ Kymberly K. Hockman, DuPont Company

DuPont Company, Experimental Station Bldg. 268/107, Rte. 141 and Henry Clay, Wilmington, DE 19880-0268

kymm.k.hockman@usa.dupont.com

Key Words: *design for Six Sigma, marketing, development, Six Sigma, sales, growth*

Use of Six Sigma has expanded into the arenas of development of all kinds and strategic marketing. Come share your challenges and successes, tools and trials with colleagues with similar goals. Networking that can occur at roundtable luncheons often results in breakthrough ideas for participants. Come chew your food and digest some new thoughts with us. Do the same tools and rules apply? What kind of data are we using? Does anything really work? YES, and come see how!

129 Section on Statistical Graphics Roundtable Luncheon (Fee Event)

Section on Statistical Graphics

Monday, August 9, 12:30 pm–1:50 pm

Large Dataset Visualization in Bioinformatics

◆ Vadim Kutsyy, Cytokinetics, Inc.

Cytokinetics, Inc., 280 E. Grand Ave., South San Francisco, CA 94022

vadim@kutsyy.com

Key Words: *large data, visualization, bioinformatics*

Most of the data analysis in bioinformatics today is done on large datasets. This creates interesting statistical and computational problems. However, in addition to analyzing the data, it is very important to visualize it during the discovery process. Visualization of the data can help statistician to understand underlying structure and perform better analysis. Also, it can help to present data findings back to the scientists and let them explore data on their own. We will discuss statistical and computational aspects of large dataset visualization in bioinformatics. In addition, we will discuss tools and techniques on how to make visualization user-friendly and understandable.

130 Section on Statistics and the Environment Roundtable Luncheons (Fee Event)

Section on Statistics and the Environment

Monday, August 9, 12:30 pm–1:50 pm

The Statistics of Biodiversity and Extinction

◆ Philip Dixon, Iowa State University

Iowa State University, Dept. of Statistics, 120 Snedecor Hall, Ames, IA 50011-1210

pdixon@iastate.edu

The popular press is full of stories about a biodiversity crisis. The numbers vary, but one common one is that a 1,000 species are lost each year. What is the scientific basis for these estimates? Where do they come from? We will examine the source of some popular "estimates" of the total number of species in the world and the current extinction rate, then discuss alternative methods to estimate these numbers.

Practical Aspects of Designing an Environmental Monitoring System

◆ Richard L. Smith, University of North Carolina

University of North Carolina, Dept. of Statistics and OR, Chapel Hill, NC 27599-3260

rls@email.unc.edu

Key Words: *spatial statistics, network design, environmental monitoring*

There is by now much theoretical research on different aspects of experimental design as they apply to the design of spatial networks that are used in applications such as monitoring environmental pollutants. But what are the practical aspects of such designs? Are the theoretical approaches to network design useful in designing real networks? How do statistical criteria such as minimizing prediction error variances fit in with the regulatory and political aspects of designing a network? How much do regulatory organizations such as the U.S. Environmental Protection Agency use statistical criteria in designing their networks? And how could the expertise of statisticians best be applied to help the USEPA and other similar organizations to design better networks?

131 Section on Statistics in Epidemiology Roundtable Luncheon (Fee Event)

Section on Statistics in Epidemiology

Monday, August 9, 12:30 pm–1:50 pm

Dealing with Missing Data in Longitudinal Studies

◆ Geert Molenberghs, Limburgs Universitair Centrum

Limburgs Universitair Centrum, Universitaire Campus, Building D, Diepenbeek, B-3590 Belgium

geert.molenberghs@luc.ac.be

Key Words: *missing data, longitudinal data, randomized clinical trial, observational study, ignorability*

Incomplete data is omnipresent in longitudinal studies, especially with human subjects. Over the last decades, a number of strategies have become the norm to deal with this problem, including complete case analysis and last observation carried forward. More recently, a wide variety of more advanced missing data handling methods have become available, together with software tools in standard statistical packages. At the same time, increasing concern has been expressed about the appropriateness of the simpler strategies. Therefore, it looks like a paradigm shift is necessary from the simpler methods to, for example, likelihood- (or Bayesian) based ignorable analyses, perhaps supplemented with well-chosen sensitivity analysis tools.

132 Section on Survey Research Methods Roundtable Luncheons (Fee Event)

Section on Survey Research Methods

Monday, August 9, 12:30 pm–1:50 pm

Projecting the 2004 Elections: What's Changed?

◆ Warren J. Mitofsky, Mitofsky International

Mitofsky International, 1776 Broadway, Suite 1708, New York, NY 10024

mitofsky@mindspring.com

Key Words: *election projections, presidential election, television networks, 2004 election*

In 2000, projections of the presidential race by the television networks went astray. After congressional hearings, scrutiny by the media and academics and technical review by expert statisticians changes were planned, but were not successfully used by the networks on election night 2002. Now that we are in the midst of a new presidential election year, what will be different when we elect a president on November 2, 2004? There have been changes in methods and approach that will be discussed.

IRBs, Surveys, and the Protection of Human Subjects

◆ Eleanor Singer, University of Michigan

University of Michigan, Institute for Social Research, Box 1248 University of Michigan, Ann Arbor, MI 48106

esinger@isr.umich.edu

Key Words: *IRBs, human subjects, federal regulations*

This roundtable will discuss recent recommendations for IRB review of surveys made by a National Academy of Sciences panel. It will also discuss problems participants have encountered with IRBs, and consider successful strategies for dealing with them. The Chair will review recommendations issued by the Panel on Institutional Review Boards, Surveys, and Social Science Research in 2003. To enliven the discussion, roundtable participants are urged to send in problems they have encountered with IRB review of research (as well as solutions!) before the meetings.

133 Social Statistics Section Roundtable Luncheon (Fee Event)

Social Statistics Section

Monday, August 9, 12:30 pm–1:50 pm

Rolling Out the American Community Survey

◆ Deborah H. Griffin, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Room 3066-3, Washington, DC 20233-7000

deborah.h.griffin@census.gov

Key Words: *demographics, decennial census, quality, data*

Since 1994, the Census Bureau has been designing, testing, and developing the methods needed to collect detailed demographic, social, economic, and housing data in an annual survey, rather than as part of the decennial census. When fully implemented, the American Community Survey or ACS, will sample 3 million addresses each year. Using rolling samples, the ACS will accumulate multiple years of sample to produce data for the lowest levels of geography that were historically served by the decennial census long form. In July 2004, the ACS is scheduled to begin "ramping up" from a national sample of about 67,000 addresses each month to a monthly sample of about 250,000 addresses each month. The ACS will become the largest demographic survey ever fielded by the Census Bureau. In 2005, the ACS will be in full production, collecting data from all counties in the nation as well as Puerto Rico. This session will discuss the readiness of the ACS for full implementation and the current plans for data products. Issues related to the shift from the decennial census to the ACS, including quality concerns and the need to educate data users will be addressed.

134 Section on Teaching of Statistics in the Health Sciences Roundtable Luncheon (Fee Event)

Section on Teaching Statistics in the Health Sciences

Monday, August 9, 12:30 pm–1:50 pm

Teaching Intermediate-level Biostatistics to Clinicians Using General Statistical Software Packages

◆ William D. Dupont, Vanderbilt University School of Medicine

Vanderbilt University School of Medicine, Dept. of Biostatistics, S-2323 Medical Center North, Nashville, TN 37232-2158

william.dupont@vanderbilt.edu

Key Words: *teaching statistics in the health sciences, multiple regression methods, intermediate-level biostatistics, statistical software*

The development of easy-to-use general-purpose statistical software packages increases the range of topics that can be taught to biomedical researchers with limited backgrounds in statistics or mathematics. The range of topics that can be effectively taught to this audience will be discussed and experiences shared. We will also debate the strengths and weaknesses of the different software packages that can be used for this instruction. Packages that will be discussed include Stata, SPSS, R, and SAS.

135 Section on Statistics in Defense and National Security Roundtable Luncheons (Fee Event)

Section on Statistics in Defense and National Security

Monday, August 9, 12:30 pm–1:50 pm

Monitoring and Responding to Health Effects of Terrorism

◆ Douglas A. Samuelson, InfoLogix, Inc.

InfoLogix, Inc., 8711 Chippendale Ct., Annandale, VA 22003

dsamuel@seas.upenn.edu

Preparedness for terrorism in the health services raises a number of interesting quantitative policy-related questions, including effects of bioterror attacks and alerts on emergency response capacity and on epidemiology; effects of mass casualties on emergency facilities; optimal positioning of resources to meet threats; appropriate alerts and warnings; mental health and family effects, some of which have long lag times from traumatic event to first report of symptoms; and re-examination of the proper balance between privacy and needed reporting.

136 Section on Statistical Education Roundtable Luncheons (Fee Event)

Section on Statistical Education

Monday, August 9, 12:30 pm–1:50 pm

Preparing Students for Graduate School in Statistics

◆ L. Marlin Eby, Messiah College

Messiah College, PO Box 3041, One College Ave., Grantham, PA 17027

eby@messiah.edu

Key Words: *undergraduate, graduate, mentoring*

Successful preparation usually begins early in the undergraduate program. Topics to be discussed include creating student interest in statistics as a discipline and a career, identifying students with graduate school potential, the role of mentoring, relevant courses and internships, guiding students in choosing graduate programs, and followup with students after they enter graduate school. The host will draw on his successful experiences as a statistician within the mathematical sciences department at a college of the liberal and applied arts and sciences.

Teaching Statistics Online

◆ Robert Gould, University of California, Los Angeles

University of California, Los Angeles, MS 8130,MC 155404, Dept. of Statistics, Los Angeles, CA 90095-1554

rgould@stat.ucla.edu

Key Words: *distance learning, blended instruction, online, AP Statistics, active learning*

Online courses—whether intended as a distance learning course or as part of a blended instruction program—are becoming increasingly popular. There are a variety of challenges involved in translating statistics education pedagogy into an online format. For example, how to encourage constructive discussions? How to design collaboration-friendly environments? How to create activities that develop understanding? The luncheon host has experience with these issues through the INSPIRE program, an online program to teach statistics to first-time AP Statistics teachers.

Save the Equation for Last

◆ Steve C. Wang, Swarthmore College

Swarthmore College, Dept. of Mathematics and Statistics, 500 College Ave., Swarthmore, PA 19081

scwang@swarthmore.edu

Key Words: *teaching, regression, pedagogy, education*

When introducing a new topic in an introductory statistics course, we often begin by giving a formula and then explaining what the formula measures and how it measures it. We will discuss an alternate approach: saving the formula for last. Instead of beginning with a formula, one can stress the concepts or intuition first, and only later give a specific equation. I will demonstrate this approach using a series of questions on the conditional standard deviation in a simple linear regression setting. These questions would be trivial to solve with a formula, but without the formula they encourage students to think about the meaning of the conditional standard deviation and how it relates to the marginal standard deviation.

137 Quantile Regression

Business and Economics Statistics Section, Section on Quality and Productivity

Monday, August 9, 2:00 pm–3:50 pm

Quantile Regression Methods for Recursive Structural Equation Models

◆ Roger Koenker, University of Illinois

University of Illinois, Economics/330 Commerce W, 1206 South 6th St., Box 111, Champaign, IL 61820-6978

rkoenker@uiuc.edu

Key Words: *quantile, regression*

Two classes of quantile regression estimation methods for the recursive structural equation models of Chesher (2003) are investigated. A class of weighted-average derivative estimators based directly on the identification strategy of Chesher is contrasted with a control variate estimation method. The latter imposes stronger restrictions achieving an asymptotic efficiency bound with respect to the former class. An application of the methods to the study of the effect of class size on the performance of Dutch primary school students shows that (1) reductions in class size are beneficial for good students in language and for weaker students in mathematics, (2) larger classes appear beneficial for weaker language students, and (3) the impact of class size on both mean and median performance is negligible.

Quantile Regression and Identification of Structural Features

◆ Andrew Chesher, University College London

University College London, Dept. of Economics, Gower St., London, WC1E 6BT UK

andrew.chesher@ucl.ac.uk

Key Words: *identification, structural modeling, quantile regression*

This paper studies the identifying power of models with nonparametrically specified structural equations from which latent variables may not be additively separable. An example is a model for the joint determination of wages and schooling in which the focus is on the identification of the value of the returns to schooling at quantiles of the distribution of "ability." Adding local conditional quantile independence conditions, together with some monotonicity and local exclusion restrictions, produces models that locally identify partial differences of a structural function, and partial derivatives when there is sufficient smoothness and continuous variation. Application of the analog principle points to estimators of these structural features, which are functionals of estimated quantile regression functions. When arguments of structural functions have discrete variation there may be interval, but not point, identification and the analog principle points to quantile-regression-based interval estimators.

138 Improving the Measurement of Research and Development in the United States ▲ ⌘

Section on Government Statistics, Social Statistics Section, SSC, Section on Survey Research Methods

Monday, August 9, 2:00 pm–3:50 pm

Improving the Measurement of Research and Development in the United States

◆ Lawrence D. Brown, University of Pennsylvania

University of Pennsylvania, Statistics Dept., Philadelphia, PA 19104

lbrown@wharton.upenn.edu

The National Science Foundation's Statistical Resources Division conducts a portfolio of surveys and censuses that measure expenditures on Research and Development in the U.S. economy. A National Academy of Sciences panel has recently conducted a study of this data collection effort. The talk will summarize the panel's view of this portfolio and present highlights from the report of the panel. Issues to be discussed involve both topics of measurement related to the concepts of research and development, and the uses for R&D data, and issues of statistical methodology and practice in collecting such data. This talk will also introduce concerns that will be covered in more depth by the other speakers in this session.

International Comparability of Data on Research and Development and on Innovation

◆ Fred Gault, Innovation and Electronic Information Division

Innovation and Electronic Information Division, Statistics Canada R.H. Coats, Building 7A, Ottawa, K1A 0T6 Canada

Fred.Gault@statcan.ca

Key Words: *research, development, innovation, international standards, data comparability*

The United States pioneered in the collection and publication of data on R&D expenditures. Today, that collection takes place against the backdrop of OECD guidelines contained in the Frascatti and Oslo Manuals, and can be compared with programs conducted by statistical agencies in other countries. The National Academy of Science findings and recommendations were made with an appreciation of the need for supporting international comparability, as well as for informing international bodies as to the need for and possibilities of collecting R&D and innovation data from public and private agencies. This paper summarizes the international context of the collection of these statistics, and draws lessons for the international community.

Statistical Quality in the NSF Data Collections

◆ Richard Valliant, University of Maryland

University of Maryland, Joint Program for Survey Methodology, 1218 Lefrak Hall, College Park, MD 20742

rvalliant@survey.umd.edu

The rich variety of data collection modes involved with the various R&D statistics programs and surveys, censuses, and administrative data compilations affords a unique opportunity to employ standardized error profile techniques developed by Brooks and Bailar, and promulgated by the Federal Committee on Statistical Methodology across a range of applications. The National Academy of Science panel looked at the statistical quality of each of NSF's major data collections. The major findings of the panel review in terms of questionnaire design, electronic collection means, estimation with rare populations, imputation, and nonresponse adjustment are discussed.

139 Confidence Distributions, Bootstrap, and Bayes

Section on Bayesian Statistical Science, General Methodology

Monday, August 9, 2:00 pm–3:50 pm

Confidence Distributions, Confidence Nets, and Likelihoods

◆ Tore Schweder, University of Oslo

University of Oslo, Box 1095 Blindern, Oslo, 0317 Norway

tore.schweder@econ.uio.no

Key Words: *confidence distribution, confidence net, prediction, dimension reduction, bootstrap, bias*

A confidence distribution for a scalar parameter spans nets of confidence intervals, confidence nets, by its quantiles. A particular confidence net is obtained by indexing the likelihood ratio by way of the chi-square distribution, or a simulated null distribution. Confidence distributions for vector parameters are obtained from joint pivots when they exist. The contours of a likelihood ratio for a vector parameter, or of another objective function, make up a confidence net when indexed by a fixed null distribution. Likelihood functions represent cumulative confidence, not confidence density, and integrated and scaled likelihoods are generally not confidence densities. A confidence net for a derived parameter from a profiled or integrated likelihood typically requires simulation to obtain an approximate null distribution, and perhaps to remove bias. A confidence net for an unobserved variable might be found from a predictive likelihood. A simultaneous set of confidence nets is a box-shaped confidence net for the vector parameter or variable. Abundance estimation of bowhead whales off Alaska will illustrate the methodology.

Combining Information from Independent Sources through Confidence Distributions

◆ Kesar Singh, Rutgers University; Minge Xie, Rutgers University; William E. Strawderman, Rutgers University

Rutgers University, Dept. of Statistics, 501 Hill Center, Busch Campus, Piscataway, NJ 08854

kesar@stat.rutgers.edu

Key Words: *combining information, bootstrap, confidence distribution, frequentist inference, p value function, adaptive meta-analysis*

This paper develops a new methodology, together with related theories, to combine informations from independent studies through confidence distributions. It provides a formal version of definition of a confidence distribution and its asymptotic counterpart (i.e., asymptotic confidence distribution), which is convenient for combination purpose. A general combination method is developed along the lines of combining p values, with some notable differences in the optimality part in terms of Bahadur type efficiency. This paper also contains a development of practical interest on adaptive combining methods, with supporting theories provided under large sample settings. The key point of this development is that the adaptive methods combine only the right information, downweighting or excluding studies containing little or wrong information about the true parameter of interest. The combination methodologies are illustrated in three simulated and real data examples with a variety of applications.

Objective Validation of Priors through Probability Matching

◆ Jayanta K. Ghosh, Purdue University; Malay Ghosh, University of Florida

Purdue University, Dept. of Statistics, 150 N. University St., West Lafayette, IN 47907

ghosh@stat.purdue.edu

One way of validating a nonsubjective prior is to show that its posterior credibility intervals are also confidence intervals with the same (frequentist) probability of coverage. If one demands this to a certain degree of approximation, it leads to a probability matching differential equation.This does require regularity conditions which are violated in the discrete case. We would first provide a brief overview in the continuous case and show the old result that the Jeffreys prior is the unique probability matching prior in the one diomensional case. We will then focus on such results in the discrete case exploring both methods due to Judith Rousseau, as well as other approaches. We will look at several examples. This is joint work with Malay Ghosh.

140 Resampling Methods in Surveys

Section on Government Statistics, Social Statistics Section, Business and Economics Statistics Section, Section on Survey Research Methods

Monday, August 9, 2:00 pm–3:50 pm

Replication Variance Estimation for the Two-phase Regression Estimator

◆ Wayne A. Fuller, Iowa State University

Iowa State University, 202B Snedecor Hall, Ames, IA 50010

waf@iastate.edu

Key Words: *double sampling, reweighted expansion estimator, multiphase sampling*

The properties of a replication variance estimation method for the two-phase regression estimator is described. The procedure reproduces the estimated variance of the first-phase sample for those attributes of the first-phase sample used as controls for the second-phase sample. Only the second-phase sample is required by the analyst for variance estimation. The procedure is applied to the National Resources Inventory, a study of land use based on a large area sample of the United States.

Mean Square Error Approximation in Small-area Estimation by Use of Nonparametric Bootstrap

◆ Danny Pfeffermann, Hebrew University and University of Southampton; ◆ Hagit Glickman, Central Bureau of Statistics Israel

Central Bureau of Statistics Israel, 66 Kanfey Nesharim St., Jerusalem, 96190 Israel

hagit@dev.cbs.gov.il

Key Words: *small-area estimation, mean square error, nonparametric bootstrap*

The computation of reliable MSE estimators is a complicated process in small-area estimation problems. This is so because of the models in use and the small sample sizes within the areas that requires accounting for the contribution to the error resulting from hyper-parameter estimation. Pfeffermann and Tiller (2001) developed a general parametric bootstrap method for MSE approximation of correct order in the context of state-space models but basically the same procedure can be applied for small-area estimation problems. The use of this procedure requires, however, the specification of the parametric distributions of the model error terms. We study the use of nonparametric bootstrap methods for MSE approximations, which use the empirical residuals. The performance of the proposed method will be illustrated empirically and compared to other methods of MSE estimation in common use.

Inference Based on Nearest Neighbor Imputation and the Bootstrap

◆ Hansheng Wang, Peking University; Jun Shao, University of Wisconsin

Peking University, Guanghua School of Management, Beijing, 100871 P.R. China

hansheng@gsm.pku.edu.cn

Key Words: *confidence intervals, hot-deck, quantiles, re-imputation, smoothed bootstrap, variance estimation*

Nearest neighbor imputation (NNI) is a popular hot-deck imputation method used to compensate for item nonresponse in sample surveys. Although there exist results showing that NNI estimators are asymptotically unbiased and perform well in empirical studies, theoretical results for asymptotic inference (variance estimation and confidence intervals) are not available. We first establish the asymptotic normality of NNI estimators (such as the sample means and sample quantiles), which sets a foundation for large sample inference based on NNI. For variance estimation and confidence intervals, we consider a smoothed bootstrap method that addresses the effect of imputation by re-imputing nonrespondents in bootstrap datasets. The smoothed bootstrap method provides consistent variance estimators and asymptotically valid confidence intervals. We also carry out a simulation study to examine the finite sample performance of the smoothed bootstrap method.

141 Mining Massive Data and Risk Analysis ▲

General Methodology, Section on Risk Analysis, Section on Statistical Graphics, Section on Statistical Computing, Section on Physical and Engineering Sciences, Section on Quality and Productivity

Monday, August 9, 2:00 pm–3:50 pm

Stochastic Analysis of Single Molecule Experiments

◆ Samuel Kou, Harvard University; Sunney Xie, Harvard University; Jun S. Liu, Harvard University

Harvard University, Science Center, 1 Oxford St., Dept. of Statistics, Cambridge, MA 02138

kou@stat.harvard.edu

Key Words: *likelihood, Brownian diffusion, continuous-time Markov chain, Cox process, Ornstein-Uhlenbeck process*

Recent technological advances allow scientists for the first time to follow a biochemical process on a single molecule basis, which raises many challenging data-analysis problems. This paper provides the first likelihood-based analysis of the single-molecule fluorescence lifetime experiment, in which the conformational dynamics of a single DNA hairpin molecule is of interest. The conformational change is not directly observable and has to be inferred from changes in photon emissions from a dye attached to the DNA hairpin molecule. In addition to the hidden structure, the presence of molecular Brownian diffusion further complicates the matter. We show that the data augmentation technique can be utilized to handle both the Brownian diffusion and the issue of model discrimination. Our results increase the estimating resolution by several folds. The success of this analysis indicates there is an urgent need to bring modern statistical techniques to the analysis of data produced by modern technologies.

A Statistics of Extremes Approach to the Monitoring of Multiple Risk Indicators

◆ John H.J. Einmahl, Tilburg University

Tilburg University, Dept. of Econometrics and OR, PO Box 90153, 5045 XG Tilburg, The Netherlands

j.h.j.einmahl@uvt.nl

Key Words: *extremes, risk, quantile, aviation safety*

In risk assessment one often deals with extreme settings with very few or no occurrences in reality. Inferences about risk indicators for this type of occurrences face the problem of insufficient data. The methodology of statistics of extremes is particularly suitable for this problem. We will discuss the general theory of statistics of extremes and apply it to deriving thresholds for flagging extreme risks. Two aviation risk indicators for monitoring the performance of air carriers will be used to demonstrate our approach. From the performance measures of the carriers, we use statistics of extremes to devise a threshold system which divides the sample space into regions with increasing levels of risk. These regions, for example, are referred to as "expected," "warning," and "unacceptable" in aviation safety analysis. We discuss in detail the construction of thresholds for monitoring multiple risk indicators, the properties for the regions devised from the thresholds, and the application of such thresholds to a real dataset.

Mining Massive Report Data and Developing Tracking Statistics

Daniel R. Jeske, University of California, Riverside; ◆ Regina Y. Liu, Rutgers University

Rutgers University, Dept. of Statistics, Hill Center, Piscataway, NJ 08854-8019

rliu@stat.rutgers.edu

Key Words: *text-mining, report data, risk indicators, test classification, tracking statistics, error measurements*

We outline a systematic data-mining procedure for exploring large free-style text datasets to discover useful features and develop tracking statistics, generally referred to as performance measures or risk indicators. The procedure includes text mining, risk analysis, classification for error measurements and nonparametric multivariate analysis. Two aviation safety report repositories from the FAA and the NTSB will be used to illustrate applications of our research to aviation risk management and general decision-support systems. Some specific text analysis methodologies and tracking statistics will be discussed. Approaches to incorporating misclassified data or error measurements into tracking statistics will be discussed as well.

142 Flexible or Adaptive Designs ▲

Biopharmaceutical Section, Section on Physical and Engineering Sciences

Monday, August 9, 2:00 pm–3:50 pm

Adaptive Designs—Design Modifications Beyond Sample-size Reassessment

◆ Peter Bauer, University of Vienna

Peter.bauer@univie.ac.at

Key Words: *adaptive, flexible, interim, multistage, clinical*

Adaptive (flexible) multistage designs allow for midtrial design modifications based on all type of information from the interim data without compromising on the Type I error rate. The methodology can be derived either from the principle of combining stage-wise test statistics according to a prefixed rule or from the invariance principle for designs adhering to the same conditional error function. The situation is considered when design modifications may have the profound effect of changing the null hypothesis to be tested, e.g., when modifying the set of treatments or doses to be compared or adapting endpoints. The way how to perform multiple inference on the different null hypotheses tested throughout the trial is outlined for different scenarios. The price to be paid for a flexibility which allows to deal with the unexpected is discussed.

ASTIN—A Bayesian Dose-response Finding Study in Acute Stroke

◆ Michael Krams, Pfizer Inc.

Pfizer Inc., ClinSci CNS, Eastern Point Rd., Groton, CT 06340

michael_krams@groton.pfizer.com

Key Words: *Bayesian, dose-response, adaptive, clinical, sequential, interim*

ASTIN was an international phase II study to establish proof-of-concept for a novel neuroprotective therapy in acute stroke patients. We successfully deployed a Bayesian sequential design with real-time efficacy data capture, continuous reassessment of the dose-response and adaptive allocation of patients to one of 16 treatment arms. An optimal stopping rule was intended to facilitate termination of the trial for either futility or efficacy at the earliest possible timepoint. Over the period of one year, 966 stroke patients were entered. An independent data monitoring committee reviewed estimates of the dose-response and the probability of futility or efficacy on a weekly basis and stopped the trial prematurely for futility. This talk will highlight the doability of real-time learning in clinical drug development and discuss our experience in developing and implementing the design.

Bayesian Innovations in Clinical Trial Design and Analysis

◆ Donald Berry, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77030-4009

dberry@mdanderson.org

Key Words: *adaptive designs, sequential designs, clinical trials, Bayesian approach*

I will describe recent innovations in the design and analysis of clinical trials. The goals are creating more efficient clinical trials and clinical development programs and treating patients more effectively, both those in and those outside of trials. These designs are developed from the Bayesian perspective but are frequentist in the sense that they meet the usual standards for operating characteristics. These designs have been effected at my home institution, in national oncology studies. Or in pharmaceutical and medical device industry-sponsored trials. The revolutionary aspect is hardly earth-shattering in nonmedical science: we pay heed to the accumulating data and let it guide the course of the trial! I will give case studies showing actual designs and in analyses presented to the FDA. They include the possibility of early stopping, assigning patients to better performing therapies, and variations on themes such as seamless phase II/III trials with sequential sampling and using early endpoints. The savings of such an approach in terms of effectively using patient resources are substantial.

Hybrid Adaptive Designs for Clinical Trials

◆ Peter F. Thall, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, Biostatistics & Applied Mathematics Dept., Box 447, 1515 Holcombe Blvd., Houston, TX 77030

rex@mdanderson.org

Key Words: *clinical trials, Bayesian design, adaptive decision rules, dose-finding, adaptive randomization*

Clinical trials can be quite complex, often involving multiple stages of therapy, multivariate outcomes, and multiple goals. Consequently, the conventional “phase I, phase II, phase III” paradigm often is inadequate and a hybrid design, tailored to the particular trial at hand, must be constructed. I will describe two such hybrid designs, both Bayesian and both utilizing outcome-adaptive decision rules. The first is a method for dose-finding that combines efficacy and toxicity in terms of a function quantifying the trade-off between the probabilities of these outcomes. This will be illustrated by a trial of an agent for treatment of graft-versus-host disease in stem cell transplantation. The second is a design for a multicenter trial to compare gemcitabine +/- docetaxel for unresectable soft tissue sarcoma. The trial design uses Bayesian, covariate-adjusted adaptive randomization probabilities based on a trinary outcome observed in each of up to four courses of therapy. A computer simulation study of each design will be described.

143 Emerging Issues in Longitudinal Data Analysis ⌘

Section on Statistics in Epidemiology, International Chinese Statistical Association

Monday, August 9, 2:00 pm–3:50 pm

An Overview of Joint Modeling of Longitudinal and Time-to-event Data

◆ Anastasios A. Tsiatis, North Carolina State University; Marie Davidian, North Carolina State University

North Carolina State University, Dept. of Statistics, Box 8203, Raleigh, NC 276595-8203

tsiatis@stat.ncsu.edu

Key Words: *conditional score, likelihood, random effects*

A common objective in longitudinal studies is to characterize the relationship between a failure time process and time-independent and time-dependent covariates. Often, the proportional hazards model is used for this purpose, which necessitates knowledge of the time-dependent covariates for each individual at each failure time. In most studies, the longitudinally measured covariate is only collected at a finite set of time points and may be subject to measurement error and biological variation. “Naïve” methods, such as last value carried forward, lead to biased estimators for the underlying relationship of survival to the time-dependent covariates. Recently, joint modeling of both the survival distribution and the longitudinally measured time-dependent covariates as a function of a common set of random effects has been advocated for this problem. We consider such models, give the rationale for their use, and trace some of the analytic methods proposed for estimation.

Nonparametric Analysis of Multisample Longitudinal Data Based on Time-varying Generalized Odds Ratios

◆ Colin O. Wu, National Heart, Lung, and Blood Institute; Jianhua Z. Huang, University of Pennsylvania; Lan Zhou, University of Pennsylvania

National Heart, Lung, and Blood Institute, 2 Rockledge Center, Room 8218, 6701 Rockledge Dr., MSC 7938, Bethesda, MD 20892

wuc@nhlbi.nih.gov

Key Words: *basis approximations, B-splines, generalized odds ratios, longitudinal observations, nonparametric regression, time-varying models*

We propose a basis expansion approach for nonparametric estimation of the time-varying generalized odds ratios with multisample longitudinal observations. This approach provides a natural means for charactering the treatment effects and time trends of the conditional distributions of the longitudinal outcomes. When the outcomes have highly skewed distributions, ordinal scales or mixed distributions, our models enjoy the advantages in interpretability and robustness over the conditional mean-based basis approximation approaches. Our main results include a two-step B-spline smoothing estimator for the odds ratios, and its large sample and practical properties.

New Statistical Methods for the Analysis of High-dimensional Longitudinal Data

◆ James Robins, Harvard School of Public Health; Andrea Rotnitzky, Harvard School of Public Health; Aad van der Vaart, Vrije University

Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115

robins@hsph.harvard.edu

Key Words: *longitudinal data, causality, Bayes-frequentist, semi-parametric*

We discuss difficult or nonstandard problems that arise in estimation of optimal treatment strategies from high-dimensional longitudinal data, including interval estimation for nonregular parameters. We describe a new form of Bayes-frequentist compromise that allows us to incorporate doubly robust (DR) estimating functions (with their desirable frequentist performance) into a formal Bayesian decision analysis in which we explicitly allow for (1) confounding by unmeasured factors, (2) misspecification of models for the conditional expectation of the counterfactual response under the “optimal regime” and models for treatment assignment that are necessary for construction of DR estimators, and (3) misspecification of the optimal-regime structural nested model itself.

144 Frailty Models In Genetics Research

ENAR, Section on Statistics in Epidemiology, WNAR

Monday, August 9, 2:00 pm–3:50 pm

The Role of Frailty Models in Genetics

◆ Philip Hougaard, H. Lundbeck A/S

H. Lundbeck A/S, Dept. of Biostatistics, Ottiliavej 9, DK-2500 Copenhagen Valby, Denmark

phou@lundbeck.com

Key Words: *survival data, frailty, genetics, multivariate*

Usually genetic dependence is evaluated by normal distribution random effects models. Such models are, however, not sensible for survival data. First, censoring requires numerical integration; second, the normal distribution does not fit well to lifetimes; and third, it does not allow the flexibility as the standard semiparametric models. Finally, truncation is difficult. Frailty models are random-effects models designed for censored survival data, with dependence modeled by a factor on the hazard. I will discuss to which extent they can serve the same purpose as normal models. That is, basic questions like what is genotype and phenotype, and complicated questions like how to split variance into components. Both shared frailty and correlated frailty models are considered. Two applications are used. First, survival of like-sex Danish twins born 1870-1930, with both alive by age 6 and follow-up to 1995. Second, recurrent hypoglycemic episodes for diabetic patients according to the ACE gene. Frailty models are well suited to evaluate dependence, but evaluation of genetic variance components is not possible due to lack of a sensible scale, on which the terms act additively.

The Additive Genetic Gamma Frailty Models: Alternative Estimation and Inferences

◆ Hongzhe Li, University of California, Davis

University of California, Davis, Rowe Progam in Human Genetics, 1 Shields Ave., Davis, CA 95616

hli@ucdavis.edu

Key Words: *additive gamma frailty model, genetic linkage analysis, age of onset, pseudo-likelihood, ascertainment*

The additive genetic gamma frailty has been proposed for genetic linkage analysis to take into account variable age of onset and covariates information. The retrospective likelihood approach is often used for parameter estimation and for the test of genetic linkage. I will discuss some alternative methods for parameter estimation and hypothesis-testing, including a pseudo-likelihood approach and an ascertaint-adjusted likelihood approach for estimating the model parameters. Some thoretical and simulations results will be presented.

Unified Qualitative and Quantitative Linkage Analysis for Age at Onset Data in Affected Relative Pairs

◆ David V. Glidden, University of California, San Francisco

University of California, San Francisco, 500 Parnassus Ave., Division of Biostatistics, San Francisco, CA 94143-0560

dave@biostat.ucsf.edu

Key Words: *linkage analysis, multipoint mapping, statistical genetics*

We consider the problem of culling linkage information from a binary disease trait where there may be a tightly related quantitative biological marker or risk factor. Huang and Jiang (2003) used a likelihood-based approach and showed that evidence for linkage is increased by pooling binary and quantitative information. This talk considers a slightly different approach. The information is pooled through a linear combination of allele sharing and Haseman-Elston statistics. A linkage test and a multipoint approach to estimating the gene position are developed. The techniques are natural for age of onset data collected in affected relative pairs. Disease status is captured as binary and the onset time is used a quantitative trait. Simulations and a data example illustrate the method.

Bivariate Linkage Analysis of Disease and Age of Onset Using Copula Models

◆ Jian Huang, University of Iowa

University of Iowa, Dept.of Statistics and Act. Science, 241 SH, Iowa City, IA 52242

jian@stat.uiowa.edu

Key Words: *age of onset, gene-mapping, likelihood, modifier genes, pleiotropy, semiparametrics*

In the analysis of linkage between a disease and genetic markers, the primary outcome variable is usually the disease status (affected/unaffected). For many diseases with variable ages of onset, such as breast cancer, Huntington's disease, type 1 diabetes, among others, it is useful to incorporate age of onset information in the linkage analysis. Indeed, age of onset has been successfully used in the linkage analysis of breast cancer as a covariate to classify a heterogeneous sample into more homogeneous groups, so that the power is increased in detecting linkage of the dichotomous trait. We consider the case in which age of onset is considered as a quantitative trait and propose a semiparametric model based on copulas that incorporate age of onset variable into the linkage analysis of dichotomous traits. Two genetic models will used in the consideration of the statistical model: pleiotropy: the affection status and age of onset are affect by the same gene(s); and modifier genes—there are modifier genes different from the disease gene that influnce age of onset. Simulation studies and an example will be used to illustrate the proposed method.

145 The Role of the ASA in Statistics Development Worldwide

Cmte on International Relations in Statistics

Monday, August 9, 2:00 pm–3:50 pm

The Role of the ASA in Statistics Development Worldwide

◆ Stephen M. Stigler, University of Chicago; ◆ Daniel Pena, Universidad Carlos III de Madrid; ◆ Enrique de Alba, Instituto Tecnologico Autonomo de Mexico

Instituto Tecnologico Autonomo de Mexico, Division of Actuarial Science, Statistics and Mathematics, Río Hondo No. 1, Mexico, D.F. 01000

bayesiano@yahoo.com

The ASA is the largest and most influential statistical association in the world. Many members of the association are from foreign countries. The ASA Board has made "outreach to the international community" a major goal of the association. The purpose of this panel session is to bring together leaders from different continents to exchange ideas on what the ASA can and should do to further the development of statistics worldwide.

146 Forest Wildfires: A Variety of Modeling Applications ▲

Section on Statistics and the Environment

Monday, August 9, 2:00 pm–3:50 pm

Wildfire as Seen through a Statistical Eye

◆ Sim Larkin, U.S. Dept. of Agriculture Forest Service; Timothy Brown, Desert Research Institute

U.S. Dept. of Agriculture Forest Service, Pacific Wildland Fire Sciences Laboratory, 400 N. 34th St., Suite 201, Seattle, WA 98103

larkin@fs.fed.us

Key Words: *wildfire, statistical application*

In 1905, the Transfer Act delivered vast reserves of American forest to the U.S. Forest Service for administration, in large part because of a perceived wildfire problem. This action began the formalization of fire management from which modern fire protection has evolved. From the beginning, fire control was not taken strictly as administrative—scientific research was considered an integral part of solving the problem. In 1911, future chief forester William Greeley stated that "firefighting is a matter of scientific management, just as much as silviculture or range improvement." With strong European influence, fire research supported fire policy for many years. In the 1950s and '60s, fire research became a more physical-based scientific problem. In the 1970s and '80s, fire behavior and fire ecology dominated a profuse scientific effort. From the 1990s to present, fire science is often viewed as a global topic, intermingled with climate change, regional air quality and dynamical ecosystem feedbacks. Through all of the eras, statistics has been a vital part of both the scientific research and management decisions.

A State-space Model for Forecasting Wildland Fire Risk

◆ Haiganoush K. Preisler, PSW Forest Service

PSW Forest Service, 800 Buchanan St., West Annex, Albany, CA 94710

hpreisler@fs.fed.us

Key Words: *generalized additive models, fire danger indices, fire occurrence probabilities, fire weather, Oregon, sampled data*

Wildland fire managers have long desired to know the risks of severe fire events well in advance of its happening. A number of actions are available to address severe fire seasons. However, contingency actions have associated costs and timeliness issues. These issues require information about the likelihood of fire occurrence. We describe a probability model for forecasting wildland fire risks based on nonlinear state-space models. By the choice of variables included in the state space, we can handle many situations, for example, the history of the process (locations and times of fires) up to the present in addition to characteristics of the environment that might serve as explanatories. We found the model to be very useful for assessing the importance of commonly used fire danger indices and for forecasting expected numbers of fires in a region.

Exploring Lightning and Fire Ignition Data as Point Processes

◆ Willard J. Braun, University of Western Ontario

University of Western Ontario, Dept. of Statistical and Actuarial Sciences, Western Science Centre, Room 262, London, ON N6A 5B7 Canada

braun@stats.uwo.ca

Key Words: *lightning, forest fires, intensity, exploratory data analysis, data-sharpening*

Lightning accounts for a high proportion of forest fires that cause substantial damage. Efforts are under way in the forest fire research community to better understand the relationship between lightning and fire ignition. Statistical models, such as those based on regression or generalized linear models, are being applied to large amounts of data on lightning strikes and fire ignitions. This talk will describe attempts to understand these data as point processes. It will attempt to answer some very basic questions. For example, is a Poisson cluster process model useful for modeling lightning strikes in a district? If so, what are the clusters, and how can they be identified? Once identified, what can we say about the spatial pattern of the lightning strikes in such a cluster? Viewing the lightning cluster centers, rain occurrence, and fire ignitions in the same district as temporal processes, can we use intensity functions and the notion of coherence to study the relationships among these processes?

A Geo-spatial Approach to Allocating Wildfire Mitigation Funds

◆ Ronald E. McRoberts, U.S. Dept. of Agriculture Forest Service

U.S. Dept. of Agriculture Forest Service, 1992 Folwell Ave., St. Paul, MN 55038

rmcroberts@fs.fed.us

Key Words: *logistic model, satellite imagery*

In recent years, the forests of the western United States have seemingly been ablaze due to wildfires. These fires have consumed vast quantities of timber resources, devastated wildlife habitat, destroyed recreational opportunities, and endangered rural communities. Multiple federal programs, including Cooperative Forestry of the USDA Forest Service, make funds available to rural communities to simultaneously treat forested areas to mitigate wildfire risk and build infrastructure in economically depressed areas. Because sufficient funds are not available to treat all forest lands thought to be at risk of forest wildfire, a defensible approach to allocating available funds was required. A geographic information system (GIS) was used to identify forested lands in each region of the U.S. that simultaneously satisfied three criteria: (1) in close proximity to rural communities, (2) in need of economic assistance, and (3) at risk of wildfire. Digital layers corresponding to the first two criteria were constructed using census data. For the third criterion, a digital layer of logistic model predictions of the relative risk of forest wildfire was used.

What's the Probability that Wildfire Will Threaten My Home?

◆ David L. Martell, University of Toronto; Maria M. Sanchez-Guisandez, University of Toronto; Cui Wenbin, University of Toronto

University of Toronto, 33 Willcocks St., Faculty of Forestry, Toronto, ON M5S 3B3 Canada

martell@smokey.forestry.utoronto.ca

Key Words: *wildland fire occurrence, forest, fuel management*

Wildfire threatens homes in many Wildland Urban Interface (WUI) areas but such losses can often be mitigated by implementing fire prevention measures, initial attack strategies and fuel management measures near threatened structures. We describe a WUI Decision Support System (WUIDSS) that can be used to predict the probability that wildfire will burn any point on a landscape and thereby threaten specific homes and other structures. We simulate daily fire occurrence and spread processes which are influenced by fuel, weather, topography, and human behavior. Managers can use the WUIDSS to interactively delineate proposed spatially explicit fuel treatments and other measures on a digital map of a landscape and then evaluate those proposed strategies in terms of their potential impact on burn probabilities across the landscape.

147 Savage Award 2003 Winners

Section on Bayesian Statistical Science

Monday, August 9, 2:00 pm–3:50 pm

Noncentered Parameterizations for Hierarchical Models

◆ Omiros Papaspiliopoulos, Lancaster University

Lancaster University, Fylde College, Lancaster, LA1 4YF UK

o.papaspiliopoulos@lancaster.ac.uk

Key Words: *parametrization, hierarchical models, Levy processes, representations, convergence rates*

We give an overview of noncentered parametrizations for arbitrary three-stage hierarchical models. Such parametrizations are employed to improve the convergence of MCMC algorithms which are used for Bayesian inference. We argue, mathematically and intuitively, why such parametrizations are appealing in a wide range of applications, but particularly for models that involve high-dimensional latent stochastic processes. Such models are commonly used in econometrics, spatial statistics, and Bayesian nonparametrics.

Random Probability Measures Derived from Increasing Additive Processes and Their Application to Bayesian Statistics

◆ Igor Pruenster, University of Pavia

University of Pavia, Dip. to di Economia Politica e Metodi Quantitativi, Universitá degli Sudi di Pavia, via S. Felice 5, Pavia, 27100 Italy

igor.pruenster@unipv.it

Key Words: *Bayesian nonparametrics, Dirichlet process, increasing additive process, neutral to the right process, normalized random measure*

Increasing additive processes (IAP), i.e. processes with non-negative independent increments, represent a natural tool for defining random probability measures whose distributions act as nonparametric priors for Bayesian inference. It is well known that the Dirichlet process can be obtained either by normalizing a time-changed gamma process or, as a particular case of neutral to the right process, by the exponential transformation of a suitable IAP. The former construction is generalized to any IAP yielding the class of normalized random measures with independent increments (normalized RMI). Simple rules for the prior specification of normalized RMI and expressions for their posterior and predictive distributions are obtained. Moreover, results concerning the distribution of their means are provided. Due to their potential relevance in applications, normalized inverse Gaussian processes are considered in detail. Normalized RMI are then further generalized to normalized IAP driven random measures, which contain the mixture of Dirichlet process as a particular case. Finally, means of neutral to the right processes are studied.

Spatial Smoothing Using a New Class of Nonstationary Covariance Functions

◆ Christopher J. Paciorek, Harvard School of Public Health

Harvard School of Public Health, 655 Huntington Ave., Dept. of Biostatistics, Boston, MA 02115

paciorek@hsph.harvard.edu

Key Words: *nonstationary covariance, Gaussian process, spatial smoothing, matern covariance, nonstationary kriging*

We introduce a class of nonstationary covariance functions for spatial modeling. Nonstationary covariance functions allow Gaussian process (GP) models to adapt to spatial surfaces whose smoothness varies with location. The class includes a nonstationary version of the matern covariance, which parameterizes the differentiability of the spatial surface. We employ this new nonstationary covariance in a full Bayesian spatial model, parameterizing the nonstationarity in a computationally efficient way that produces nearly stationary local behavior. We also use the new covariance in an ad hoc “nonstationary kriging” method. We perform an extensive assessment of the full Bayesian model and compare it to a stationary GP model, as well as various spatial and general smoothing methods. In simulations, the nonstationary GP model adapts to variable smoothness while standard spatial methods do not. On real datasets, the nonstationary GP model outperforms other nonstationary smoothers, as well as the stationary GP model under certain conditions, but also shows evidence of overfitting when the spatial surface is complicated.

Perceptually Motivated Approaches to Audio Signal Enhancement: Broadband Noise Reduction via Bayesian Modeling of Time-frequency Coefficients

◆ Patrick J. Wolfe, University of Cambridge

University of Cambridge, Engineering Dept., Trumpington St., Cambridge, CB2 1PZ UK

p.wolfe@ieee.org

Key Words: *regression, model selection, Bayesian estimation, regularization, risk theory, Markov chain Monte Carlo*

This paper describes perceptually motivated statistical models for audio data, formulated to effect broadband noise reduction in natural sound signals such as speech and music. In contrast to most other Bayesian approaches to date, however, the methodology pursued here involves modeling of time-frequency coefficients—thus forming a natural yet novel extension of the techniques currently employed in many audio signal processing applications. Standard methods are interpreted from a Bayesian viewpoint and consequently extended to develop new, computationally efficient algorithms for online noise reduction. Working within this context, Bayesian risk theory is then employed in conjunction with perceptual optimality criteria to devise noise suppression rules motivated by psychoacoustic considerations. Masked thresholds in turn provide a basis for perceptual cost functions, under which minimum-risk spectral amplitude estimators are derived. Lastly, hierarchical models employing MCMC methods are developed to enhance speech signals degraded by noise, in which case meaningful prior information is shown to aid in the solution of the resultant ill-posed inverse problem via regularization.

Information Accumulation, Model Selection, and Rater Behavior in Constructed Response Student Assessments

◆ Louis T. Mariano, RAND Corporation

RAND Corporation, 1200 S. Hayes St., Arlington, VA 22202

loum@rand.org

Key Words: *Bayesian hierarchical models, item response theory, importance sampling, Bayes factors*

Open-ended items are now commonly used in educational testing. Responses to these items are usually evaluated by human raters, often with multiple raters judging each response. When modeling such an assessment, the additional uncertainty due to the raters must be accurately captured. The IRT Facets model is contrasted with the Bayesian Hierarchical Rater Model. A rigorous treatment of the approach to dependence and uncertainty in each model shows that the Facets model is overly optimistic in the accumulation of Fisher information about examinee ability, while information under the HRM is properly bounded. The HRM is expanded to include covariates of rater behavior as a diagnostic tool for both between-rater and within-rater effects. As an example, data from a state assessment rating study is used to demonstrate the effect of modality—the design for distributing items among raters—on rater performance. The selection among multiple covariates requires a choice among models where parameters number in the thousands. Methods based in importance sampling for computing marginal densities are extended to apply to this high-dimensional setting.

148 General Statistical Issues for Cancer Studies ▲ ⌘

Biometrics Section, Biopharmaceutical Section

Monday, August 9, 2:00 pm–3:50 pm

The Solution to the Two-sample Location Problem

◆ Michael J. Schell, University of North Carolina

University of North Carolina, Chapel Hill, NC 27599-7295

michael_schell@unc.edu

Key Words: *two-sample problem, t-test, Wilcoxon rank sum test, statistical practice*

The two-sample location problem is perhaps the most frequently encountered problem in statistical practice. Although many tests have been proposed, over 99% of applications use either the pooled t-test or Wilcoxon rank sum test. Many standard textbooks suggest using the Wilcoxon rank sum test for small samples, and the t-test for large samples. Greater statistical power is achieved by taking the opposite approach. For small samples, either a pooled or unequal variance t-test should be used, usually after a preliminary test for equality of variance has been conducted. For large samples, a Wilcoxon rank sum test with normal scores is best. A structured guideline is presented for selecting the appropriate test, while preserving definitive inference.

Statistical Analysis of Stratified Missing-at-random Designs for Large Clinical Trials

◆ Ilana Belitskaya-Levy, New York University School of Medicine; Judith D. Goldberg, New York University School of Medicine; Yongzhao Shao, New York University

New York University School of Medicine, 650 First Ave., Room 512, New York, NY 10016

beliti01@med.nyu.edu

Key Words: *missing data, maximum likelihood estimation, monotone missing pattern, clinical trial, EM algorithm, translational research*

In large clinical trials for complex diseases, it is often impossible for all investigators to include all patients in all of the ancillary translational research studies that are part of the trials. This arises due to time and budgetary constraints and other logistical considerations. We investigate a clinical trial designed with some covariates missing at random, and propose algorithms for data analysis that incorporate the features of the design.

Quality of Life Endpoints in Clinical Trials: From Start to Finish

◆ Amylou Dueck, Mayo Clinic Cancer Center; Sumithra J. Mandrekar, Mayo Clinic; Paul Novotny, Mayo Clinic Cancer Center; Pamela Atherton, Mayo Clinic Cancer Center; Jeff A. Sloan, Mayo Clinic Cancer Center

Mayo Clinic Cancer Center, 3725 N. Peseo Del Sol, Mesa, AZ 85207-1857

adueck@asu.edu

Quality of life (QOL) represents any outcome related to patient well-being, health, performance status, or symptom distress. Statisticians are often at the crossroads caught between investigators wanting to measure patient QOL and regulatory agencies looking for a sound scientific rationale for its inclusion, the proposed analysis plans and interpretation of results. The purpose of this paper is to provide a walking tour of the implementation of QOL endpoints in clinical trials from a statistician's perspective. The challenges faced by statisticians at the design (endpoints, choosing appropriate QOL tools, power calculations based on clinical significance, etc.) and analysis stage (choosing appropriate graphical and statistical procedures, methods to handle missing data, etc.) will be discussed. General guidelines to systematically address some of these issues will be proposed. With the help of examples, we show that the design and analysis of a QOL endpoint in a clinical trial is not necessarily different from other well-known and accepted endpoints like tumor response and survival. Clinical trial examples are drawn from oncology treatment and cancer control studies carried out by the North Central Cancer Treatment Group (NCCTG) and Mayo Clinic. We provide macro code to carry out the analyses undertaken.

Weighted Kaplan-Meier Statistics in the Case of Nonproportional Hazards: An Application in Cancer Prevention

◆ Paul H. Frankel, City of Hope National Medical Center; Mary E. Reid, Roswell Park Cancer Institute; James R. Marshall, Roswell Park Cancer Institute

City of Hope National Medical Center, 1500 East Duarte Rd., Duarte, CA 91010

pfrankel@coh.org

Key Words: *survival, proportional hazards, Kaplan-Meier, prevention, competing risks, Cox*

When comparing tumor progression or patient survival between two arms of a randomized study, proportional hazards always holds under the null hypothesis of no differences in treatment effect. However, if there is a difference between the two arms, the proportional hazard assumption is problematic, particularly in cancer therapeutic and prevention trials. As a result, the power of the log-rank test can be suboptimal and may actually decrease with increasing follow-up time. In therapeutic clinical trials for active disease, accelerated failure or mixed models with a cured fraction can often remedy this problem. In prevention trials, a different approach is required due to the increased magnitude of competing risks. A weighted Kaplan-Meier statistic will be presented as a viable alternative. Based on inverse variance-weighting and an empirical distribution, this method is useful in analyzing survival that does not follow proportional hazards and is not amenable to parametric methods. Data from the Nutritional Prevention of Cancer trial that tested a selenized yeast intervention on cancer prevention, will be used to demonstrate the utility of this test.

An Exact P Value from Multistage Phase II Clinical Trials

◆ Sin-Ho Jung, Duke University; Kouros Owzar, Duke University; Stephen George, Duke University

Duke University, DUMC Box 3627, Durham, NC 27710

jung0005@surgerytrials.duke.edu

Key Words: *stochastic ordering, uniformly minimum variance unbiased estimator, maximum likelihood estimator*

When we perform a statistical testing, we calculate a p value to see how significant the statistical evidence is against the null hypothesis from the data. Due to ethical and practical issues, clinical trials are conducted in multiple stages, but the reported p values often fail to reflect the design aspect of the trials. We propose an exact p value for analysis of multistage phase II clinical trials which have a binary variable, such as response, as the primary endpoint. The sample space consists of the paired observations of the stopping stage and the number of responses, which are jointly complete and sufficient statistic for the true binomial proportion. Our p value calculation is based on the ordering of sample space defined by uniformly minimum variance unbiased estimator. We will compare our p value with the one based on ordering defined by maximum likelihood estimator and the one ignoring multistage design aspect of phase II trials.

149 Statistical Approaches for Web Survey Data ▲ ⊞

Section on Survey Research Methods, Social Statistics Section

Monday, August 9, 2:00 pm–3:50 pm

Adjusting Responses in a Nonprobability Web Panel Survey by the Propensity Score-weighting

◆ Osamu Yoshimura, National Center for University Entrance Examination

National Center for University Entrance Examination, 2-19-23 Komaba, Meguro-Ku, Tokyo, 1538501 Japan

osamu@rd.dnc.ac.jp

Key Words: *web survey, nonprobability online panel, probability online panel, propensity score, weighting*

As is often the case with nonprobability web surveys, responses differ from other traditional surveys. Difference in response to the same question in two different surveys may caused by differnces in survey modes, survey administrators, survey periods, design and/or layout of questionnaires, etc., and difference in sample. If two surveys differ in only sample, a survey result could be adjusted to another survey by weighting. The propensity score weighting is a useful technique when the number of covariates to be considered is large. To examine the availability and effectiveness of the propensity score weighting in practice, an experimental web survey was conducted on two different panels, and then tried to adjust responses in nonprobability panel to that probability one. The response variables were owner rates of 14 kinds of information equipments, and the covariates adopted in estimating the propensity weights were 10 demographic variables. The possibility of a general weight that could adjust responses in a nonprobability web survey and the effective ge of the propensity score weighting in web surverys were discussed.

On the Variability of Estimates Based on Propensity-score-weighted Data from Web Panels

◆ Annica Isaksson, Linkoping University; Gosta Forsman, Linkoping University; Stig Danielsson, Linkoping University

Linkoping University, Dept. of Mathematics, Linkoping, SE-58183 Sweden

anisa@mai.liu.se

Key Words: *political opinion poll, variance estimation, propensity score adjustment*

This paper deals with variance estimation in web surveys of the general population. Such surveys often utilize a panel of web users, from which samples are selected for various surveys. Inference commonly suffers from considerable problems, including severe selection biases due to low internet penetration in the population and large nonresponse. Thus, good weighting procedures are badly needed. We restrict our attention here to an application of the "propensity-score-weighting" procedure, in which a parallel telephone survey is used to estimate the propensities of being in the web sample. The resulting weights may potentially reduce both selection bias and bias due to nonresponse. It is not obvious, however, how the estimator's variance should be estimated, and estimates are typically presented without adhesive uncertainty measures. This unsatisfying situation is the starting-point of our work. Since textbook variance formulae do not apply on the propensity score estimator, we try instead a model approach. From data collected for an opinion poll before the Swedish 2003 EURO referendum, model-based variance estimates are calculated, and compared with bootstrap estimates.

Propensity Score Matching as a Bias Correction Method for Internet-based Studies

◆ John Bremer, HI Europe/Harris Interactive; George Terhanian, HI Europe; Paul Strange, HI Europe

HI Europe/Harris Interactive, Waterman's Park, High St., Brentford, TW8 0BB UK

jbremer@hieurope.com

Key Words: *Propensity Score Adjustment, surveys, matching methods, Bayesian methods, election forecasting*

Most surveys conducted over the internet have up to three obvious forms of selection bias inherent in their sampling procedures: the bias associated with being online, the bias associated with panel based methodologies, and the bias associated with with the decision whether or not to participate in a given survey once invited. This paper investigates the use of a propensity score matching methodology using non-online studies (primarily telephone or face-to-face) as the control to greatly reduce these biases and suggests when this methodology would be applicable. The methodology will be thoroughly discussed and evidence presented as to the effectiveness of the methodology. Other methods of selection bias reduction will be discussed and compared to the propensity score methodology.

Attempting to Adjust for Selection Bias in Web Surveys with Propensity Scores: The Case of the Health and Retirement Survey

◆ Matthias Schonlau, RAND Corporation; Mick P. Couper, University of Michigan; Joachim Winter, University of Mannheim

RAND Corporation, 1700 Main St., PO Box 2138, Santa Monica, CA 90407

matt@rand.org

Key Words: *web survey, selection bias, propensity scoring, HRS*

Many web surveys allow respondents to self-select into the survey. Making inference about the population from a self-selected survey is very difficult. We analyzed HRS data as well as supplementary information about which HRS respondents volunteered to respond and responded to an additional Internet survey. The HRS is a longitudinal study of health retirement and aging. The target population of the HRS includes all adults in the contiguous United States, age 51 - 61, who reside in households. We investigate whether and to what extent it is possible to adjust for selection bias using propensity scores in the case of the HRS.

Propensity Score Adjustment for Web Surveys

◆ Sunghee Lee, University of Maryland

University of Maryland, Joint Program in Survey Methodology, 1218 LeFrak Hall, College Park, MD 20742

slee@survey.umd.edu

Key Words: *web survey, propensity score adjustment*

Most existing methodological web survey studies are focused on the questionnaire design. While findings from those studies are helpful in improving the web survey design, their implications for survey inferences are limited, because of their qualitative nature. Missing from the current literature are statistical examination of web survey data. Propensity score adjustment (PSA) has been suggested as an alternative for statistically surmounting inherent problems in web surveys. However, the link between web surveys and PSA is not well understood as studies that have examined the performance of PSA do not seem to document the statistical procedures or to clearly address the scope of adjustment. Moreover, there has been a minimal amount of evidence for its performance, and the implications are inconclusive. This study attempts to investigate PSA for web surveys. It will provide a clear description of applcation procedures of PSA for web survey data by presenting web survey protocols and stages of correction. It will investigate the applicability of PSA by focusing on volunteer panel web surveys by comparing the performace of different weighting schemes with respect to bias reduction.

150 Multiple Imputation ⌘

General Methodology, Section on Survey Research Methods

Monday, August 9, 2:00 pm–3:50 pm

Session on MI: Applications and Comparisons

◆ Wei Chen, University of Michigan; Trivellore E. Raghunathan, University of Michigan

University of Michigan, 611 Church St., Rm. 244, Ann Arbor, MI 48104

lisachen@umich.edu

Key Words: *multiple imputation, spline smoothing, generalized semiparametric regression, Gibbs sampler*

Sequential regression multiple imputation (SRMI) is a procedure to impute missing values for a relatively complex multivariate data structure when the data are missing at random. Fitting a sequence of regression models and drawing values from the corresponding posterior predictive distributions (ppd) will obtain the imputed values. This paper is to extend SRMI by relaxing the parametric regression assumption and using spline smoothing to establish the ppd. We developed an easily implemented algorithm using the Gibbs sampler to produce posterior inferences on the smoothing spline. The application of the approach is evaluated using the alcohol consumption data. A simulation study shows the proposed method is less sensitive to the sampling properties of the imputed data than the current method in SRMI.

Multiple Imputation of Missing Income Data in the National Health Interview Survey

Nathaniel Schenker, National Center for Health Statistics; Trivellore Ragunathan, University of Michigan; Pei-Lu Chiu, National Center for Health Statistics; ◆ Diane Makuc, National Center for Health Statistics; Guangyu Zhang, University of Michigan; Alan Cohen, National Center for Health Statistics

DMakuc@cdc.gov

Key Words: *multiple imputation, surveys, item nonresponse*

The National Health Interview Survey (NHIS), a household survey of the U.S. civilian noninstitutionalized population, is a major data source for studies of health status and health care access and utilization. Although item nonresponse in the NHIS is generally low, it is high for annual family income, a key variable in many analyses. In 2001, for example, 29% of families did not report family income. Starting with the 1997 NHIS, multiple imputation of unknown family income and unknown personal earnings is being carried out. Five sets of imputed values are being created for each year using sequential regression multivariate imputation. The application to the NHIS is discussed in documentation for public use files containing the 1997-2001 imputed income values. This presentation will describe the patterns of missing income data and the multiple imputation of income in the 1997-2001 NHIS. Analyses of health status and access to health care that are based on the multiply imputed data will be compared with analyses based on singly imputed data and analyses that delete respondents who did not report income.

Assessing Potential Precision Gains Using Imputation to Combine Administrative Data with Sampled Outcomes

◆ Thomas R. Belin, University of California, Los Angeles; Jun Xing, University of California, Los Angeles

University of California, Los Angeles, 51-267 UCLA Ctr. for Health Sciences, Dept. of Biostatistics, Los Angeles, CA 90095-1772

tbelin@ucla.edu

Key Words: *multiple imputation, quality of care, health services*

Research on quality of care often investigates whether certain procedures or medications are being used in treating patients. Some relevant information may be available from administrative data sources, which is apt to be relatively inexpensive, while other information may only be available from more expensive forms of inquiry. A motivating application for the present work involved a comparison of two public health systems, a county system and a Veterans Administration system located in a large urban area in the United States, regarding quality of care for schizophrenic patients, where data were available from administrative records on medication prescriptions, medical chart reviews, and patient interviews. Because the latter two methods of data collection are more expensive, it was of interest to investigate whether the more expensive data sources could be collected on a sample basis, with missing-data techniques used to support inferences comparing health systems. We investigate this question using both an available sample of 224 schizophrenic patients and a simulation evaluation that allows for alternative sample sizes and correlation structures.

A Comparison of Multiple Imputation and Other Unit-nonresponse Compensating Techniques in Fear of Crime Surveys

◆ Susanne Rässler, University of Erlangen-Nürnberg; Rainer Schnell, University of Konstanz

University of Erlangen-Nürnberg, Lange Gasse 20, Nürnberg, D-90403 Germany

susanne.raessler@wiso.uni-erlangen.de

Key Words: *regression-switching imputation, chained equations, monotone missing data pattern, weighting, iterative proportional fitting, nonresponse follow-up*

The results of nonresponse follow-up studies from national fear of crime surveys are compared with results following the use of different nonresponse correction procedures. We compare naive estimates, weighted estimates, estimates after a thorough nonresponse follow-up and estimates after multiple imputation (MI). For multiple imputation basically the flexible chained equations approach is used. A strong similarity between the MI-estimates and the follow-up-estimates is found. This suggests, that if the assumption of the data missing at random (MAR) holds, carefully selected and collected additional data used in MI procedures could yield similar estimates to a nonresponse follow-up at a much lower price and respondent burden.

Refining Multivariate Normal Imputations to Accommodate Non-normal Data

◆ Juwon Song, University of Texas M.D. Anderson Cancer Center; Thomas R. Belin, University of California, Los Angeles

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd. -447, Houston, TX 77030

jwsong@mdanderson.org

Key Words: *multiple imputation, multivariate normal distribution, importance sampling, SIR algorithm*

An MCMC algorithm based on a multivariate normal distributional assumption provides the basis for widely available statistical software (such as in SAS and S-PLUS) for conducting multiple imputation. However, when data do not fit well with the multivariate normal distribution, this technique may introduce biased estimates. We adapt the sampling-importance-resampling (SIR) algorithm to perform multiple imputation by first generating imputations based on a multivariate normal distribution and then refining the values drawn in the first stage using importance resampling, making use of a more realistic distributional assumption. We first show the feasibility of the method in a simple example where missing values are missing completely at random (MCAR). We discuss the complexity of adapting the method to the more plausible situation where missingness is not MCAR but may be missing at random (MAR). We then outline some potential extensions of the SIR idea that suggest useful Ave.s to explore.

151 Recent Developments on Skew-elliptical Distributions and Their Applications ⌘

Section on Bayesian Statistical Science

Monday, August 9, 2:00 pm–3:50 pm

A Multivariate Skew-in-mean GARCH Model

◆ Giovanni De Luca, University of Naples Parthenope; Nicola Loperfido, University of Urbino; Marc G. Genton, North Carolina State University

University of Naples Parthenope, Istituto di Statistica e Matematica, via Medina, 40, Naples, 80133 Italy

giovanni.deluca@uniparthenope.it

Key Words: *skew-normal distribution, stock returns, volatility*

The distribution of a small capitalized market index return is usually influenced by the performance of some dominant market. Some inspection revealed that the behavior of the returns at time t of a set of European stock index is affected by the performance of the U.S. stock exchange at time t-1. Taking into account the dynamics of the conditional correlations among stock exchange returns, a multivariate framework is introduced. Empirical analyses show that the multivariate distribution of financial returns is usually far from being multinormal. In particular, using Mardia's measure of multivariate skewness it is possible to stress a feature usually neglected in the statistical analysis of returns. Starting from standard assumptions, we show that the multivariate distribution of a set of European stock index returns, conditionally on the performance of the one-day lagged American stock exchange expressed through its sign, turns out to be a multivariate skew-normal distribution which is an extension of the multivariate normal distribution allowing for an additional parameter to regulate skewness. The model, denoted as Multivariate Skew-in-Mean GARCH model, is applied.

Estimation and Testing of Parameters in Multivariate Laplace Distributions

◆ Muni S. Srivastava, University of Toronto; Tonu Kollo, University of Tartu

University of Toronto, Dept. of Statistics, Toronto, ON M5S 3G3 Canada

srivasta@utstat.toronto.edu

Key Words: *kurtosis, multivariate cumulants, skew distributions, skewness*

Characteristics of Laplace distributions such as mean, covariance, third and fourth cumulants, and moments are given. Real valued skewness and kurtosis is defined in terms of multivariate cumulants and shown to be equal to Mardia's measures. An inequality between skewness and kurtosis is given which reduces to the univariate case when p=1. A test for skewness to be zero is given along with its asymptotic distributions.

Multivariate Skew-normal Distributions and Maxima of Normally Distributed Random Vectors

◆ Nicola M.R. Lopefido, Università di Urbino

Università di Urbino, Facolta di Economia, via Saffi 42, Urbino (PU), 61033 Italy

nicola@econ.uniurb.it

Key Words: *skewness, extremes, invariance*

Several authors showed that the minimum (maximum) of a bivariate normal vector with standardized marginals is univariate skew-normal. The present work shows that extensions of this result to normal vectors in higher dimensions lead to several generalizations of the univariate skew-normal: the multivariate skew-normal, the general skew-normal, and the fundamental skew-normal. From the theoretical viewpoint, the above results imply a new invariance property of skew-normal vectors. From the applied viewpoint, these results are useful in measuring visual acuity, hearing loss function and lungs' efficiency.

Nonparametric Bayesian Modeling Using Skewed Dirichlet Processes

◆ Fernando A. Quintana, Pontificia Universidad Católica de Chile; Pilar L. Iglesias, Pontificia Universidad Católica de Chile; Orellana Yasna, Pontifica Universidad Católica de Chile

Pontificia Universidad Católica de Chile, Av. V. Mackenna 4860, Macul, Santiago, 22 Chile

quintana@mat.puc.cl

Key Words: *random probability measures*

We introduce a new class of discrete random probability measures that extend the definition of Dirichlet process (DP) by explicitly incorporating skewness. The asymmetry is controlled by a single parameter in such a way that symmetric DPs are obtained as a special case of the general construction. We review the main properties of skewed DPs and develop appropriate Polya urn schemes that are useful for MCMC implementation. We illustrate the flexibility of skewed DP models with a few examples.

Linear Mixed Models With Semiparametric Generalized Skew-elliptical Random Effects

◆ Yanyuan Ma, North Carolina State University; Marc G. Genton, North Carolina State University; Marie Davidian, North Carolina State University

North Carolina State University, Raleigh, NC 27606

yma@unity.ncsu.edu

Key Words: *mixed effect, skew-elliptical*

We relax the standard normal assumption of the random effects in linear random effect models to semiparametric generalized skew-elliptical distributions (SGSE). The SGSE family is able to take into account heavy tail, skewness, and multimodality through the elliptical part, the coefficients in the skewing function part, and a tuning parameter respectively. When the tuning parameter is set to 0, the distribution reduces to an elliptical distribution. SGSE family facilitates simple sampling procedure, which reduces computation complexity for Monte Carlo type algorithms. It also avoids artifactual fluctuations which tend to occur for SNP distributions and mixture distributions.

152 Multivariate Analyses in Epidemiological Studies ⌘

Section on Government Statistics, Social Statistics Section, Section on Statistics in Epidemiology

Monday, August 9, 2:00 pm–3:50 pm

Simultaneous Modeling of Multiple Outcomes over Time: A Hierarchical Modeling Approach

◆ Abhik Das, RTI International; Charlotte Gard, RTI International; Henrietta Bada, University of Kentucky Chandler Medical Center; W. Kenneth Poole, RTI International

RTI International, 6110 Executive Blvd., Suite 420, Rockville, MD 20852

adas@rti.org

Key Words: *longitudinal data, hierarchical models, covariance structure, simultaneous inference, syndrome, multiple outcomes*

Multiple outcomes assessed longitudinally over time are common in public health research. This work, motivated by the Maternal Lifestyle Study, involves children prenatally exposed to cocaine prospectively followed up for the onset of abnormal behavior problems assessed by several standardized instruments measuring different domains of behavior. Thus, at each time point, each child contributed to multiple, possibly correlated outcomes that collectively constitute one syndrome, while specific outcomes affected by cocaine are also of scientific interest. Since it is neither substantively appropriate nor statistically efficient to fit separate longitudinal models for each outcome, we utilize the covariance structure in hierarchical models to simultaneously fit all the behavioral outcomes as a function of cocaine exposure and other covariates. The advantages of this approach are: (a) it recognizes that all outcomes together constitute one syndrome; (b) it boosts statistical efficiency by incorporating correlated outcomes and avoiding multiple comparisons; (c) it allows for outcome-specific exposure effects; so we can identify the specific behavioral domains affected by cocaine.

Subgroup Effects in Analysis of Pubertal Landmarks and Body Habitus

◆ Melinda L. Drum, University of Chicago; Robert L. Rosenfield, University of Chicago; Rebecca B. Lipton, University of Chicago

University of Chicago, 5841 S. Maryland Ave., MC 2007, Chicago, IL 60637-1470

mdrum@health.bsd.uchicago.edu

Key Words: *factorial models, interactions, NHANES-III*

Complex interactions in factorial regression may result from atypical subgroups that contribute little information to the question of interest. We encountered this situation when using

NHANES III data to model attainment of pubertal landmarks on age-sex specific BMI percentile and waist circumference in 3,197 (1,609 girls, 1,588 boys) nonhispanic black, nonhispanic white, and Mexican American youth ages 8-15, using logistic regression with NHANES survey sampling strata and probability weights. Multiple interactions involving age, race-ethnicity and body habitus were traced to two classes of atypical subgroups. Effectively all 14-15 year olds (95-100% over subgroups) had passed all the pubertal landmarks and other age groups had uniformly passed selected landmarks, providing little information. Among the leanest 2.5% (35 boys, 44 girls with BMI<5th percentile), low attainment of the pubertal landmarks was of interest, but the small numbers led to many zero cell counts, aliasing and inability to estimate contrasts of interest. When such subgroups were excluded, at most a single two-way interaction, clearly interpretable and clinically meaningful, remained for each pubertal landmark.

The Use of Classification Trees in Profiling Smokers and Smoking Intention among Adolescents

◆ Panagiota Kitsantas, East Carolina University; Trent Moore, Florida State University; David Sly, University of Miami

East Carolina University, Dept. of Mathematics, Austin 124, Greenville, NC 27858

kitsantap@mail.ecu.edu

Key Words: *classification trees, smokers, smoking intention, adolescents*

We employed a novel statistical method, classification trees, to explore the interactive nature of various predictor variables in profiling smokers (situational, current, and established), and nonsmokers who intend to smoke. The data (N=3,610) were generated from cross-sectional surveys of the Florida Anti-Tobacco Media Evaluation. Four classification trees were built based on the type of smoker and various groups of predictor variables such as sociodemographic, peer influence, social and health risks, role-modeling, and parental monitoring. The results support the important role of peer influence in smoking among adolescents. The classifier for the intention model suggests that social and health risks are important in the context of peer influence. A current smoker was profiled as one who frequently consumed alcohol and had friends that smoked. Knowing where to illegally purchase cigarettes was essential in the classification of both established and situational smokers. This study demonstrates the use of classification trees in profiling smokers and provides important information in identifying adolescents who intend to smoke.

Retention in Outpatient Methadone Treatment: Multivariate Analysis Using the Alcohol and Drug Services Study

◆ Lev S. Sverdlov, Synectics for Management Decisions, Inc.; Thoms M. Brady, Substance Abuse and Mental Health Services Administration; Sameena M. Salvucci, Synectics for Management Decisions, Inc.; Ping Yu, Synectics for Management Decisions, Inc.; Hannah Kyeyune, Synectics for Management Decision, Inc.; Sameer DeSale, Synectics for Management Decision, Inc.; Emmanuel Sikali, Synectics for Management Decision, Inc

Synectics for Management Decisions, Inc., 1901 North Moore St., Suite 900, Arlington, VA 22209

levs@smdi.com

Key Words: *substance abuse, regression analysis, methadone treatment, factor analysis, length of stay (LOS), survival analysis*

The years 1992 and 2000 saw a 40% increase in U.S. treatment episodes associated with heroin abuse. Effective methadone treatment for these clients is critical. This study explores factors affecting retention in treatment—an important measure of effectiveness—among clients discharged from outpatient methadone treatment, using the Substance Abuse and Mental Health Services Administration's nationwide Alcohol and Drug Services Study. The factors associated with length of stay (LOS) were studied using regression, survival, and factor (principal components) analyses. The average LOS was 555.0 days (n=428; N=109,973). In the three-dimensional space created by the methadone dose two weeks after treatment admission, duration of heroin addiction, and LOS, LOS had a complex configuration suggesting the presence of hidden factors associated with retention. Factor analysis of clinical, demographic, and social variables identified five factors (using the mineigen criterion) indicating differential associations of LOS with each of these. The paper will discuss the results of analyses with regard to the problem of individualization of methadone dose and optimization of the effectiveness.

The Continuation-ratio Model as a Tool for Exploratory Analysis of Clinical Research Data: An Illustrative Analysis of Delirium in a Hospitalized Elderly Cohort

◆ Peter H. Van Ness, Yale University School of Medicine

Yale University School of Medicine, 06510

peter.vanness@yale.edu

Key Words: *continuation-ratio, ordinal, risk-factor, exploratory, clinical, power*

The objective of much clinical epidemiologic research is the identification of disease risk factors. A second common goal of clinical research is to prevent morbidity that has occurred from becoming more severe, persistent, or disabling. To achieve this second goal it is important to determine if the same risk factors associated with disease occurrence are also associated with disease advancement and if the magnitude of their effect on the latter is the same as on the former. Fitting a continuation ratio regression model—a generalization of logistic regression for analyzing sequentially ordinal outcomes—can help achieve this second goal. Furthermore, it does so chiefly by calculating goodness-of-fit statistics and plotting related graphs, and thus can be implemented in an exploratory analysis. Interpretation of the results of exploratory data analyses involving the continuation-ratio model, especially when supplemented by relevant power calculations, can provide clinical researchers with valuable guidance in establishing protocols for confirmatory data analyses, even those using other models. This approach will be illustrated by an exploratory analysis of data.

153 Enhancing the Impact of Six Sigma

Section on Quality and Productivity, Section on Physical and Engineering Sciences, Business and Economics Statistics Section, Section on Statistical Education, Biopharmaceutical Section, Section on Statistical Consulting

Monday, August 9, 2:00 pm–3:50 pm

Enhancing the Impact of Six Sigma

◆ Doug Zahn, Florida State University; ◆ Heather S. Smith, California Polytechnic State University; ◆ Roger Hoerl, General Electric Company ; ◆ Dan Boroto, Florida State University; ◆ Ron Snee, Tunnell Consulting

Tunnell Consulting, King of Prussia, PA 19406

snee@tunnellconsulting.com

Key Words: *technical, interpersonal, relationship, leadership, management, attitude*

Six Sigma has produced dramatic results in many organizations. However, even its strongest proponents acknowledge that more results are possible and that we are not reaping all possible results for six sigma implementations. As several authors have noted, six sigma is not all technical tools; it is a blend of technical and interpersonal tools. We will look at ways of enhancing the impact of six sigma by looking at improving the technical and interpersonal aspects of six sigma practice.

154 Conveying the Core Concepts

Section on Statistical Education

Monday, August 9, 2:00 pm–3:50 pm

Conveying the Core Concepts

◆ Robert H. Carver, Stonehill College; ◆ Christopher J. Lacke, Rowan University; ◆ John D. McKenzie, Jr., Babson College; ◆ Jerry Moreno, John Carroll University; ◆ Ronald M. Weiers, Indiana University of Pennsylvania

Indiana University of Pennsylvania, Eberly College of Business and Information Technology, Indiana, PA 15705

rweiers@microserve.net

Key Words: *activities, applets, computational software, conceptual software*

What are the most important statistical concepts that students should learn in their first applied statistics course? How, in the context of a traditional course, especially one positioned as a service course, can we assure that students learn, understand, and internalize these concepts so that they can apply them effectively elsewhere? The panelists will propose a list of essential concepts and discuss incorporating such concepts into an introductory course. They will illustrate three basic approaches to presenting them: manipulative activities, computational software, and conceptual software. After discussion among the panelists, the audience will be invited to share comments and insights.

155 Differential Gene Expression

▲ ⌘

Biometrics Section, ENAR

Monday, August 9, 2:00 pm–3:50 pm

Degrees of Differential Gene Expression: Detecting Biologically Significant Expression Differences and Estimating Their Magnitudes

◆ David R. Bickel, Medical College of Georgia

Medical College of Georgia, Office of Biostatistics and Bioinformatics, 1120 15th St., AE-3037, Augusta, GA 30912-4900

bickel@prueba.info

Key Words: *biological significance, clinical significance, effect size, microarray gene expression data, medical significance*

Many methods of identifying differential expression in genes depend on testing the null hypotheses of exactly equal means or distributions of expression levels for each gene across groups, even though a statistically significant difference in the expression level does not imply the occurrence of any difference of biological or clinical significance. This is because a mathematical definition of “differential expression” as any nonzero difference does not correspond to the differential expression biologists seek. Furthermore, while some current methods account for multiple comparisons in hypothesis tests, they do not accordingly adjust estimates of the degrees to which genes are differentially expressed. Both problems lead to overstating the relevance of findings. Testing whether genes have relevant differential expression can be accomplished with customized null hypotheses, thereby redefining “differential expression” in a way that is more biologically meaningful. When such tests control the false discovery rate, they effectively discover genes based on a desired quantile of differential gene expression.

A Statistical Method for Identifying Differentially Expressed Genes

◆ Sunil Mathur, University of Mississippi

University of Mississippi, Hume 325, Dept. of Mathematics, University, MS 38677-1848

skmathur@olemiss.edu

Key Words: *statistical test, differential expresssion, gene, identification, robust, efficient*

A statistical test is proposed to detect differential gene expression for replicated microarray experiments conducted under two different conditions. This test is robust in nature and highly efficient when population is non-normal. The proposed test statistic is unaffected by outliers. Outliers have very serious effect on the test statistic, which is based on sample mean. The proposed test statistic is compared with Pan (2003) and Welch (1947) t-test under normal, gamma, and exponential populations set-up, using Monte Carlo technique.

A Weighted Permutation Method in Detecting Differential Gene Expression

◆ Xu Guo, Affymetrix, Inc.; Wei Pan, University of Minnesota

Affymetrix, Inc., 3380 Central Expressway, Santa Clara, CA 95051

xu_guo@affymetrix.com

Key Words: *EB, FDR, permuation, SAM, MMM*

A class of nonparametric statistical methods, including empirical Bayes (EB) method, the Significance Analysis of Microarray (SAM) and the mixture model method (MMM), have been proposed to detect differential gene expression for replicated microarray experiments. They all depend on constructing a test statistic, for example, a t-statistic, and then use permutation to draw inferences. However, due to special features of microarray data, the standard permutation may not estimate the null distribution of the test statistic well, leading to possibly too conservative inferences. We propose a weighted permutation method to overcome the problem: posterior probabilities of having no differential expression from the EB method are used as weights for genes to better estimate the null distribution of the test statistic. We also propose a weighted method to estimate the false discovery rate (FDR) using the posterior probabilities. Using simulated data and real data for two-condition and time-course microarray experiments, we show the improved performance of the proposed methods when applied in MMM, EB, and SAM.

Selection of Significant Genes in Microarray Experiments

◆ James J. Chen, National Center for Toxicological Research; Chen-An Tsai, National Center for Toxicological Research

National Center for Toxicological Research, 3900 NCTR Rd., HFT-20, Jefferson, AR 72079

Jchen@nctr.fda.gov

Key Words: *FDR, FWE, number of null hypotheses, ROC curve, classification*

An objective in microarray experiments is to select a subset of genes that are differentially expressed among different experimental samples or treatments. Selection of differentially expressed genes involves two steps: calculate a discriminatory index (score) for each gene for ranking of differential expressions and determine a cutoff from the ranked scores. This study considers using discriminative indices from the Receiver Operating Characteristic curve for the ranking of differential expressions and the significance testing approach to determine a cutoff from ranked p values. The unadjusted and adjusted p values are computed using the permutation resampling method. We discuss the use of the familywise error rate and false discovery error rate criterions for gene selection. Example datasets are used to illustrate gene selections in sample classification.

Semiparametric Logistic Regression for Classification of Gene Expression Microarrays: Support Vector Machines and Mixed Models

◆ Dawei Liu, University of Michigan; Xihong Lin, University of Michigan; Debashis Ghosh, University of Michigan

liudawei@umich.edu

Key Words: *semiparametric, logistic regression, support vector machine, mixed model*

In disease diagnosis, classification of patient samples plays a much important role. For some diseases both clinical covariates and gene expressions should be considered simultaneously in making decisions on disease status. We propose a semiparametric logistic regression model to relate a binary clinical outcome to clinical covariates and gene expressions, where the clinical covariate effects are modeled parametrically and gene expression effects are modeled nonparametrically using the support vector machine. The nonparametric function of gene expressions allows for the fact that the number of genes is likely to be large and the genes are likely to interact with each other. We show that the dual problem of the primal support vector machine problem can be formulated using a linear mixed effects model. Estimation hence can proceed within the linear mixed model framework using standard mixed model software. Both the regression coefficients of the clinical covariate effects and the support vector estimator of the nonparametric gene expression function can be obtained using the Best Linear Unbiased Predictor in linear mixed models.

Analysis of cDNA Microarray Gene Expression Data Using GenePlus

◆ Juan (Joanne) Zhang, U.S. Food and Drug Administration; Yi Tsong, U.S. Food and Drug Administration; Lue Ping Zhao, Fred Hutchinson Cancer Research Center

U.S. Food and Drug Administration, MD 20878

ZhangJua@cder.fda.gov

Key Words: *microarray, gene expression, GenePlus*

Microarray technology provides a global overview of thousands of different genes simultaneously. OPS/DAPR/FDA conducted a cDNA microarray experiment including six microarry chips. Each chip includes treated and controlled samples, which are from kidney tissues of rats. The primary objective is to identify genes that are differentially expressed between treated and controlled groups. We approach this goal from two different angles. One way is to identify those genes which have high mean differences based on six chips between two groups. Another way is to identify those genes that are consistently up regulated or down regulated based on each experiment among six chips. We apply six statistical analysis strategies to the dataset using GenePlus software.

P-values-only-based Stepwise Procedures for Multiple Testing and Their Optimality Properties

◆ Alexander Y. Gordon, University of Rochester

University of Rochester, Dept. of Biostatistics and Computational Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

Alexander_Gordon@urmc.rochester.edu

Key Words: *multiple testing procedures, p-values-only-based procedures, familywise error rate*

A multiple testing procedure is p-values-only-based, or uninformed, if its decision on which hypotheses to reject depends only on their observed p values, without relying on any information or assumptions about the joint distribution of p values. We study uninformed generalized step-down procedures and associated Type I error rates. We prove, under a natural condition of monotonicity, that the classic Holm procedure is the most rejective (most powerful) among all such procedures with strong control of familywise error rate (FWER) at a given level. On the other hand, the Bonferroni procedure cannot be improved on, while preserving FWER, in the class of all monotonic generalized step-up procedures.

156 Incomplete Data

Biopharmaceutical Section

Monday, August 9, 2:00 pm–3:50 pm

Application of Multiple Imputation in Calculating Propensity Scores in Observational Studies

◆ Ouhong Wang, Eli Lilly and Company; Brenda J. Crowe, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

wang_ouhong@lilly.com

Key Words: *multiple imputation, propensity scores, observational studies, simulation*

In the safety analysis of observational studies comparing two clinical treatment options, propensity scores are often used to control for baseline imbalances. At the mean time, missing baseline values result in exclusion of patient data due to missing propensity scores. It is hypothesized that both the power is reduced, and the point estimate of relative risk is biased depending on the missing data pattern. The use of multiple imputation in calculating propensity scores is investigated. The impact on power and estimating bias are studied through simulations and a comparison of results using data from an actual clinical study will be discussed as well.

Impute Missing Values below Precision Level

◆ Ying Wan, Merck & Co., Inc.; Cong Chen, Merck & Co., Inc.

Merck & Co., Inc., BLX 27, PO Box 4, West Point, PA 19486

ying_wan@merck.com

Key Words: *missing value, precision level, impute*

For biomarkers assessed in clinical trials, the precision level is restricted by technical limitations, leading to some missing measurements. In this situation, the missing data generated can not be treated as missing at random (MAR), but missing due to truncation. In spite of widely accepted methods, i.e., last observation carried forward (LOCF) and multiple imputation, we present a simple likelihood-based method. The idea is a version of regression calibration, using a conditional expectation to replace the unknown values. Three methods are developed for model estimation. Performances of the methods are compared when the proportion of missing values out of total observations vary. Finally, the methods are illustrated using examples of studies that included missing albuminuria measurements in a Cox regression analyses.

Estimating Relative Potency in Direct Assays Containing Nonresponders

◆ Daniel C. Bonzo, Serono, Inc.; Eugene Laska, Nathan S. Kline Institute for Psychiatric Research

Serono, Inc., One Technology Place, Rockland, MA 02370

Daniel.Bonzo@serono.com

Key Words: *relative potency, tolerance dose, nonresponders, cure models, delta method, Fieller's method*

Relative potency plays a key role in understanding the relationship between the doses of two treatments. Defined as the ratio of equally effective doses, it is central to communicating the relationship between a new drug entering the market and older medications. Because doses are not continuous, estimation in direct assays is hampered by the fact that tolerances are usually only available as grouped data which is frequently left- or right-censored information or both. This problem was considered in Bonzo, Friedman, and Laska (2003), which modeled premature discontinuation of treatment. However, further difficulties arise when there are nonresponders, regardless of dose, in the population. We propose an extension of our previous work for estimating relative potency in this setting that utilizes a cure model to characterize the tolerance dose distribution. The delta method and a large sample version of Fieller's theorem are used to construct confidence interval estimates. An example utilizing data that is typical of analgesic trials is given to illustrate the method's use.

Modeling LOCF Analysis of Continuous Endpoints in Clinical Trials

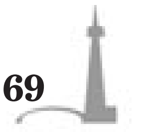
◆ Vijay Chauhan, NJ Biostatistics Inc.

NJ Biostatistics Inc., 21 Demarest Place, Piscataway, NJ 08854

vijay@njbiostatistics.com

Key Words: *LOCF, MMRM, clinical trials, study center*

Analysis of LOCF endpoint is a standard procedure for evaluating treatments in clinical trials. A model with treatment and center effects is commonly used in multicenter trials. Sample size is calculated to ensure a power 80-90% for treatment comparisons with one LOCF observation per subject. However, the precision of other fixed effects namely the study center is ignored. If such fixed effects are not estimated precisely, the treatment effects may be biased. Rather than using the LOCF data only, we propose to use all data, which will allow more precise estimation of both treatment and center effects. We suggest classifying the observations as LOCF or non-LOCF, and estimating the treatment effects within this class along with the center effect in the model. The treatment effect estimated from LOCF class only may be used for inferences about efficacy at the final time point. Analysis of the data of a clinical trial showed that the standard errors of center and treatment effects were about 58% and 18% smaller, respectively. Simulation was performed to confirm the accuracy of this procedure.



Comparison of Methods of Missing-data Imputation in Analysis of Two Phase III Trials Treating HIV1 Infection

◆ Les Huson, F. Hoffman La Roche; Jain Chung, F. Hoffman La Roche; Miklos Salgo, F. Hoffman La Roche

F. Hoffman La Roche, Broadwater Way, Welwyn, UK

les.huson@roche.com

Key Words: *missing data, imputation, HIV1*

In most longitudinal clinical trials, some patients drop out before the end of the planned follow-up, and, in order to allow an all-patient intent-to-treat analysis to be performed, it is common practice to use some method of imputation to estimate values for missing data. However, different imputation methods may give different results, and it is essential to investigate the sensitivity of the analysis using different imputation rules. Trials of treatment for HIV1 infection involve measurement of HIV1-RNA plasma values over time, and change from baseline in HIV1-RNA plasma concentration is a key measure of success. In our analysis of two trials of the new HIV1 fusion inhibitor enfuvirtide, we compared several methods of imputing and analyzing HIV1-RNA data, to check the robustness of the primary endpoint results. These were: (1) multiple imputation, (2) a nearest-neighbor hot-deck method, (3) censored regression analysis using last-observation-carried-forward, (4) last-observation-carried-forward using analysis of covariance, and (5) zero change from baseline (baseline-carried-forward).

Generally Applicable Modes of Analyses for Incomplete Binary Longitudinal Clinical Trial Data

◆ Ivy Jansen, Limburgs Universitair Centrum; Caroline Beunckens, Limburgs Universitair Centrum; Geert Molenberghs, Limburgs Universitair Centrum; Geert Verbeke, Katholieke Universiteit Leuven; Craig H. Mallinckrodt, Eli Lilly and Company

Limburgs Universitair Centrum, Universitaire Campus Bldg. D, Diepenbeek, 3590 Belgium

ivy.jansen@luc.ac.be

Key Words: *complete case analysis, generalized estimating equations, generalized linear mixed models, last observation carried forward, missing at random*

Many clinical trials result in incomplete longitudinal data. Common analysis methods are complete case (CC) and last observation carried forward (LOCF), resting on strong and unrealistic assumptions. Many full longitudinal methods, valid under MAR, have been developed. We focus on non-Gaussian outcomes, a setting more complicated than the Gaussian counterpart, due to the lack of an analogy for the linear mixed model. Model choices include the random-effects based generalized linear mixed models (GLMM) and the marginal generalized estimating equations (GEE). Since the latter is non-likelihood-based, it requires modification (weighted GEE) to be valid under MAR. Both methods provide similar results for hypothesis testing, but the estimated parameters have different interpretation. Current statistical computing brings GLMM and WGEE within reach and their implementation in depression trials is presented, showing they are viable alternatives for CC and LOCF, even when a single time point only (e.g., the last) is of interest. Even then, all information from all profiles, complete and incomplete, is used, showing this approach is fully compatible with the intention-to-treat principle.

Modified Intention-to-treat Analysis in Clinical Trials

◆ Chunqin Deng, PPD Development, Inc.

PPD Development, Inc., Research Triangle Park, 3900 Paramount Pkwy., Morrisville, NC 27560

chunqin.deng@rtp.ppdi.com

Key Words: *intention-to-treat, modified intention-to-treat, clinical trial, per-protocol*

Intention-to-treat (ITT) analysis has been widely accepted for the analysis of controlled clinical trials and is now part of several guidelines and recommendations. Strictly, the ITT analysis should include all randomized subjects, regardless of their adherence with the entry criteria, the treatment they actually received, and subsequent withdrawal from treatment or deviation from the protocol. While ITT preserves the initial randomization and minimizes the bias, there are situations where the ITT analysis is not practical or other alternatives may be more appropriate. Modified intention-to-treat (mITT) is a subset of the ITT population and allows the exclusion of some randomized subjects in a justified way. In the recently published results from several controlled clinical trials, mITT analysis has been used. In the pharmaceutical industry, it is not uncommon that under the cap of the ITT, the mITT is actually defined and used for the primary analysis. Using the examples from the real world in clinical trials, we discuss how mITT is defined, when mITT can be used, and how mITT is different from ITT, per-protocol, and other analysis populations.

157 Nonstandard Data and Applications in Longitudinal Data

ENAR, Biometrics Section, Section on Statistics in Epidemiology

Monday, August 9, 2:00 pm–3:50 pm

Longitudinal Modeling of Incomplete Data: Example of Plasma Viral Load in HIV Infection

◆ Rodolphe Thiebaut, Université Bordeaux 2; Hélène Jacqmin-Gadda, Université Bordeaux 2

Université Bordeaux 2, INSERM E0338 Bordeaux France, 146 rue Léo Saignat, ISPED, Bordeaux, 33076 France

rodolphe.thiebaut@isped.u-bordeaux2.fr

Key Words: *longitudinal data, joint modeling, informative dropout, left-censoring, HIV infection*

Longitudinal data may be incomplete because of left-censoring of the response variable (e.g., in case of a lack of sensibility of the assay used to measure the outcome) or may be missing due to informative dropout. An example of such data is the plasma viral load in HIV infection that is a good prognostic factor of disease progression but may be left-censored (undetectable viral load). Moreover, follow-up of patients may also be censored due to clinical progression (i.e., opportunistic infection occurrence) leading to informative dropout. To study longitudinal evolution of such marker, we proposed a joint model including a piecewise linear mixed model for the marker evolution and an accelerated failure time model for the occurrence of informative drop-out. Parameters were estimated by a direct maximization of the likelihood (Marquardt algorithm) handling for left-censored viral load measurements (based on Gaussian or mixture distributions). We presented some checks of model assumptions. The model was applied to a dataset of HIV-1 infected patients followed in an observational cohort (Aquitaine Cohort).

The Performance of Binary Mixed-effects Models in Settings with Correlation between Random Effects and Model Covariates

◆ John Neuhaus, University of California, San Francisco; Charles E. McCulloch, University of California, San Francisco

john@biostat.ucsf.edu

Key Words: *clustered data, conditional likelihood, misspecification*

Standard mixed-effects regression models typically assume that the random effects are independent of model covariates but this is frequently not the case in practice. Clustered datasets often exhibit variability in cluster means of the response that cannot be explained by covariate effects and independent random effects. We consider binary models with random intercepts and examine the effect of ignoring such correlations on standard analysis methods. These investigations suggest that while standard mixed-effects models fit by maximum likelihood yield biased estimates of covariate effects, conditional likelihood methods and mixed-effects models that partition covariate effects into between- and within-cluster components often provide nearly unbiased estimates of parameters of interest. We further illustrate the results with model fits to real data.

Shared Random Effects Analysis of Multistate Markov Models—Application to the Nun Study

◆ Juan C. Salazar, University of Kentucky; Suzanne L. Tyas, University of Kentucky; Mark F. Desrosiers, University of Kentucky; Kathryn P. Riley, University of Kentucky; Marta S. Mendiondo, University of Kentucky; David A. Snowdon, University of Kentucky; Richard J. Kryscio, University of Kentucky

University of Kentucky, Center on Aging, 800 South Limestone St., Lexington, KY 40536-0230

salazar@ms.uky.edu

Key Words: *multistate models, shared random effect, polytomous logistic regression, Markov chain, importance sampling, Alzheimer's disease*

Multistate models are appealing tools for analyzing data about the progression of a disease over time. We consider a multistate Markov chain with two competing absorbing states (Alzheimer's disease and death) and three transient nondemented states (Cognitively Intact, Mild Cognitive Impairments, and Global Impairments). The likelihood function for the data is derived and estimates for the effects of the covariates on transitions are determined when the process can be viewed as a polytomous logistic regression model with shared random effects. The presence of a shared random effect not only complicates the formulation of the likelihood but also its evaluation and maximization. Three approaches for maximizing the likelihood are compared using a simulation study; the first method is based on the Gauss-quadrature technique, the second is based on importance sampling ideas, and the third method is based on an expansion by Taylor series. These approaches are illustrated using a longitudinal study on aging and Alzheimer's disease conducted in a population of 678 catholic sisters (The Nun Study) aged 75 to 102 when the study began in 1991.

Comparing Treatments in a Longitudinal Study with Multiple Responses: The Effects of Bypass Surgery on Cognition

◆ Sarah J.E. Barry, Johns Hopkins University; Scott L. Zeger, Johns Hopkins University; Guy M. McKhann, Johns Hopkins University

Johns Hopkins University, 615 N. Wolfe St., Baltimore, MD 21205

sbarry@jhsph.edu

Key Words: *longitudinal, multiple responses*

In a study of the effects of coronary artery bypass graft (CABG) surgery on cognition, the response comprises 16 standardized results of neuropsychological tests. The study observes CABG patients and nonsurgical controls at baseline and 3, 12, and 36 months later. The scientific question is whether there is a decline in cognitive function in the CABG group relative to the controls. We compare and contrast three approaches. The first is the standard approach of summarizing the 16 NP tests into a single measure of cognition and then conducting standard longitudinal data analysis (LDA). The second is to fit an LDA model separately to each of the 16 NP tests and then to summarize the results. The final method is an hierarchical model that comprises random effects for both subjects and tests and directly estimates a mean (over tests) contrast of the time trends for the two groups. We contrast the methods in terms of statistical efficiency, ease of implementation and interpretation.

Longitudinal Analysis of Predictors of Depressive Symptoms in Old Age with Adjustment for Heritability

◆ Amy L. Byers, Yale University; Marilyn Stolar, Yale University; Heather G. Allore, Yale University

Yale University, Program on Aging, 1 Church St., 7th Floor, New Haven, CT 06510

amy.byers@yale.edu

Key Words: *heritability, genetic, longitudinal data analysis, psychosocial, family studies, depressive symptoms*

We present a case study using a multilevel model for a longitudinal continuous outcome measure obtained from the USC Longitudinal Study of Generations. We examine the association between non-genetic intergenerational factors and depressive symptoms in aging parents (age 50 - 72) with adjustment for heritability. The outcome, parent's depressive symptoms, was measured by the Center for Epidemiologic Studies Depression scale (CESD). Both the outcome variable and predictor variables were measured at baseline (in 1985) with two waves of follow-up at years 3 and 6. The heritability of the phenotypic trait of depressive symptoms was found to be 8.6% (95% CI: 0 - 57%; p=0.71) using a multilevel mixed effects model for 230 families. In summary, heritability was found not to be an important covariate in the model. This study raises the possibility that depressive symptoms measured by the CESD are not heritable but are influenced by nongenetic factors.

Generalized Estimating Equation for Longitudinal Cognitive Data

◆ Juan Li, Stony Brook University; Wei Zhu, SUNY, Stony Brook; Susan DeSanti, New York University; Mony J. DeLeon, New York University

juan.li@med.nyu.edu

Key Words: *generalized estimating equation, longitudinal, alternating logistic regression, marginal model*

In longitudinal studies of cognitive data, traditional likelihood-based methods are less tractable since these approaches either assume multivariate normal distributions or have difficulties with time-varying covariates and missing data. An alternative strategy is the generalized estimating equation, which is most often applicable for dichotomous outcomes. GEE is an extension to generalized linear model and it accounts for the structure of the covariance of the response variables. In a longitudinal cognitive study conducted at the NYU Center for Brain Health, 210 subjects including normal and patients who converted to mild cognitive impairment were chosen. We used the first-order GEE model to characterize the marginal expectation of the set of diagnosis outcomes as a function of longitudinal cognitive test scores and other prognostic factors. Second-order GEE (alternating logistic regression) with log-odds ratio structures was applied to the data also to investigate the longitudinal associations within the diagnostic outcomes themselves.

A Marginalized Pattern-mixture Model for Longitudinal Binary Data with Possibly Nonignorable Nonresponse

◆ Kenneth J. Wilkins, Harvard School of Public Health; Garrett M. Fitzmaurice, Harvard School of Public Health

Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115

kwilkins@hsph.harvard.edu

Key Words: *binary data, missing data, dropout, nonignorable nonresponse, longitudinal data*

This paper proposes a method for modeling longitudinal binary responses that are subject to nonignorable nonresponse. The approach presumes that the target of inference is the marginal distribution of the response at each occasion and its dependence on covariates, and can handle both monotone nd nonmonotone missingness. The approach involves a marginally specified pattern-mixture model that directly parameterizes both the marginal means at each occasion and the dependence of each response on indicators of nonresponse pattern. This formulation readily incorporates a variety of nonignorable nonresponse processes assumed within a sensitivity analysis. With identifying constraints in place, estimation of model parameters proceeds via solution to a set of modified generalized estimating equations. The proposed method provides an alternative to standard selection and pattern-mixture modeling frameworks, while featuring the advantages of each. The paper concludes with applications of the method to data from two longitudinal studies: a contraceptive clinical trial of dosage with substantial dropout, and a study of obesity in children for which there was intermittent nonresponse.

158 EM and Related Algorithms

Section on Statistical Computing

Monday, August 9, 2:00 pm–3:50 pm

A Numerical Method for the MLE of an NLMM

◆ Jing Wang, North Carolina State University

North Carolina State University, 2702 Vanderbilt Ave., Raleigh, NC 27607

jwang6@stat.ncsu.edu

Key Words: *NLMM, MLE, stochastic approximation, importance sampling*

Nonlinear mixed-effects models have received a great deal of attention in the statistical literature in recent years because of the flexibility they offer in handling the unbalanced repeated-measurements data that arise in different areas of investigation, such as pharmacokinetics. A numerical method involving a combination of stochastic approximation and importance sampling is proposed for the maximum likelihood estimation of the parameters in nonlinear mixed-effects models. Real data and simulation results have shown this method works well in most linear mixed-effects models and a few nonlinear mixed-effects models presented in the statistical literature.

Combining Data Augmentation and Over-relaxation to Speed Up the EM Algorithm

◆ Robert Creecy, U.S. Census Bureau

U.S. Census Bureau, SRD 3000 FB-4, Stop 9100, Washington, DC 20233

robert.h.creecy@census.gov

Key Words: *EM, algorithm, optimization*

The EM algorithm is popular primarily because it is easy to implement and has stable convergence properties. However, in some applications it can converge quite slowly. Various methods proposed to speed up the algorithm work by significantly complicating the writing of the computer code to achieve speed-up. Meng and van Dyke demonstrate that it is possible to improve the speed of the standard EM algorithm without sacrificing simplicity or stability by a class of data augmentation methods and introducing a working parameter to index the class. Another approach to speeding up EM while maintaining its simplicity has more recently been proposed by Salkhutidinov and Roweis that uses an adaptive over-relaxed bound optimization method. They derive a general adaptive over-relaxed EM algorithm that can be applied to any specific model for which a standard EM algorithm can be defined. This paper takes these two ideas and combines them in a single algorithm that speeds up EM while maintaining its simplicity. The method is illustrated on several examples.

Parallel Expectation-maximization Learning for Text Document Classification

◆ Shibin Qiu, University of New Mexico; Mei Qiu, University of New Mexico

sqiu@unm.edu

Key Words: *expectation-maximization, statistical learning, parallel data-mining, text classification, Bayes classifier*

Text classification has many important and practical applications given the massive amount of documents, especially online information that proliferates rapidly. In text classification, training a classifier with labeled documents is expensive. Expectation-maximization (EM) learning allows us to train classifier parameters using unlabeled data with only a small amount of labeled data while maintaining classification quality. However, EM training requires considerable amount of processing power. We analyze computational performance of EM learning and propose a parallel computing strategy for EM learning applied to text document classification. We apply a compact and fast data structure during parameter estimation. Grouped communication with optimal group size is used to reduce communication overhead and increase speedup. We also utilize the producer-consumer model for parallel classification. Consequently, computation for parameter estimation gets faster, storage space is saved and communication overhead is reduced. More than 100-fold speed-up has been achieved.

Empirical Likelihood with Arbitrary Censored/Truncated Data by Constrained EM algorithm

◆ Mai Zhou, University of Kentucky; Min Chen, University of Kentucky; Jingyu Luan, University of Kentucky

University of Kentucky, Statistics Dept., 849 Patterson Office Tower,
Lexington, KY 40506

mai@ms.uky.edu

Key Words: *maximization, weighted hazard, hazard constraints, proportional hazard regression*

Empirical likelihood ratio method is a general nonparametric inference procedure with many desirable statistical properties. To apply this method, a crucial computational step is to find the nonparametric likelihood estimator (NPMLE) and the maximum log likelihood under some constraints, which was achieved by using the Lagrange Multiplier in a lot of cases. But if the data are arbitrary censored or truncated with a constraint on the hazard, the Lagrange Multiplier no longer works. Instead, the self-consistent/EM algorithm introduced by Turnbull (1976) can be modified to handle the computation here. We present a modified EM algorithm with a constraint on the hazard of the unknown distribution such that the maximum empirical likelihood can be calculated easily. Moreover, we extend the EM algorithm to the Cox proportional hazards regression model. The empirical likelihood under hypothesis of regression coefficient and constraint of hazard can be found so using of empirical likelihood ratio method to test the hypothesis of regression coefficient for arbitrary censored or truncated data becomes computationally viable.

Missing Value Imputation with Generalized EM Algorithm and Information Complexity Measure

◆ Rui Zhang, University of Tennessee; Hamparsum Bozdogan, University of Tennessee

University of Tennessee, 331 Stokely Management Center, Dept. of
Statistics, Knoxville, TN 37996-0532

rzhang1@utk.edu

Key Words: *missing observations, imputation, generalized EM algorithm, information complexity criterion*

For a dataset with missing values, a direct estimation approach of the mean and covariance from the available data usually is not available. Using the estimation method for complete datasets will give negative eigenvalues, which might mislead many multivariate statistical techniques such as Principal Component Analysis (PCA), Factor Analysis, and Classification and Clustering techniques, and many more. Ignoring the incomplete observations takes the risk of losing valuable information, especially for the datasets with limited number of observations. The generalized expectation- maximization (GEM) algorithm is applied to perform the missing value imputation for Gaussian data. We introduce and propose a regularization parameter selection method using Bozdogan's information theoretic measure of complexity (ICOMP) criterion hybridized with the GEM. Numerical examples will be shown to illustrate the versatility of this new approach on benchmark datasets.

159 Statistical Applications in Government Agencies ▲ ⌘

Business and Economics Statistics Section

Monday, August 9, 2:00 pm–3:50 pm

Accounting for Imputation in Estimating the Variance of CPS Earnings

◆ Anne E. Polivka, Bureau of Labor Statistics; Martha Duff, Bureau of Labor Statistics

Bureau of Labor Statistics, Suite 4945, 2 Massachusetts Ave., NE,
Washington, DC 20212

polivka_a@bls.gov

Key Words: *missing data, CPS, imputation, wage and salary, variance estimation*

The Current Population Survey (CPS) provides timely data on the earnings of a large sample of wage and salary workers every month along with information about workers' demographic characteristics, their industry and occupation of employment, their hours of work, and their union status. Depending on the stringency of the criterion used, the nonresponse rate to the earnings questions in the CPS recently has been between 20-30%. Data for CPS responders who do not answer the earnings questions is generated using a hot-deck methodology. After documenting the hot-deck methodology used to impute earnings data, this paper examines the effect of imputation on earnings estimates derived from the CPS including means and coefficients from OLS regressions with a particular emphasis on the effect of imputation on the variance of these estimates. The paper concludes with exploration of ways to incorporate imputation into the calculation of the variance estimates.

An Evaluation of the Quarterly Financial Report's Computerized Self-administered Questionnaire

◆ Irene L. Evans, U.S. Census Bureau

U.S. Census Bureau, 2188 S. Glebe Rd., Arlington, VA 22204

irene.l.evans@census.gov

Key Words: *electronic reporting, computerized self-administered questionnaire, business statistics*

The U.S. Census Bureau conducts the Quarterly Financial Report to collect income statement and balance sheet data for U.S. manufacturing, mining, wholesale trade, and retail trade corporations. Beginning in the third quarter of 2000, select companies were given the opportunity to report electronically using a Computerized Self-Administered Questionnaire (CSAQ) instrument. Over the next several quarters, additional companies were given the opportunity to respond using the CSAQ. Respondents install and run the CSAQ on their personal computer. The automated questionnaire controls the flow of survey questions, provides on-screen instructions and help, includes edit checks, and reconciliation of data items performed as the respondent enters data. Results of an evaluation of the impact of the CSAQ on survey response time, response rate, data accuracy, and data quality are presented.

State Seasonal Adjustment under NAICS in the Current Employment Statistics Program

◆ John Stewart, Bureau of Labor Statistics

stewart_john@bls.gov

Key Words: *Current Employment Statistics, state and area seasonal adjustment*

In January 2003 the Current Employment Statistics (CES) Program, as a part of the U.S. Bureau of Labor Statistics' Payroll Survey, transitioned from the Standard Industrial Classification System to the North American Industrial Classification System. Data used to reconstruct the unadjusted series did not use sample-based data, and as such, limited the publication and creation of statewide seasonally adjusted data for such series. Historical studies of statewide seasonally adjusted data indicate the presence of distinct and differing seasonal patterns (being divided into historical, or benchmarked data, and current, or sample data) of the same time series, and as such, seasonal factors could not be forecasted without sample-based data. This challenge was met through the development of a methodology for reconstructing sample-based data to provide a forecast of seasonal factors for Supersectors and total nonfarm time series at the statewide level. This methodology, as well as a review of the historical methodology of adjusting both a historical and current time series will be discussed, with conclusions focused upon possible improvements for future years.

Accounting for Business Births and Deaths in the Bureau of Labor Statistics CES State and Area Program

◆ Brian Dahlin, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 4860, Washington, DC 20212

dahlin.brian@bls.gov

Key Words: *birth-death estimation, Current Employment Statistics, Bureau of Labor Statistics*

The Current Employment Statistics (CES) Program of the U.S. Bureau of Labor Statistics publishes national, state, and metropolitan area employment data by industry on a monthly basis. Selection of our survey sample occurs on an annual basis. An unavoidable lag exists from the date of the sample draw to the monthly estimation of industry employment. During the lag period, new businesses may form and existing ones may disappear—referred to as "births" and "deaths," respectively. Business births and deaths may contribute to nonsampling error in CES estimates. To reduce this error, an internally developed estimation and modeling technique is applied. The technique utilizes establishment micro-data from from the Bureau of Labor Statistics Longitudinal Database. This article summarizes improvements made to the net births model in recent years and discusses ongoing efforts to improve its accuracy.

New Metropolitan Area Data in the Current Employment Statistics Program

◆ Kenneth Shipp, Bureau of Labor Statistics

Bureau of Labor Statistics, Room 4860, 2 Massachusetts Ave. NE, Washington, DC 20212

shipp_k@bls.gov

Key Words: *Current Employment Statistics, metropolitan areas, small-domain estimation*

The Current Employment Statistics (CES) Program, commonly known as the U.S. Bureau of Labor Statistics' Monthly Payroll Survey, publishes national, state, and metropolitan area employment data by industrial sectors. Current publications of CES metropolitan area data include only 274 of the 337 officially recognized areas from the Office of Management and Budget's (OMB) designations. Using the recent introduction of revised metropolitan area definitions based on data from the 2000 Census, the CES program will expand its metropolitan area publication scope to nearly 400 separate areas. These will include all official Metropolitan Statistical Areas, Metropolitan Divisions within the largest Metropolitan Areas, and New York City. This paper will highlight the publication plans, estimation procedures, and the construction of historical estimates for these new data.

In the Right Place at the Wrong Time—The Role of Firms and Luck in Young Workers Careers

◆ Stefan Bender, Institute for Employment Research; Till Von Wachter, Columbia University

Institute for Employment Research, Regensburger Str. 104,
Nuremberg, D-90478 Germany

stefan.bender@iab.de

Key Words: *job displacement, instrumental variables, young workers, matched employer-employee data, past dependence, administrative data*

Do early job losses permanently reduce the earnings and career prospects of young workers? Simple estimates may overstate the true effects of early displacements, especially if less able workers sort into firms with high turnover rates. The bias from initial assignment of workers between firms is compounded by biases from selection within firms, which arise if employers selectively displace their least able workers, or if workers move voluntarily to take better jobs. This paper uses longitudinal social security data on German apprentices and their training firms to obtain estimates of the long-term effects of an early job loss that account for nonrandom assignment between firms and selection within firms. We use differences over time in the fraction of graduating apprentices that are retained by the training firm as an instrument for job displacement. These should reflect exogenous changes in firm-specific labor demand that are independent of individual ability or permanent firm characteristics. Using this strategy, we find that wage losses from leaving the training firm at graduation are initially strong but fade within the first five years in the labor market.

Combined PPS and Stratified Sampling in Modern Tax Audits

◆ Stephen J. Miller, Ernst & Young LLP; Wendy Rotz, Ernst & Young LLP; Eric Falk, Ernst & Young LLP; Ryan Petska, Ernst & Young LLP

Ernst & Young LLP, 1225 Connecticut Ave. NW, Washington, DC
20036

steve.miller@ey.com

Key Words: *pps, stratified sampling, rare events*

Theory tells us that in audit situations, probability proportionate to size (pps) sampling can be advantageous over stratified random sampling when sampling in settings with rarely occurring errors. However, federal and state regulators, as well as the big four accounting firms, often use stratified approaches in many tax audit settings, even in the face of low error rates. The simulation study of Rotz et al. concluded that there was no overwhelming evidence to convert from a purely stratified to a purely pps approach. This paper extends the work of Rotz et al. (2002) by exploring the advantages of a combined stratified and pps approach over standard stratified and pps designs.

160 Design Issues for Health Surveys ▲

Section on Survey Research Methods, SSC, Section on Government Statistics, Section on Health Policy Statistics, Social Statistics Section

Monday, August 9, 2:00 pm–3:50 pm

Statistics Canada, the Croatian Ministry of Health, and the Central Bureau of Statistics: A Joint Effort in Implementing the 2003 Croatian Adult Health Survey

◆ Yves Béland, Statistics Canada

Statistics Canada, R.H. Coats Bldg., 16th Floor, Tunney's Pasture,
Ottawa, ON K1A 0T6 Canada

yves.beland@statcan.ca

Key Words: *health survey, cross-sectional*

The Croatian Ministry of Health as part of an overall strategy for health system reform to support Croatia's capacity to achieve a more efficient health system, launched the Health Systems Project. Sponsored by the World Bank, the contract requires that the work be done in accordance with internationally accepted quality standards. Statistics Canada (STC) has been asked to join the Canadian Society of International Health project team to offer technical assistance, and statistical expertise on best practices for ensuring quality. In partnership with the Andrija Stampar School for Public Health and the Central Bureau of Statistics of the Croatia Republic, STC has coordinated the development and the implementation of the 2003 Croatian Adult Health Survey (CAHS). The 2003 CAHS collected information to ensure a comprehensive assessment of the health of Croatians, including health services utilization, health status, and determinants of health. The 2001 Croatian Census of Households was used to select a representative sample of households under a multistage stratified cluster design in which one adult per household was selected to participate in an interview.

Adapting to Emerging Health Information Needs with the Canadian Community Health Survey

◆ Marc Hamel, Statistics Canada; Yves BÈland, Statistics Canada; Johane Dufour, Statistics Canada

marc.hamel@statcan.ca

Key Words: *health surveys, sampling approaches, collection methodology, cross-sectional*

The need for health information is more critical than ever in Canada, with the ageing of the population and the anticipated increased demand for health care services. Statistics Canada launched the Canadian Community Health Survey (CCHS) in 2000 to fill many of the data gaps in health information. The CCHS survey program, consisting of two cross-sectional surveys conducted over a two-year repeating cycle, has been designed to address priority health information needs at regional, provincial, and national levels. The survey also allows for the study of special populations such as immigrants, seniors, children, and others. An important distinctive feature of the CCHS is its flexibility to respond and adapt to emerging health information needs. During the course of the first two cycles of the CCHS, the survey program has had to adapt to requests for additional health-related information for smaller levels of geography than originally designed for. This paper will describe how the CCHS survey program has adapted its collection methodology, sampling approaches, questionnaire content, and data outputs to respond to those numerous requests while maintaining high quality standards.

NYC Hanes: Design of a Community Health and Nutrition Examination Survey

◆ Jill M. Montaquila, Westat; Leyla Mohadjer, Westat; Lester R. Curtin, Centers for Disease Control and Prevention; R. Charon Gwynn, NYC Dept. of Health and Mental Hygiene; Lorna Thorpe, NYC Dept. of Health and Mental Hygiene

Westat, 1650 Research Blvd., Rockville, MD 20850

jillmontaquila@westat.com

Key Words: *local area study, sample design, area sample*

The National Health and Nutrition Examination Surveys (NHANES) are one of several health related programs sponsored by the National Center for Health Statistics (NCHS). A unique feature of these surveys is the collection of health data by means of medical examinations on a nationally representative sample of the U.S. population. Because health and nutrition characteristics vary considerably by demographic and geographic characteristics, intensive study of particular subgroups through the “community” HANES model has been a longstanding interest. The first survey that followed this model was the Hispanic HANES (HHANES, 1982-1984), which produced estimates for three major Hispanic subgroups. In 2003, NCHS and Westat agreed to provide technical expertise on a project spearheaded by the New York City Department of Health and Mental Hygiene to conduct a community HANES in New York City. Interviews and examinations for NYC HANES will be conducted during the summer of 2004 on a representative sample of the NYC adult population. This paper presents an overview of the community HANES model, and gives a summary of the objectives, sample design, and analytic capabilities of NYC HANES.

Why Large Design Effects Can Occur in Complex Sample Designs: Examples from the NHANES 1999-2000 Survey

◆ David A. Lacher, National Center for Health Statistics; Lester R. Curtin, Centers for Disease Control and Prevention; Jeffery P. Hughes, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 4215, Hyattsville, MD 20782

dol2@cdc.gov

Key Words: *design effect, complex sample, variance estimation, NHANES, laboratory test*

The design effect is defined as the ratio of the “true” sampling variance to the hypothetical simple random sample variance assuming the same point estimate. Design effects are used to compare alternative designs to determine the effective sample size for analysis and to adjust confidence intervals for estimates based on complex designs. Design effects have been examined for the National Health and Nutrition Examination Survey (NHANES) for 1999-2000. Although design effects for most estimates based on NHANES were less than three, design effects for means of laboratory tests were much higher, ranging up to 40. Design effects of means were larger for laboratory tests with small between-person coefficients of variation. It was determined that small systematic shifts in laboratory test values apparently led to larger between-sampling unit component of sampling error and increased the design effect. Simulations were done to compare the effect of analytical bias and the effect of precision of laboratory methods on the design effect of the mean. For the NHANES data, the bias of the laboratory method influenced the design effects of means more than the precision of the laboratory test.

Analysis of NHANES 1999-2002: One Four-year or Two Two-year Surveys?

◆ Lester R. Curtin, Centers for Disease Control and Prevention; Margaret D. Carroll, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 3311 Toledo Rd., Room 4206, Hyattsville, MD 20782

lcurtin@cdc.gov

Key Words: *complex survey, variance estimation, sample size, percentiles, design-based inference*

The current National Health and Nutrition Examination survey is designed to be a nationally representative sample for each year and for combinations of years. Public use data files for the combined years 2001-2002 were released in late spring 2004 while data files for 1999-2000 have been previously released. Estimation and analysis of current NHANES data is complicated by several factors including the change in survey design for 2002 due to the 2000 Census of population becoming available, confidentiality restrictions on two-year data releases, and the need to examine relatively rare events. The CDC/NCHS recommends that the two-year datasets be combined and primarily analyzed as a single (four-year) survey. This paper provides support for that recommendation and compares the two-year versus four-year survey in terms of the demographic detail for reliable estimation for a range of commonly used statistics such as means, proportions and percentile distributions. The paper also examines the impact of small sample size and design constraints on annual variation to establish conditions under which the data may be appropriately analyzed as two two-year surveys.

Dissemination Strategy for Statistics Canada’s Canadian Community Health Survey

◆ Mario BÈdard, Statistics Canada; Yves Béland, Statistics Canada; Larry Macnabb, Statistics Canada

Statistics Canada, Main Bldg., Room 2600, Tunney’s Pasture, Ottawa, ON K1A 0T6 Canada

mario.bedard@statcan.ca

Key Words: *Canadian community health surveys, dissemination, health*

The CCHS consists of two distinct cross-sectional surveys conducted over a recurring two-year cycle. The cycle’s first survey, known as the regional or general component, is designed to collect data from a sample large enough to provide data specific to each of Canada's 100+ health regions. This survey provides a wide range of information, arrayed in sets of core, optional, and sub-sample modules. In the second year, the focus survey, the CCHS collects in-depth provincial-level information on a specific topic. The CCHS is designed to reach a wider, more diversified user base than any earlier population health surveys. The comprehensive dissemination strategy implemented to this effect includes the provision of microdata to various share partners, as well as public use microdata, filtered to avoid disclosure issues yet detailed enough to allow sophisticated analysis. The internet is also put to contribution. This paper looks at the content and dissemination of the first two complete cycles of the CCHS: the 2000-01 general, the 2002 on mental health and well-being, the in-processing 2003 general, and the in-collection 2004 focus on nutrition.

161 Time Series Models and Analysis ⌘

General Methodology

Monday, August 9, 2:00 pm–3:50 pm

Geometric Process Modeling on Binary Data

◆ Pak Kei Ho, University of Hong Kong

University of Hong Kong, Hong Kong, NA

h9914468@graduate.hku.hk

Key Words: *geometric process, maximum likelihood, Bayesian*

Much research has been done in the analysis of observations with trend. One of the widely used techniques is the nonhomogeneous Poisson (NHPP) process. If the successive interarrival times are monotone, the Cox-Lewis model and Weibull Process model are commonly used. We focus on the geometric process modeling, a more direct approach to model the data by a monotone process. A stochastic process {Xi, i=1,2,.} is a geometric process if there exists a>0 such that $\{Yi=a^{(i-1)Xi}\}$ generates a renewal process. If the observations {Wi} are binary, we define $Wi = I(Xi >1) = I(Yi > a^{(i-1)})$ with the indicator function (the probit link function) I. Under the assumption that Yi follows Weibull distribution, we study the statistical inference for the geometric process on a binary dataset. Two approaches will be investigated: maximum likelihood and the second is Bayesian. Then, some suggestions and discussions will be made based on the simulation and real data analysis.

Nonparametric Empirical Bayes Method in Some Panel Time Series Models

◆ Ka Yee Liu, University of Hong Kong; Wai Keung Li, University of Hong Kong

University of Hong Kong, Room 515, Meng Wah Complex, Hong Kong

karenliu@graduate.hku.hk

Key Words: *empirical Bayes, panel time series, threshold autoregressive model, non-Gaussian time series*

Empirical Bayes method has been proposed to improve the least squares estimates in the analysis of panel autoregressive time series. We extend the idea to analyze the so-called panel Threshold Autoregressive (TAR) time series and panel non-Gaussian time series. Simulation studies have shown that the Empirical Bayes method can improve the least squares estimates sufficiently well to give better estimates of the true parameter in the analysis of panel TAR time series and panel non-Gaussian time series.

Temporal Aggregation of Stationary and Nonstationary Discrete-time Processes

◆ Henghsiu Tsai, Institute of Statistical Science, Academia Sinica; Kung-sik Chan, University of Iowa

Institute of Statistical Science, Academia Sinica, 128, Academia Rd. Sec. 2, Taipei, 115 Taiwan

htsai@stat.sinica.edu.tw

Key Words: *asymptotic efficiency of prediction, autocorrelation, ARFIMA models, long memory, spectral density*

We study the autocorrelation structure and the spectral density function of aggregates from a discrete-time process. The underlying discrete-time process is assumed to be a stationary AutoRegressive Fractionally Integrated Moving-Average (ARFIMA) process, after suitable number of differencing if necessary. We derive closed-form expressions for the limiting autocorrelation function and the normalized spectral density of the aggregates, as the extent of aggregation increases to infinity. These results are then used to assess the loss of forecasting efficiency due to aggregation.

G(λ)-stationary Processes and Their Application to Wave Match Filtering

◆ Huiping Jiang, New York State Psychiatric Institute; Henry Gray, Southern Methodist University; Wayne A. Woodward, Southern Methodist University

New York State Psychiatric Institute, 1051 Riverside Rd. #48, Division of Biostatistics, New York, NY 10032

Jianghu@pi.cpmc.columbia.edu

Key Words: *G(λ)-stationary processes, filtering, time-varying frequencies, instantaneous spectrum*

Methods such as wavelets short-term Fourier transforms and M-stationary processes have been developed to analyze the time-frequency properties of a process where the frequency changes with time. We review a new approach to this problem based on time deformation. In particular, we review G(λ)-stationary processes and demonstrate their applicability to a wide variety of problems with time varying frequencies. These problems include bat echolocation, dolphin and whale chirp signals, as well as a variety of seismic signals. Some simulated examples are also included. Several examples are also given that show the application of these new methods to filtering signals with time varying frequencies. This latter application is referred to as Wave Match Filtering. Several new ideas are introduced, including instantaneous spectrum of a G(λ)-stationary process.

Bispectral-Based Goodness-of-fit Test for Time Series Linearity

◆ Nusrat Jahan, Mississippi State University; Jane L. Harvill, Mississippi State University

Mississippi State University, Dept. of Mathematics and Statistics, PO Box MA, Mississippi State, MS 39762

nj1@msstate.edu

Key Words: *nonlinear time series, bispectral density function, noncentral chi-squared distribution, goodness-of-fit test*

We discuss a bispectral-based goodness-of-fit test for time series linearity. The two-stage procedure first tests Gaussianity. If this null is rejected, the linearity of the time series is tested. The tests are constructed from properties of the normalized bispectrum under the corresponding null hypothesis. Under Gaussianity, the test statistic has a (central) chi-square distribution with two degrees of freedom; under linearity a noncentral chi-square 2. Size and power of the test are examined. We conclude by applying the test to a dataset from signal processing literature.

Time Series Models with Infinitely Divisible Margins

◆ Xuefeng Li, University of Pennsylvania

University of Pennsylvania, Suite 370, 3600 Market St., Philadelphia, PA 19104

xuefeng@grasp.cis.upenn.edu

Key Words: *decomposability, Markov, time series models, infinitely divisible, AR, MA*

We study time series models with infinitely divisible (ID) marginal distributions. New constructions of two classes of autoregressive models and a class of moving average models with ID margins are proposed. Their property and relationship with existing models are explored. All of them yield positively correlated ID processes. Related to these constructions we describe a notation of decomposability for infinitely divisible processes.

Computationally Intensive Spectrum Estimation Methods and Nonstationarity

◆ Juana Sanchez, University of California, Los Angeles

University of California, Los Angeles, Dept. of Statistics, 8130 Math Sciences Bldg, Box 951554, Los Angeles, CA 90095-1554

jsanchez@stat.ucla.edu

Key Words: *bootstrap, splines, spectrum, ensemble, nonstationarity, seasonality*

Considerable progress has been made in estimating the spectral distribution of a stationary time series. New computationally intensive methods such as log-spline, ensemble and bootstrap are now easy to implement. When a time series is nonstationary, however, the user is at the mercy of the controversy on the appropriate way to make the data stationary so that the spectrum can be estimated accurately. If, in addition to nonstationarity, the data present seasonal frequencies that are so dominant that it is very hard to discern the presence of other frequencies, one has to worry about what to do with those. The nonexistence of a unified way to approach the problem of making a nonstationary dataset stationary and the problem of dominant frequencies without losing information about the data make it hard to avoid that controversy. I build on the latest computationally intensive spectrum estimation methodology to explore ways to integrate the spectrum estimation and handling of nonstationarity in a unified framework that retains all the information about the data.

162 Model-based Estimation ▲

Section on Government Statistics, Section on Survey Research Methods

Monday, August 9, 2:00 pm–3:50 pm

Model-based Estimates of the Finite Population Mean for Two-stage Cluster Samples with Unit Nonresponse

◆ Ying Yuan, University of Michigan; Roderick J. Little, University of Michigan

University of Michigan, 1420 Washington Heights, M4048D, Dept. of Biostatistics, Ann Arbor, MI 48109

yuany@umich.edu

We propose new model-based methods for unit nonresponse in two-stage survey samples. A standard design-based weighting adjustment (WT) is potentially inefficient when the estimated response rates within clusters are very variable. In addition, we show that the usual random-effects model based estimator of the population mean (RE) is biased in the setting of unit nonresponse unless nonresponse is missing completely at random, which makes the often unrealistic assumption that the response rates are unrelated to cluster characteristics. This fact motivates modifications of RE that allow the cluster means to depend on the response rates in the clusters. Two approaches are considered to correct the bias of RE, one that includes the observed response rate as a cluster-level covariate (RERR), and one based on a probit model for response (NI1). We also consider another nonginorable model estimate of the mean (NI2) that removes the bias of WT, NI1, and RERR when there is association between response and the survey outcome within the clusters.

Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Microlevel Latent Class Analysis

Clyde Tucker, Bureau of Labor Statistics; Paul P. Biemer, RTI International and UNC-CH; Brian J. Meekins, Bureau of Labor Statistics; ◆ Jennifer Shields, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Suite 1950, Washington, DC 20212

shields_j@bls.gov

Key Words: *latent class analysis, Measurement error, bias, longitudinal surveys, underreporting, Consumer Expenditure Surveys*

This paper uses latent class analysis (LCA) to estimate the amount of underreporting on the BLS Consumer Expenditure Quarterly Survey (CEQ). Specifically, it models underreporting in a given commodity category by those reporting a purchase of any item within that category. This work builds on previous research by these authors using LCA to model the extent of erroneous

nonreports and estimate the amount of underreporting from nonreporters. It also builds on research by Tucker (1992) that models patterns of consumer expenditure reporting based on microlevel, procedural indicators. Data from the CEQ for 1996 to 2003 are used in the analysis. A series of LCA models are used to evaluate observed expenditure reporting patterns. Model covariates include characteristics of the interview, the respondent, and the household. Best-fitting models are determined from statistical and subjective diagnostics developed by the authors. Diagnostic tools developed in previous research are expanded, and issues related to correlated classification errors are discussed. Finally, weighted estimates of the amount of underreported expenditures are produced.

The Posterior Likelihood

◆ Hee-Choon Shin, NORC, University of Chicago

NORC, University of Chicago, 55 East Monroe St., Suite 4800, Chicago, IL 60603

shin-hee-choon@norcmail.uchicago.edu

Key Words: *likelihood, Bayes theorem*

The Bayes theorem can be stated that the posterior probability is proportional to the product of prior probability and the likelihood. The principle of maximum likelihood states that, when confronted with a choice of hypotheses, we choose that one which maximizes the likelihood. Now consider a likelihood function with known parameter(s). We call this prior likelihood. The posterior likelihood is the product of the prior likelihood and the likelihood. The desired parameter(s) will be obtained by maximizing the posterior likelihood.

Measurement of Mail Volumes—An Application of Model Assisted Estimation

◆ Stephen Woodruff, United States Postal Service; Feng Lan, United States Postal Service

United States Postal Service, 475 L'Enfant Plaza SW, Room 1830, Washington, DC 20260-5328

swoodruf@email.usps.gov

Key Words: *robust estimation, superpopulation model, regression*

When foreign countries dispatch mail to the United States, they pay the U.S. Postal Service a portion of the costs to sort, transport, and deliver that mail to its final U.S. address. The System for International Revenue and Volume Inbound (SIRVI) captures pieces and weight data from foreign origin mail entering the United States. Unfortunately, mail processing generally has a higher priority than measuring the quantity of mail delivered. This means that sample control is marginal and a probability sample based estimator could not be properly implemented. This paper describes a more robust procedure based on a regression superpopulation model, an application of work done by Royall, et al. The constraints that mail processing places on survey design and inference are also briefly described. Several estimators are evaluated by simulation and theoretically in the product space of the sampling distribution and the superpopulation model.

Comparative Effectiveness of Administrative Data and County-level Aggregates in Nonresponse Adjustments for Surveys of Low-income Populations

◆ Frank Potter, Mathematica Policy Research, Inc.; Mourad Touzani, Mathematica Policy Research, Inc.; Ronghua Lu, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., PO Box 2393, Princeton, NJ 08543-2393

FPotter@Mathematica-MPR.com

Key Words: *nonresponse adjustments, response propensity, low-income populations*

In the evaluation of social programs, low-income populations are surveyed to study the program effectiveness. In these surveys, limited data are available from administrative data to use in the nonresponse adjustments (generally, age, gender, and sometimes race of the program participant). The data can be supplemented by using aggregated data at the county level or at the zip code level (e.g., Area Resource File or Census Bureau data). These data provide contextual information on where the program participant resides, but are not directly related to the individual's propensity to respond. In addition, using these aggregated data can add to the costs of the nonresponse adjustment activities. Our paper will investigate the benefits of using the additional aggregated data in terms of response variation explained by the response propensity models when the aggregated data are used relative to models using only the available administrative data. We will also evaluate measures of model fit for applicability to similar nonresponse logistic modeling efforts. Our study is based on more than 50 nonresponse models developed for a 10-state evaluation of a social program targeting children.

On Fitting the Proportional Hazards Model to Data from Complex Surveys

◆ Susana Rubin Bleuer, Statistics Canada

Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A-0T6 Canada

rubisus@statcan.ca

Key Words: *survey data, counting processes, joint design-model, proportional hazards*

Binder (1992) proposed a method of fitting the proportional hazards regression model to survey data with complex sampling designs via a sample regression estimator defined as the solution of the weighted sample partial likelihood score (SPLS) function. Lin (2000) gave a formal justification of Binder's method under the super-population approach. However, he did not provide conditions for the asymptotic theory to hold. I use Lin's (2000) set up of the super-population approach and develop counting process methodology for a joint design-model space, to obtain, under stated sufficient model and design conditions, a rigorous proof of the weak convergence of the SPLS process and the asymptotic normality of the sample weighted regression estimator in the joint design-model space envisaged by Lin (2000).

Predicted and Conditional Marginals for Cox's Proportional Hazards Model Using SUDAAN

◆ Elvessa Aragon-Logan, RTI International; G. Gordon Brown, RTI International; Beth G. Barnwell, RTI International; Babubhai V. Shah, SAFAL Institute Inc.

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

earagon@rti.org

Key Words: *survival, SUDAAN, adjusted treatment means, predictedmMargins, direct standardization*

Obtaining adjusted treatment means for regression models is often of interest to the researcher. The primary goal of our work is to produce functional forms for two types of adjusted treatment means—predicted marginals and conditional marginals—for Cox's proportional hazards model for cluster-correlated data with tied event times. We consider two methods, Breslow's and Efron's, that attempt to adjust the estimation from Cox's model in situations where ties occur. The predicted and conditional marginals are presented for both of these methods. The point estimates for the two types of marginals are derived using pseudo-likelihood equations, and variance estimates are determined using Taylorized deviations. The variance estimates derived from the Taylorized deviations tend to be robust for cluster-correlated data. We will present analytic results comparing the variance estimates using the Taylorized deviations and the model-based (naïve) variance estimates. All calculations for the analytic examples are derived using new features in SUDAAN's SURVIVAL procedure.

163 Repeated Measures and Correlated Data

Biopharmaceutical Section

Monday, August 9, 2:00 pm–3:50 pm

Optimal s^n Factorial Design When Observations within Blocks Are Correlated

◆ Venkat Sethuraman, GlaxoSmithKline; Damaraju Raghavarao, Temple University; Bikas K. Sinha, Indian Statistical Institute

GlaxoSmithKline, One Franklin Plaza, FP 1375, Philadelphia, PA 19101

venkat.s.sethuraman@gsk.com

Key Words: *optimal designs, correlated observations, factorial designs*

We characterize D-optimal designs for s^n symmetric factorial experiments when observations within blocks are correlated. Motivation for this problem lies in a pharmaceutical experiment for estimating the bioavailability, measured by plasma AUC, in healthy human subjects when administered several formulations. Each subject is considered as a block receiving several different tablet formulations (combination of n-factors). An s^n experiment is designed in $s^{\{n-1\}}$ blocks such that observations are correlated within blocks, and uncorrelated between blocks. We provide construction of D-optimal designs for estimating slopes when the block size is equal to s. Further when s=2, we give details of the construction of D-optimal designs for 2^n, even if the block sizes are unequal and odd.

D-optimum Designs for Dynamic Systems in the Presence of Auto-correlated Errors

◆ Barbara Bogacka, Queen Mary, University of London; Bartosz Kuczewski, University of Zielona Gora; Maciej Patan, University of Zielona Gora

Queen Mary, University of London, Mile End Rd., London, E1 4NS UK

B.Bogacka@qmul.ac.uk

Key Words: *nonlinear regression, chemical kinetic models*

Optimum experimental designs for parameter estimation are widely considered in the statistical literature, particularly for linear models of observations. However, there is not much work done for models with complicated error structures, though in many practical situations in which data are collected sequentially over time, substantial serial correlation in the errors appears. This phenomenon presents a serious difficulty in building constructive algorithms for determining optimal experimental designs. Some efficient algorithms will be proposed to determine D-optimum design in the presence of the auto-correlated errors and are validated via computer simulations for a class of nonlinear dynamic models coming from chemical kinetics. The paper closes with discussion of related work and some suggestions for further developments.

Generalized Estimating Equations Analysis of Correlated Multinomial Data Using a New Class of Working Correlation Structures

◆ Fangyi Luo, Procter & Gamble Company; Jorge G. Morel, Procter & Gamble Company

Procter & Gamble Company, 11450 Grooms Rd., Cincinnati, OH 45242

luo.f@pg.com

Key Words: *GEE, correlated data, categorical analysis, working correlation structure*

Generalized Estimating equations (GEE) methodology has been used in the analysis of longitudinal or clustered categorical response data in clinical trials. GEE requires the specification of a working correlation matrix for the repeated measurements within a subject. A new class of working correlation structures for correlated multinomial data is proposed in this paper and incorporated into the GEE analysis using a One Step Gauss-Newton (OSGN) procedure. Under mild regularity conditions and assuming the link function of the response probabilities is correctly specified, the OSGN procedure provides consistent and asymptotically normal estimates of the regression parameters, and consistent estimates of the intracluster correlation parameters. The proposed method is illustrated with a clinical trial example and compared to GEE analysis using independent and exchangeable working correlation structures.

Bootstrap Approach for Computing Standard Error of Estimated Coefficients in Proportional Odds Model Applied to Repeated Assessments in Psychiatric Clinical Trials

◆ Qiuhu Shi, New York Medical College; Young Zhu, Janssen Pharmaceutical Products, L.P.; Jianhong Lu, Data Solutions, LLC

New York Medical College, School of Public Health, Valhalla, NY 10595

qiuhu_shi@nymc.edu

Key Words: *bootstrap, proportional odds model, clinical global impression, positive and negative syndrome scale*

In this paper, bootstrap method is described for computing standard error of estimated coefficients in building a proportional odds model to explore clinical relevance of Positive and Negative Syndrome Scale (PANSS) in schizophrenia patients. In our model, patient's Clinical Global Impression (CGI) was categorically rated as an outcome variable. Patient's PANSS total score was treated as a continuous independent variable. During a clinical trial, CGI rating and PANSS scale were typically measured more than one time. It was certain that using all measurements would a be benefit in building a prediction model. There were two potential problems. One was that measurements within patient were usually not independent each other, another was that there might be large proportion of missing data. GEE with repeated measure method might not work properly for large proportion of missing data. However, the estimated coefficients in proportional odds model with this type of data were still unbiased, but not for standard error of estimated coefficients. Bootstrap approach on patient level would provide a method to compute standard error of estimated coefficients.

Influence of Withdrawal Criteria and Treatment Time-response Profiles on Dropout Process in Clinical Trials: A Simulation Approach

◆ Naum M. Khutoryansky, Novo Nordisk Pharmaceuticals, Inc.; Andrew Liu, Novo Nordisk Pharmaceuticals, Inc.

Novo Nordisk Pharmaceuticals, Inc., 100 College Rd. West, Princeton, NJ 08540

nakh@novonordisk.com

Key Words: *longitudinal clinical trials, incomplete data, dropouts, dropout rate, simulation*

We propose a simulation approach to evaluate the effect of different withdrawal criteria and treatment time-response profiles on the dropout process and dropout rate in longitudinal clinical trials, particularly in diabetes clinical trials of phase III. The goal of simulation is to estimate in advance (before the trial) the average dropout rate for feasible combinations of withdrawal criteria and the treatment titration schedule. The probability of dropping out of the trial is presented as a sum of two probabilities associated and not associated with the effectiveness of the treatment and withdrawal criteria. The first probability is obtained by simulation and the second one is estimated using data from previous similar trials. The simulated clinical trials resemble diabetes clinical trials with repeated measures design where there is a measured continuous variable (e.g., fasting plasma glucose) that is used both as an efficacy parameter and as the indicator for withdrawal if it exceeds a certain threshold (constant or variable in time). Some illustrative examples based on those simulations are given to investigate the sensitivity of the dropout rate to the threshold variations.

A Unifying Approach for Surrogate Marker Validation Based on Prentice's Criteria

◆ Helena M. Geys, Limburgs Universitair Centrum; Ariel Alonso, Limburgs Universitair Centrum; Geert Molenberghs, Limburgs Universitair Centrum

Limburgs Universitair Centrum, Center for Statistics, Universitaire Campus, Bldg. D, Diepenbeek, B-3590 Belgium

helena.geys@luc.ac.be

Key Words: *hierarchical models, surrogate markers, likelihood reduction factor, Prentice criteria*

Part of the recent literature on the evaluation of surrogate endpoints starts from a multitrial context which leads to a definition of validity in terms of the quality of both trial-level and individual-level association. Those authors proposed their methodology based on the simplest cross-sectional case in which both the surrogate and the true endpoint, are continuous and normally distributed. Different variations to the theme have been implemented for binary responses, times to event, combinations of binary and continuous endpoints, etc. However, a drawback of this methodology is that different settings have led to different definitions to quantify the association at the individual level. In the longitudinal setting, Alonso (2003b) defined a class of canonical correlation functions that can be used to study surrogacy at the trial and individual level. We propose a new approach to evaluate surrogacy in the repeated measurements framework, we also show the connection between this proposal and the previous ones reported in the literature and finally we extend this concept to the non-normal case using some of the criteria defined by Prentice.

Longitudinal Data Analysis—Is "Optimal" Modeling of the Covariance Structure Necessary?

◆ Devan V. Mehrotra, Merck & Co., Inc.; Geert Molenberghs, Limburgs Universitair Centrum; Russell D. Wolfinger, SAS Institute Inc.

Merck & Co., Inc., UN-A102, 785 Jolly Rd., Bldg. C, Blue Bell, PA 19422

devan_mehrotra@merck.com

Key Words: *compound symmetry, missing data, REML, repeated measures, sandwich estimator, unstructured covariance*

In a typical comparative clinical trial, subjects are randomized to receive either treatment A or B. For each subject, the response of interest (e.g., blood pressure) is measured at baseline and at fixed time points after initiation of treatment. The presumed primary objective is to test the hypothesis that the true mean response at the end of the treatment period is the same for A and B. To address this objective, the response data are commonly analyzed using a "cell means" linear model approach based on the restricted maximum likelihood (REML) methodology. A key component of the REML-based analysis is modeling of the dependency between

intrasubject responses at different time points, often referred to as the "covariance structure." Two strategies are commonly used in practice. In the first, the covariance structure to be used is specified in advance (compound symmetry, AR(1), etc.). In the second, an attempt is made to select the "optimal" covariance structure for the data in hand from a menu of available structures, based on a particular information criterion (e.g., AIC). We discuss two alternate strategies, one based on the so-called "sandwich" covariance.

164 Estimation and Testing in Clinical Trials

Biometrics Section, ENAR, Section on Statistics in Epidemiology

Monday, August 9, 2:00 pm–3:50 pm

Symmetrized Percent Change for Treatment Comparisons

◆ Gregory D. Ayers, University of Texas M.D. Anderson Cancer Center; Donald Berry, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, Dept. of Biostatistics and Applied Mathematics, 1515 Holcombe Blvd.-447, Houston, TX 77030-4009

gayers@mail.mdanderson.org

Key Words: *change from baseline, expanded baseline eligibility, measure of treatment effect, percent change, required sample size, robust percent change*

A standard way of addressing treatment effect is to relate post-treatment (F) with the corresponding baseline (B) measurement. Commonly used measures of effect are absolute difference, D = F-B, and percent change, PC = 100*(F-B)/B. A less commonly used measure was suggested by Berry (1989): symmetrized percent change, SPC=100*(F-B)/(B+F). We examine the interpretability and relative performances of these three measures in the context of a parallel, two-armed clinical trial. The conclusions are sometimes the same for all three measures. But they may well be different. We show by simulating a realistic setting that analyses based on SPC tend to be more powerful than those employing the other two measures. For the cases we consider the savings in sample size range from 20-70% when using SPC as compared with using one of the other measures.

Performance of Floating Absolute Risks

◆ Patrick G. Arbogast, Vanderbilt University

Vanderbilt University, Dept. of Biostatistics, S-2323 Medical Center North, Nashville, TN 37232-2158

patrick.arbogast@vanderbilt.edu

Key Words: *confidence intervals, hazard ratio, proportional hazards regression, relative risk, survival analysis*

A recent investigation of hormone-replacement therapy and breast cancer risk used a method called "floating absolute risks" (FARs) to compute confidence intervals for relative hazards. This method has been reported in other medical studies and has received controversy as to its appropriateness. However, there has been no direct comparison of the FAR method to the standard approach for computing confidence intervals from proportional hazards regression. I report simulation results comparing these two methods and demonstrate that the FAR method can substantially underestimate the width of confidence intervals for relative hazards.

On Appropriate Sample Size for Using Normal Approximation in Clinical Trials

◆ Jun Zhao, Organon Pharmaceuticals Inc.; Gang Li, Organon Pharmaceuticals Inc.

Organon Pharmaceuticals Inc., 56 Livingston Ave., Roseland, NJ 07068

j.zhao@organonusa.com

Key Words: *sample size, skewness, Type I error, normal approximation, p value*

To evaluate the effectiveness of study drug in clinical trials, p values for hypothesis testing are often obtained by using normal approximation, such as in t-test, ANOVA, and chi-square test. But clinical data are rarely normally distributed. A question often raised is how large a sample is appropriate for using normal approximation to obtain p values. One important factor that hinders normal approximation is the skewness of data. The larger the skewness, the greater the sample size is needed for normal approximation. Instead of examining the approximation on a whole distribution function, our focus is on the tail probability (the Type I error), which is used as a criterion to evaluate the difference of treatment effects. For each level of skewness, a minimal sample size for which the normal approximation approach can control the Type I error is suggested.

A Novelty Test for Testing Treatment Differences under Unparallel Studies

◆ Yi-Ting Hwang, National Taipei University; Chien-Hua Wu, Chung Yuan Christian University; Yu-Sheng Wang, National Taipei University

National Taipei University, 67, Sec. 3, Ming-Sheng E. Rd., Dept. of Statistics, Taipei, Taiwan

hwangyt@mail.ntpu.edu.tw

Key Words: *analysis of covariance, unparallel studies, F test, Johnson-Neyman technique*

The Analysis of Covariance (ANCOVA) has been widely implemented in the various areas to assess the effect of the treatments on the variable of interest controlling for the effects of prognostic variables. Although ANCOVA is a major tool in the statistical analysis, it requires many assumptions such as normality assumption of the error distribution and homogeneity of variance within groups. Furthermore, the coefficient on the prognostic variable is usually assumed to be the same accross treatment groups since the derivation is simple. There are only limited studies that can handle the unparallel problem. However, it requires the mean values of the two prognostic variables are equal. A new test adopting the idea from regression estimators is proposed for testing treatment differences under unparallel studies. The

proposed method preserves descent power and Type I error under parallel or unparallel studies, whereas the regular F test is only good for the parallel studies.

Sample Size Requirements for Randomization within Clusters

◆ Miu Lam, Queen's University; Richard Birtwhistle, Queen's University; Marshall Godwin, Queen's University

Queen's University, Dept. of Community Health & Epidemiology, Kingston, ON K7L 3N6 Canada

lamm@post.queensu.ca

Key Words: *stratified randomization, cluster randomization, random effects, intraclass correlation*

Much of the development for sample size requirements have been either on stratified randomization, cluster randomization, or their combined stratified cluster randomization designs. Sample size calculation for a study design involving randomization within clusters has not been found in the literature. Sample size requirements will be derived when randomization of subjects to two treatment groups is carried out within each cluster. Since subjects within the same cluster are not independent of each other, intracluster correlation is involved in determining the total number of clusters required to compare the two treatment groups. Formulas are derived for the comparison of two means and the comparison of two proportions. Results show that if observations are independent, in particular, the formulas reduce to the usual formulas for the two independent sample comparisons. For a constant cluster size of two, results are comparable to the usual sample size calculation for paired-sample comparisons.

Data Analysis under Case-cohort Designs for Cluster Randomized Trials

◆ Shou-En Lu, University of Medicine and Dentistry of New Jersey

University of Medicine and Dentistry of New Jersey, 683 Hoes Lane West, Piscataway, NJ 08854

lus2@umdnj.edu

Key Words: *case-cohort design, clustered binary outcome data, clustered failure-time data, marginal models*

Case-cohort design is an efficient and economical design to study risk factors for infrequent disease in a large cohort. In the literature, the case-cohort design has been extensively studied, but was exclusively considered for univariate outcome data. We propose the case-cohort designs for cluster-randomized trials where the investigators wish to evaluate the population averaged treatment effect. Estimation procedures for clustered binary and failure time outcomes are studied, respectively. The performance of the proposed estimators and comparisons of statistical efficiencies are investigated with simulation studies. A data example from a child vitamin A supplementation trial in Nepal (Nepal Nutrition Intervention Project-Sarlahi) is used to illustrate the proposed methodology.

165 Questionnaire Design and Reducing Respondent Error ▲

Social Statistics Section, Section on Government Statistics, Section on Survey Research Methods

Monday, August 9, 2:00 pm–3:50 pm

When Less is More: Are Reluctant Respondents Poor Reporters?

◆ Ting Yan, JPSM, University of Maryland/Abt Associates; Roger Tourangeau, University of Maryland; Zac Arens, University of Maryland

JPSM, University of Maryland/Abt Associates, 1218 LeFrak Hall, College Park, MD 20742

tyan@survey.umd.edu

Key Words: *response rate, nonresponse error, measurement error, response bias*

Survey research has been witnessing declining response rates across a wide range of surveys. As a result, extensive resources have been invested in boosting response rates under the assumption that a higher response rate will lead to lower survey error. However, the extra attempts to reduce nonresponse are costly. Given limited budgets, survey researchers might be more confident about allocating a larger share of survey budgets to decreasing nonresponse if they could be assured that an improved response rate does in fact lower survey error. In fact, the empirical evidence in this regard is mixed. Some studies show respondents interviewed on early calls differ from those interviewed on later calls. However, other recent studies seem to demonstrate that large differences in response rate have only minor effects on cross-sectional analyses. This study, analyzing a national RDD survey data that measures the public knowledge of and attitudes towards science, replicates the work of Curtin et al. (2000) by looking at the impact of response rates on key survey variables from another topic area. In addition, this study extends that work by exploring the relationship between the nonresponse errors and the measurement errors in a national survey. It will focus on answering the question whether reluctant respondents have more or less measurement error.

Anchoring Questions in the Respondent-generated Intervals Protocol

◆ Judith M. Tanur, SUNY, Stony Brook; S. James Press, University of California, Riverside; LiPing Chu, University of California, Riverside

SUNY, Stony Brook, PO Box 280, Montauk, NY 11954

jtanur@notes.cc.sunysb.edu

Key Words: *anchoring, Bayes surveys, bounds, RGI, hierarchical model*

The Respondent-Generated Intervals protocol (RGI) has been used to have respondents (Rs) answer a numerical factual question by giving both a point estimate and also bounds within which they feel almost certain that the true value of the quantity being reported on falls. Some new anchoring questions elaborate the RGI protocol, aiming to improve the accuracy of estimators derived from it. Because point estimates of Rs who give short intervals are weighted more heavily in the Bayesian RGI estimator than point estimates of Rs who give longer ones, it is advantageous to cue and encourage more accurate Rs to give shorter intervals and less accurate Rs to give longer ones. We describe preliminary results of an experiment embedded in a survey to test this new thinking. We introduce mechanisms to direct confident (thus presumably accurate) Rs to give shorter intervals and less confident (thus presumably less accurate) Rs to give longer ones. The experimental design varies the instructions about how Rs should construct their intervals. We also present a new Bayesian estimation procedure that depends upon both the means and the standard deviations of the bounds to further improve accuracy.

A Comparison of Perception-based and Behavior-based Measures on Sexual Orientation in a Telephone Health Survey

◆ Wei Yen, University of California, Los Angeles; Charles DiSogra, University of California, Los Angeles; David Grant, University of California, Los Angeles; Rong Huang, University of California, Los Angeles

University of California, Los Angeles, 10911 Weyburn Ave., Suite 300, Center for Health Policy Research, Los Angeles, CA 90024

weiyen@ucla.edu

Key Words: *sexual orientation, California Health Interview Survey, data quality, homosexual, gay/lesbian, bisexual*

The 2003 California Health Interview Survey included two measures on sexual orientation: perception-based (PB) and behavior-based (BB). The PB measure is based on the question "Do you think of yourself as straight or heterosexual, as gay, lesbian, or homosexual, or as bisexual?" The BB measure consists of a question on the number of sexual partners during the past 12 months and a question on sex of the partners. Common categories of homosexual (including gay and lesbian), bisexual, and heterosexual are constructed for each measure and compared between the two measures. Results from the first half of the survey show that with the PB measure, 2% of the respondents reported homosexual, 1% bisexual, and 95% heterosexual among adults age 18-70. Correspondingly, with the BB measure, 2% of the respondents reported homosexual, 0.3% bisexual, 76% heterosexual, and another 20% reporting no sexual partners in the past 12 months. The concordance between the two measures for homosexual persons is 60% and a very low 15% for bisexual persons. This paper further discusses possible factors, including data limitations, that may contribute to the discordance between the two measures.

Patterns of Proxy Usage in the 2001 National Household Travel Survey

◆ Jonaki Bose, Bureau of Transportation Statistics; Lee Giesbrecht, Bureau of Transportation Statistics

Bureau of Transportation Statistics, 400 7th St. SW, Room 4432, Washington, DC 20590

jonaki.bose@bts.gov

Key Words: *NHTS, travel data, proxy responses*

The 2001 National Household Travel Survey is a nationally representative survey on the daily and long distance travel patterns of the nation. Data were collected from about 26,000 households about the travel behavior of all household residents. For adults, 18 and older, a proxy response was accepted if the person was unavailable to participate. Overall, data were collected from proxies for 22% of the adult respondents. This paper evaluates the quality of the data collected from proxy respondents, compares complete households with and without any proxy data, and examines the potential characteristics that might predict whether a person will have a proxy response. The ideal way to evaluate the quality of proxy data would be to have proxy- and self-reported data for each respondent. Despite the absence of such an experimental set-up, this study attempts to understand the phenomenon using survey data. It highlights some of the difficulties in deciphering whether differences in estimates between proxy and self respondents are due to true differences or measurement error, especially since the need for a proxy respondent may be directly linked to the topic of the survey.

Score Redevelopment Utilizing Factor Analysis and Expert Knowledge: Creating a Briefer BERS

◆ Laura S. Morris, Policy Research Associates; Karen Hennigan, Social Science Research Institute, Inc.; Lisa Russell, ETR Associates; Debbie Rinehart, Arapahoe House, Inc.; Nina Kammerer, Health and Addictions Research, Inc.

Policy Research Associates, 345 Delaware Ave., Delmar, NY 12054

lmorris@prainc.com

Key Words: *scale development, strength-based assessment, multivariate analysis*

The Behavioral and Emotional Rating Scale (BERS) is one of the leading strength-based instruments used to assess children's functioning. As part of the Women, Co-occurring Disorders and Violence Children's Subset Study, the BERS was administered to 253 women with co-occurring mental health and substance use disorders and histories of trauma, to describe their children. We were unable to test the subscale structure using confirmatory factor analysis due to a singular matrix. An exploratory factor analysis (EFA), however, did not support the published subscale structure. Additionally, feedback from interviewers and clinicians raised concerns about the relevance and conciseness of the BERS's 52 individual items and five subscales. To address these analytic and content-oriented concerns, feedback from consumers, clinicians, and researchers was integrated with EFA and reliability analyses to redevelop the scale. Through this process a new scale was constructed containing 17 items and four subscales, resulting in an increased percentage of explained variation. Preliminary validation suggests that this briefer BERS is comparable to the original BERS for assessing this population.

Achieving the Connection between Usability and Advanced Visual Design Principles for Designing Establishment Surveys

◆ Don A. Dillman, Washington State University; Arina Gersteva, Washington State University

Washington State University, 133 Wilson Hall, Pullman, WA 99164-4014

dillman@wsu.edu

Key Words: *visual layout, establishment surveys, usability, survey*

Published experimental evidence now exists that relatively simple changes in the visual layouts of questionnaires can significantly influence how respondents interpret and respond to items in self-administered questionnaires. However, past research has tended to emphasize relatively simple manipulations such as the addition of arrows such as linear vs. nonlinear layouts of answer categories. To date virtually no attempts have been made to apply more sophisticated or complex visual appearance concepts to redesigning self-administered questionnaires to ease the task of completion. Our purpose is to develop the relevance of several of these more advanced visual design concepts that go well beyond those available in the survey literature and demonstrate their applicability to designing self-administered establishment surveys that use matricies and other complicated response elements. A conceptual linkage is made between usability concepts developed by Donald Norman and their realization through the advanced visual design concepts discussed.

Family Income Response Patterns for Varying Levels of Income Detail: An Analysis of the National Health Interview Survey

◆ John R. Pleis, Centers for Disease Control and Prevention; James Dahlhamer, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, National Center for Health Statistics, Division of Health Interview Statistics, 3311 Toledo Rd., Room 2319, Hyattsville, MD 20782

JPleis@cdc.gov

Key Words: *family income, socioeconomic status (SES), ordinal*

The NHIS, an ongoing population-based health survey conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention, allows respondents to report family income information in a variety of ways (exact dollar amount, income interval, income in reference to \$20,000 (greater than or equal to \$20,000, less than \$20,000), as well as no income reporting). Using data from 2001-2002, this analysis examines patterns of income level reporting by respondent, family, household, and geographic characteristics. Preliminary results indicate that age, race/ethnicity, education, marital status, employment history, number of adults, region, and the number of income sources are associated with the level of income detail provided. For example, lower educated respondents generally provided less income information than their higher educated counterparts. Final analytic results will provide information that could prove useful for future question design efforts.

166 Family Studies in Genetic Epidemiology ▲

Section on Statistics in Epidemiology, Biometrics Section

Monday, August 9, 2:00 pm–3:50 pm

Distinguishing the Presence of Common Risk Factors from a Causal Relationship When Two Traits Co-aggregate

◆ Fanesca Young, Columbia University; Daniel Rabinowitz, Columbia University

fanesca@yahoo.com

Key Words: *familial co-aggregation, permutation test, febrile seizure, epilepsy*

Familial co-aggregation of two traits may result from the influence of common risk factors or from the influence of one trait on the other in the presence of aggregation of the former. It is shown that there are sources of information in the patterns of co-aggregation of the traits in family data for detecting one or another of the two kinds of influences, and for detecting, in the presence of the influence of one trait on the other, the influence of common risk factors. A systematic investigation of the sources of information reveal that in many cases, methods based on known population rates for the traits are vastly superior to permutation tests. The influence of various trait-based sampling schemes on the power to distinguish the two types of influence is examined and the results are used to provide guidance for study design. Data on co-aggregation of febrile seizure and epilepsy are used for illustration.

When Does Studying Only the Higher Relative Risk Subgroup Increase Efficiency for Evaluating an Effect?

◆ Abhijit Dasgupta, National Institutes of Health; Sholom Wacholder, National Institutes of Health

National Institutes of Health, Biostatistics Branch, DCEG, NCI, DHHS, 6120 Executive Blvd., Suite 8022, Bethesda, MD 20852-7244

dasgupab@mail.nih.gov

Key Words: *subgroup analysis, study design, sample size, relative efficiency, gene-environment interaction*

Many studies are designed to assess the joint effects of genetic and environmental factors in the development of disease. We consider the efficiency of different designs for testing whether a dichotomous environmental factor X is related to disease. The alternative hypothesis is that X is related to disease in at least one subgroup defined by dichotomous genetic factor G. When the relative effect of X is the same in both subgroups, clearly there is no efficiency advantage in determining G. On the other hand, if the effect of X is present only when G=1, a strategy that excludes all cases and controls with G=0 will be more efficient. We investigate the relative efficiency of these two strategies over a wide range of realistic scenarios. We find that often there is no benefit in efficiency from using G to choose subjects for measurement of X, even when the relative risk due to X is much greater in G=1 than in G=0. We provide a spreadsheet to assess the relative efficiencies of the strategies for arbitrary joint distributions of X and G and arbitrary joint effects models. We also identify situations where a particular strategy is at least 10% more efficient.

Nonparametric Estimation from Kin-cohort Data

◆ Yuanjia Wang, Columbia University; Daniel Rabinowitz, Columbia University

yw2016@columbia.edu

Key Words: *kin-cohort, missing covariats, effciency bounds, nonparametrics*

In kin-cohort studies, genotype and disease status are obtained from a group of volunteers (probands), and disease history is obtained from relatives of the probands. The relatives of the probands are not genotyped but form a retrospective cohort for the disease. Investigators are often interested in assessing the distribution of gene-related phenotypes as a function of deleterious mutation. For rare genetic variants, it can be difficult to obtain adequate number of carriers who develop disease for accurate estimation of the effect. Several parametric and semiparametric approaches have been proposed to estimate the influence of genotype on the distribution of disease status. Here, a class of nonparametric estimators is developed, and the optimal member of the class is derived. The optimal estimator in the class lies in the tangent space of the nonparametric model, and is therefore efficient among the class of regular unbiased estimators. A two-step approach to computing the estimator is described. Some simulation results are shown, and the method will be applied to Parkinson's disease.

Examining Relationships in Heterozygous Triplets

◆ Gene S. Fisch, North Shore LIJ Research Institute; Nina Kohn, North Shore LIJ Research Institute; Catherine Derom, Catholic University, Leuven; Martin L. Lesser, North Shore LIJ Research Institute

North Shore LIJ Research Institute, 1129 Northern Blvd., Suite 302, Manhasset, NY 11030

gfisch@nshs.edu

Key Words: *intraclass correlation, triplets, twins*

The intraclass correlation coefficient (ICC) computes the correlation between $k>2$ unordered measures, e.g., IQ in twins. In triplets, the ICC is applied when triplets are monozygotic (MZ) or trizygotic (TZ). When the triplet is heterogeneous—identical MZ twins plus a DZ sib—the ICC is not valid; nor is there valid procedure to examine the relationship. We propose to measure relationship in $n=7$ such sets of triplets by constructing all possible 2^n vectors of MZ twins that could be paired with their respective DZ sib, and compute their ICCs. Mean ICC for all 27 ICCs was used to measure relationship between MZs and the DZ sib. Our heterogeneous dataset had higher than expected mean ICC=.75, the distribution of which was unexpectedly bimodal. We used a jackknife procedure to identify the source of bimodality and found one influential triplet. The DZ from that triplet had an unusually high IQ compared to its MZ siblings. In 6/7 triplets, the DZ IQ was either higher or lower than the MZ IQs (p=.26). To analyze the effect on the ICC distribution, we lowered DZ IQ score in that triplet and found that, as it decreased, the distribution of ICCs became increasingly unimodal.

Modeling the Effects of Maternal-fetal Genotype Incompatibility on Disease Risk When Fetal Viability is Reduced

◆ Hsin-Ju Hsieh, University of California, Los Angeles; Christina Palmer, University of California, Los Angeles; Janet Sinsheimer, University of California, Los Angeles

University of California, Los Angeles, 3767 Mentone Ave. #414, Dept. of Biostatistics, Los Angeles, CA 90034

hjhsieh@ucla.edu

Key Words: *log-linear model, genetic epidemiology, genotype interaction, conditional likelihood, gene-mapping, confounders*

Genotype incompatibility between a mother and fetus can create adverse prenatal conditions that can potentially increase disease susceptibility, even into adulthood. Maternal-fetal genotype (MFG) incompatibility may also reduce fetal viability, so that potentially affected individuals are not observed. Sinsheimer et al. (2003) proposed the MFG test, a loglinear model, using case-parent trios to examine incompatibility but the effect estimated from their method cannot distinguish between a fetal viability and an incompatibility effect. We propose a new method that extends the multiple sibling version of the conditional likelihood model of Kraft et al. (in press). We model both incompatibility and fetal viability in a joint distribution of the children's and parents' genotypes conditional on the children's affection status. The model is robust and has the appropriate Type I error rate. Through simulation, we show that moderate sample sizes can provide significant power to detect MFG incompatibility effect on disease risk when fetal viability is reduced, and ignoring fetal viability in analysis may underestimate disease relative risk due to the MFG incompatibility.

Apo-E Genotypes and Cardiovascular Diseases: A Sensitivity Study

◆ Mark E. Glickman, Boston University

Boston University, EN Rogers Mem. Hospital (152), Bldg. 70, 200 Springs Rd., Bedford, MA 01730

mg@bu.edu

The Apo-E gene, a gene that produces proteins which help to regulate lipid levels in the bloodstream, is of interest in the study of cardiovascular diseases. An approach to making inferences about the genetic effects of the Apo-E gene has been developed by Glickman and Gagnon (2002). The framework describes the role of genetic and risk factors on the onset ages of multiple diseases, and accounts for missing data due to censoring, and the possibility that an individual was censored for reasons related to the diseases of interest. We apply an extension to this framework to the original cohort of the Framingham Heart Study for measuring the effects of different Apo-E genotypes on the onset age of various cardiovascular disease events. In particular, we compare the fit of univariate vs. multivariate onset age components to the model, whether to incorporate health covariates measured at baseline or at a point later in the study, and whether to assume a heritability model for Apo-E genotype frequencies. The results of the best-fitting model are presented.

Fine-scale Mapping of Disease—Extending the Bayesian Partition Model

◆ John Molitor, University of Southern California; Paul Marjoram, University of Southern California; Duncan Thomas, University of Southern California

University of Southern California, 1540 Alcazar St., Los Angeles, CA 90089

jmolitor@usc.edu

Key Words: *haplotype, Markov chain Monte Carlo, fine-mapping, Voronoi tesselations, Bayesian methods*

We present a number of extensions to our fine-scale mapping method. Specifically, we examine the effectiveness of the method with regards to its ability to handle diploid only data. This extension entails the use of a latent structure to estimate, at each step of the MCMC algorithm, individual haplotypes based upon available genotype data. The amount of information lost when only genotypical information is available will be explored. Further, if time permits, we will examine ways in which more population genetic theory can be incorporated into the model and examine the amount of improvement that is gained with the introduction of this genetic theory versus the increase in simulation time.

167 President's Invited Address

ASA, ENAR, IMS, SSC, WNAR

Monday, August 9, 4:00 pm–5:50 pm

The Romance of Hidden Components

◆ David L. Donoho, Stanford University

Stanford University, Humanities and Sciences Statistics Dept., Stanford, CA 94305

donoho@stanford.edu

Perhaps the most romantic and seductive idea in all of science is that, hiding behind the enormously complex structures we see in the world around us, there are hidden components that are, on the one hand, very simple and even elegant, and on the other hand easily combine to generate all the variety we see about us. Classical examples include Newton and the spectrum of light, eugenecists and the idea of IQ; modern examples include wavelets and quarks. I will review some of the classical ideas of hidden components, starting from principal components or even before, and describe some of the most recent notions, such as independent components analysis, sparse components analysis, nonnegative matrix factorizations, and cumulant components. I will try to keep things at an elementary level, communicating the attractiveness of these ideas to scientists and engineers outside of statistics, the wide-ranging impact these ideas are having from high-tech industry to neuroscience and astronomy, and describing what I think is the much greater role that statisticians should be playing in developing and deploying these methods.

168 Introductory Overview Lecture on Bayesian Statistical Computation and Its Applications

ASA, ENAR, IMS, SSC, W.H. Freeman & Company

Tuesday, August 10, 8:30 am–10:20 am

Bayesian Statistical Computation

◆ Steve Brooks, University of Cambridge

University of Cambridge, Mathematics Dept., Cambridge, CB2 1TN UK

s.p.brooks@statslab.cam.ac.uk

We illustrate the general principles and practical issues of Bayesian computation through a series of examples from our applied work in the social and biological sciences. Modern computational methods allow us to fit increasingly complicated and realistic models to data. We discuss issues of fitting, checking, and expanding models from a Bayesian perspective.

Applied Bayesian Data Analysis

◆ Andrew Gelman, Columbia University

Columbia University, Dept. of Statistics, 23 Mathematics Bldg., New York, NY 10027

gelman@stat.columbia.edu

Computational developments over the past 20 years have accompanied a resurgence of interest in Bayesian methodology which, though once the dominating paradigm, had lain almost dormant for much of the last century. We introduce modern Bayesian computational methods, starting with the Gibbs sampler and the Metropolis algorithms that improve simulation efficiency. We discuss the theoretical and practical issues of implementation and illustrate with a series of examples of Bayesian inference in our applied research in biology and the social sciences.

169 Recent Development of Statistical Methods for Recurrent Events Data ⌘

Biopharmaceutical Section, WNAR, ENAR

Tuesday, August 10, 8:30 am–10:20 am

Maximum Likelihood Estimation in Transformation Models for Counting Processes

◆ Danyu Lin, University of North Carolina; Donglin Zeng, University of North Carolina

University of North Carolina, Chapel Hill, NC 27599

lin@bios.unc.edu

Key Words: *censoring, intensity function, multivariate failure times, recurrent events, semiparametric efficiency, survival analysis*

We propose a class of semiparametric transformation models with random effects for the regression analysis of counting processes with censored observations. We establish that the nonparametric maximum likelihood estimators for the parameters of interest are consistent and asymptotically Gaussian. Furthermore, the limiting variances achieve the semiparametric efficiency bounds and can be consistently estimated. Simulation studies show that the likelihood-based inference procedures perform well in practical settings. Illustration with a real medical study is provided.

Analyzing Longitudinal Data with Informative Censoring: Repeated Measurements vs. Recurrent-event Data

◆ Mei-Cheng Wang, Johns Hopkins University

Johns Hopkins University, Dept. of Biostatistics, School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205

mcwang@jhsph.edu

Key Words: *informative censoring, repeated measurements, recurrent events*

Repeated measurements and recurrent events can both be classified as longitudinal data since they are commonly observed in longitudinal follow-up studies. In general, statistical methods developed for data from one of the two categories are not automatically applicable to data from the other category because of differences in the data characteristics. This talk presents an informative censoring model for data of each category, and shows that under the specified models these two kinds of data share some common features. These features give us advantages to develop statistical methods which are applicable to both kinds of data. The "shape" and "size" parameters are introduced to characterize, model, and analyze repeated measurements and recurrent event data.

Robust Tests for Recurrent Events Occurring over Multiple Treatment Periods

◆ Richard J. Cook, University of Waterloo; Wei Wei, University of Michigan; Grace Y. Yi, University of Waterloo

University of Waterloo, 200 University Ave. West, Waterloo, ON N2L 3G1 Canada

rjcook@uwaterloo.ca

Key Words: *recurrent events, marginal rate, conditioning, robustness, pseudo-score test*

Robust methods for the analysis of recurrent events have been developed based on Poisson estimating functions. We describe an alternative robust approach suitable for settings in which recurrent events occur over multiple treatment periods and treatment comparisons are of primary interest. The approach is motivated by a "working" mixed Poisson model in which suitable conditioning can eliminate the subject-specific effects. A robust variance is derived for the resulting pseudo-score statistic. The relative efficiency of the conditional analyses versus a marginal analysis is examined, and the robustness of the new test is demonstrated through simulations. The method is illustrated by application to data from an asthma trial.

Regression Methods for Multiple-sequence Recurrent-event Data

◆ Douglas E. Schaubel, University of Michigan; Jianwen Cai, University of North Carolina, Chapel Hill

University of Michigan, Dept. of Biostatistics, 1420 Washington Hts, SPH-II, M4140, Ann Arbor, MI 48109-2029

deschau@umich.edu

Key Words: *missing data, multiple imputation, proportional means model, weighted estimating equations*

Censored recurrent event data frequently arise in biomedical studies. Often, the events are not homogenous, and may be categorized. An analysis which incorporates event categorization may be much more informative than one which aggregates across event categories, particularly when estimating covariate effects. The resulting data structure may be viewed as multiple recurrent event sequences. We propose semiparametric regression methods for analyzing such multiple-sequence recurrent event data. We also consider the setting where event times are known, but the event category may be missing. Parameter estimators are shown to be consistent and asymptotically normal, while finite sample properties are examined through simulation. The proposed methods are applied to an end-stage renal disease dataset obtained from a national organ failure registry.

170 Evaluating the Quality of Small Area Data from the American Community Survey: Evidence from Four Test Sites △

Social Statistics Section, Section on Government Statistics, Section on Statistics in Epidemiology, Section on Survey Research Methods

Tuesday, August 10, 8:30 am–10:20 am

From City to Farm: The View from California

◆ Linda Gage, California Dept. of Finance

California Dept. of Finance, Demographic Research Unit, 915 L St., Sacramento, CA 95814

Linda.Gage@dof.ca.gov

Key Words: *census, American Commuity Survey, response rates, quality measures, population*

The American Community Survey is designed to replace the census 2010 long form. The survey has been conducted in several counties across the country for several years. This research explores the measurement of population, social, economic, and housing data at the county and census-tract level for the American Community Survey sites in California: San Francisco and Tulare counties. Population and housing attributes from the 2000 decennial census long form are compared to individual and three-year pooled data from the American Community Survey to assess the differences and accuracy of the two measurements. In addition to the census and survey data, several data quality measures are reviewed.

These measures include standard errors, self-response rates, sample unit nonresponse rates, and sample completeness ratios, as well as total, self-response, and interviewer-response population and housing item allocation rates. Additionally, several characteristics, such as nativity and citizenship are compared at the census-tract level.

Seasonal Swings in Population: What Can We Learn from the ACS?

◆ Paul R. Voss, University of Wisconsin, Madison; Roger B. Hammer, University of Wisconsin, Madison; Paul M. Van Auken, University of Wisconsin, Madison; Daniel L. Veroff, University of Wisconsin, Madison

University of Wisconsin, Madison, Dept. of Rural Sociology, 1450 Linden Dr., Madison, WI 53706

voss@ssc.wisc.edu

Key Words: *American Community Survey, seasonal populations, Census long form, demographic analysis*

This paper examines monthly pooled data from the Census Bureau's American Community Survey (ACS) for the period 1999 through 2001, and compares attribute and quality measures from the ACS with comparable information from the 2000 Census. The purpose is to understand differences between the ACS and census data with an eye on the critical question of whether the ACS can be viewed as an adequate replacement in the year 2010 for the census long-form sample traditionally carried out in conjunction with the decennial census enumeration. Our analysis focuses on four counties that are part of the Census Bureau's ACS test sites: Oneida and Vilas counties in Wisconsin and Flathead and Lake counties in Montana. Portions of our analysis focus specifically on differences between census results and ACS results that may arise because of the increases in summer seasonal populations that are known to affect the monthly demographic metabolism of these northern U.S. counties blessed with environmental attributes that attract recreational visitors.

An Evaluation of the American Community Survey: Results from the Oregon Test Site

◆ David A. Swanson, University of Mississippi; George Hough, Portland State University

University of Mississippi, Dept. of Sociology & Anthropology, PO Box 1848, University, MS 38677-1848

dswanson@olemiss.edu

Key Words: *Census Long Form, American Community Survey, loss function, Census 2000, Multnomah County, Oregon*

Operational plans for the 2010 census call for ACS to replace the decennial Census long form (Census LF). Using a Loss Function Analysis and other tools, this paper reports findings from a comparison of ACS and Census LF data centered on the year 2000 for Multnomah County, Oregon. The findings suggest that there are notable differences between some of the corresponding variables—race and disability status are prominent in this regard. Other, less notable, differences are found within each of the four major areas of interest for which data are collected by ACS and Census LF: demographic; social; economic; and housing, with housing showing the greatest overall difrerences. The findings suggest reasons for some of the differences, which are described along with suggestions for future research on the sources of differences and ideas on how to use, display, and interpret ACS results.

How Does the ACS Measure Up in the Bronx, New York, Test Site?

◆ Joseph J. Salvo, NYC Dept. of City Planning; Arun P. Lobo, New York City Dept. of City Planning; Timothy Calabrese, NYC Dept. of City Planning

NYC Dept. of City Planning, 22 Reade St., New York, NY 10007

jsalvo@planning.nyc.gov

Key Words: *American Community Survey, data quality, small-area data, item allocation, city planning, neighborhoods*

The American Community Survey (ACS) is slated to replace the 2010 Census long form, based on the premise that the ACS is capable of creating an accurate and useful socioeconomic picture of sub-county areas, such as neighborhoods in large cities. This study compares estimates from the 1999-2001 ACS and the 2000 Census for 88 neighborhoods in Bronx County, New York. It asks whether estimates from the ACS can replace those from the census sample by drawing a picture of similarities and differences; meaningful differences are defined in this study as those that can affect the work of a large city planning agency. The results show that, despite some important conceptual differences, most estimates from the 1999-2001 ACS do not differ from those in the 2000 Census in a meaningful way; however, there are some important exceptions. Examining differences in the context of the data collection effort (mail return rates, household nonresponse rates) and in terms of data quality (item allocation levels) leads to a better understanding of differences between the two surveys.

171 *JBES* Invited Session

Business and Economics Statistics Section, *Journal of Business and Economic Statistics* (*JBES*)

Tuesday, August 10, 8:30 am–10:20 am

The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers

◆ Lars Vilhuber, Cornell University; John M. Abowd, Cornell University and U.S. Census Bureau

Cornell University, 6900 Rue France, Montreal, PQ H2G 2Z6 Canada

lars.vilhuber@cornell.edu

Key Words: *flow statistics, probabilistic matching, transitions, tenure, job flow, job creation, QWI*

We describe the sensitivity of small-cell flow statistics to coding errors in the identity of the underlying entities. Specifically, we present results based on a comparison of the U.S. Census Bureau's Quarterly Workforce Indicators (QWI) before and after correcting for such errors in SSN-based identifiers in the underlying individual wage records. The correction used involves a novel application of existing statistical matching techniques. It is found that even a very conservative correction procedure has a sizable impact on the statistics. The average bias ranges from 0.25% to 15% for flow statistics, and up to 5% for payroll aggregates.

172 Disclosure Limitation Methods for Protecting the Confidentiality of Statistical Data ▲

Section on Survey Research Methods, Section on Government Statistics, Section on Health Policy Statistics, Social Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Statistical Disclosure Techniques Based on Multiple Imputation

◆ Trivellore E. Raghunathan, University of Michigan; Roderick J. Little, University of Michigan; Fang Liu, University of Michigan

University of Michigan, Dept. of Biostatistics, School of Public Health, Ann Arbor, MI 48109

teraghu@umich.edu

Key Words: *multiple imputation, combining rules, partial synthesis, full synthesis, inferentially valid, Bayesian method*

Statistical disclosure control (SDC) is concerned with the modification of statistical data that contain confidential information on individual entities (persons, households, businesses, etc.), to prevent third parties from revealing sensitive information about identifiable individuals. In recent years, SDC has received increasing attention. Today's sophisticated computer technology and the increased access to data via the internet and electronic media allow data intruders to access information and identify individuals more easily, increasing concerns about respondent privacy. Confidentiality is vital for the future cooperation of respondents to provide high-quality data. On the other hand, policymakers need maximum information to make efficient and timely decisions. This paper summarizes a cluster of SDC techniques that protect against disclosure by deleting values of variables in the dataset and replacing them by values drawn from their predictive distribution. The imputation uncertainty is reflected by the method of multiple imputation (MI). The repeated sampling properties of inferences from the altered data are evaluated.

Characterizing Multidimensional Contingency Tables with Applications to Confidentiality and Statistical Disclosure Limitation

◆ Stephen E. Fienberg, Carnegie Mellon University; Aleksandra Slavkovic, Carnegie Mellon University

Carnegie Mellon University, Dept. of Statistics, Pittsburgh, PA 15213

fienberg@stat.cmu.edu

Key Words: *confidentiality, log-linear models, marginals, algebraic geometry, characterizations*

We describe alternative ways to provide partial characterizations of probability distributions associated with two-way tables of counts (contingency tables), using marginals, conditionals, and odds ratios and their generalizations to higher dimensions. Partial specification of such distributions arising in statistical disclosure limitation usually involves the dropping of components from complete specifications. We link several partial specifications to statistical models for contingency tables and to recent developments in algebraic geometry. These links offer insights for developing statistical methodology for disclosure limitation.

Ensuring Data Quality and Confidentiality for Tabular Data

◆ Lawrence H. Cox, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd. Room 3211, Hyattsville, MD 20782

lcox@cdc.gov

Key Words: *disclosure limitation, controlled tabular adjustment, mathematical programming*

Controlled tabular adjustment, a new disclosure limitation method for tabular data, eliminates the need to suppress tabular data cells for confidentiality purposes. Instead, cell values representing statistical disclosure are replaced by safe values and typically small adjustments are made to some or all nondisclosure cells to rebalance the tabulations. This solves the confidentiality problem and provides a complete set of tabulations on which the analyst can perform the full range of analyses as would have been performed on the original tabulations had disclosure not been present. The critical issue, then, is: Is it possible to perform the adjustment so that adjusted data will produce comparable analytical outcomes? We present approaches to this problem based on mathematical programming.

173 Where the Wild Things Are: Modeling Animal Movement ▲

Section on Statistical Graphics, Section on Statistics and the Environment

Tuesday, August 10, 8:30 am–10:20 am

Stochastic Differential Equations in the Analysis of Wildlife Motion

◆ David Brillinger, University of California, Berkeley

University of California, Berkeley, Dept. of Statistics, Berkeley, CA 94720-3860

brill@stat.berkeley.edu

Key Words: *diffusion models, nonparametric regression, potential functions, stochastic differential equations, vector field, wildlife examples*

The concern is with the use of stochastic differential equations (SDEs) to describe the motion of wildlife. There is special consideration of the case where the drift function comes from a potential function. There is study of SDEs whose parameters are periodic in time as these are useful for incorporating circadian rhythm in the

behavior. Nonparametric estimates of the drift and variance terms of an SDE model are discussed and evaluated and the fit of the model assessed. One issue is how to include explanatories, beyond location and time. Data from Hubbs Seaworld and the U.S. Forest Service's Starkey Experimental Station in Oregon are studied. The work is collaborative with researchers from those two institutions.

Incorporating Movement into Hidden Process Models

◆ Ken B. Newman, University of St. Andrews

University of St. Andrews, Centre for Research into Ecol. and Envir. Modellin, The Observatory, St. Andrews, KY16 9LZ Scotland

ken@mcs.st-and.ac.uk

Key Words: *sequential importance sampling, state-space model*

Hidden process models, which include state-space models as a special case, are a useful framework for simultaneously describing the evolution of an animal population and observations made of the population. Movement, natural mortality, birth, maturation, and man-induced mortality can be characterized as a sequence of stochastic subprocesses that link the state of the animal population at time t-1 to its state at time t. Computer-intensive procedures, such as sequential importance sampling, make inference for complex, nonlinear,and non-Gaussian processes feasible. A model for the British grey seal metapopulation which includes movement of maturing females from their natal colony will be presented.

State-space Models for Leatherback Turtle Migration

◆ Christopher Field, Dalhousie University; Joanna Flemming, University of Geneva; Ian Jonsen, Dalhousie University; Ransom Myers, Dalhousie University; Mike James, Dalhousie University

Dalhousie University, Mathematics and Statistics, Halifax, NS B3H 3J5 Canada

field@mathstat.dal.ca

Key Words: *state-space models, animal migration, leatherback turtle, robustness*

We discuss methods for analyzing animal trajectory data based on satellite tracking devices. To address the complexity of animal/environment interaction, and to accommodate the measurement error and process noise inherent in trajectory data, state-space models provide an appropriate mechanism. A state-space model focused specifically on quantifying the underlying navigational mechanisms inherent in animal movements is proposed for modeling the behavior of animals who are migrating, often over large distances. We utilize the term Circle of Confusion, first proposed by Kendall (1974), to describe the imprecision of animals in following a desired course. The modeling strategy enables us to statistically describe the Circle of Confusion, and allows us to build in robustness to deal with the outlying observations that characterize this type of data. One specific goal is to improve knowledge of the navigational abilities of the endangered leatherback turtles, something which has traditionally been poorly understood.

An Animal Movement Model Incorporating Resource Selection and Home Range

◆ Jay M. Ver Hoef, Alaska Dept. of Fish and Game; Aaron M. Christ, Alaska Dept. of Fish and Game

Alaska Dept. of Fish and Game, 1300 College Rd., Fairbanks, AK 99701

ffjmv@uaf.edu

Key Words: *Bayesian methods, hierarchical model, MCMC, utilization distribution, vector autoregression, simulation*

How an animal uses space and how it uses the habitat resources within that space are two common questions in biology. One major factor influencing the path an animal follows is the need for certain resources along that path. These resources may fill requirements for nutrition, shelter, reproduction, etc. We develop a unified model that accounts for both movement and the use of habitat implied by that movement. A vector autoregressive model is developed and subsequently modified by a habitat-indexed multiplicative constant such that the likelihood depends on a home-range distribution, a movement distribution, a habitat distribution. The resulting utilization distribution incorporates preference for and aversion to the various habitat types. Model parameters are estimated using Bayesian methods and MCMC. The method is generalized to multiple animals using a hierarchical framework. A simulation study shows that the current method outperforms previous resource selection methods. The method is applied to brown bear data from Southeast Alaska. We find that for these data, distance to stream is the most important covariate, while vegetation classes are also important.

174 Contemporary Challenges and Developments in Design ▲ ⌘

Section on Quality and Productivity, Section on Physical and Engineering Sciences, IMS

Tuesday, August 10, 8:30 am–10:20 am

Design Challenges in Microarray Experiments

◆ Patty Solomon, University of Adelaide

University of Adelaide, School of Mathematical Sciences, North Terrace, Adelaide, 5005 Australia

patty.solomon@adelaide.edu.au

Key Words: *microarrays, optimal experimental design, gene expression*

Microarrays are powerful tools for surveying the expression levels of thousands of genes simultaneously and have found wide application in the biomedical, agricultural, and pharmaceutical sciences. There are, however, many sources of uncertainty in microarray experiments and good experimental design is essential for answering the questions of interest to biologists accurately and precisely. Unfortunately, the matter of choosing a suitable design cannot be addressed intuitively and although there is a large and well-established literature on the design of experiments, little is

directly relevant to the microarray context. This presents a challenge to statisticians to find new methodology for generating and choosing designs. I will outline the key challenges and describe joint research with Gary Glonek on optimal designs for complex two-color microarray experiments. Our principles are to first identify the parameters of interest, then to seek designs that minimize the variance of the estimates subject to resource constraints. I will elucidate the issues with case studies on measuring gene expression in osteoarthritis, leukaemogenesis, and embryonic stem cells.

Design of Experiments for Screening, Exploration, and Optimization

◆ C.F. Jeff Wu, Georgia Institute of Technology

Georgia Institute of Technology, School of Industrial and Systems Engineering, 755 Ferst Dr., ISYE, Atlanta, GA 30332-0205

jeff.wu@isye.gatech.edu

Key Words: *factor screening, response surface, orthogonal arrays, optimization, projection aberration*

Orthogonal arrays are widely used in industrial experiments for factor screening. In many practical situations only a few of the factors will turn out to be important. Previous work shows that an orthogonal array can be used not only for screening factors but also for detecting interactions among a subset of active factors. A set of optimality criteria is proposed to assess the performance of designs for factor screening, projection, and interaction detection, and a three-step approach is proposed to search for optimal designs. Combinatorial and algorithmic construction methods are proposed for generating new designs. Level permutation methods are used for improving the eligibility and estimation efficiency of the projected designs. The techniques are then applied to search for best three-level designs with 18 and 27 runs. Many new, efficient and practically useful nonregular designs are found and their properties discussed. (Joint work with Hongquan Xu and Shaowei Cheng.)

The Design of Spatial Monitor Networks

◆ Richard L. Smith, University of North Carolina; Zhengyuan Zhu, University of North Carolina, Chapel Hill

University of North Carolina, Dept. of Statistics and OR, Chapel Hill, NC 27599-3260

rls@email.unc.edu

Key Words: *Bayesian statistics, spatial statistics, asymptotic approximations*

Spatial network design is concerned with the optimal placement of a network of monitors for a spatial field; a typical problem would be deciding where to place ozone monitors to obtain the best information about the ozone field over a region. Recent research has highlighted the contrast between "estimative" and "predictive" criteria: the optimal placement of monitors for achieving optimal prediction, given known parameters of the spatial field, is not necessarily the same as the optimal design for estimating those parameters. In a University of Chicago PhD thesis, Zhu (2002) proposed a compromise criterion that combines both aspects. Here we suggest an alternative approach based on Bayesian considerations. By focusing on the expected length of Bayesian predictive intervals, we obtain a criterion that effectively combines both estimative and predictive aspects. Implementation, however, requires the develop of suitable asymptotic approximations. These approximations, and their application to the design problem, will form the main focus of this talk.

175 Statistical Issues Arising in Model Selection and Model Uncertainty

Section on Statistical Computing, Section on Statistics in Epidemiology, Business and Economics Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Spike and Slab Variable Selection: Frequentist and Bayesian Strategies

◆ J. Sunil Rao, Case Western Reserve University; Hemant Ishwaran, Cleveland Clinic Foundation

Case Western Reserve University, Desk WG-57, School of Medicine, 10900 Euclid Ave., Cleveland, OH 44106

sunil@hal.cwru.edu

Key Words: *generalized ridge regression, model averaging, hypervariance, shrinkage*

Variable selection in the linear regression model takes many apparent faces from both frequentist and Bayesian standpoints. We introduce a class of variable selection models we refer to as rescaled spike and slab models. We study the importance of the prior hierarchical specifications and draw connections to frequentist generalized ridge regression estimation. Specifically, we study the usefulness of continuous priors to model hypervariance parameters, and the effect scaling has on the posterior mean through its relationship to penalization. We demonstrate the importance of selective shrinkage for effective variable selection in terms of risk performance. Using specialized forward and backward selection strategies, we study the effects of selection bias and model uncertainty and compare these to an ordinary least squares (OLS) model estimator formed without the benefit of model averaging. Simulations are used to illustrate the performance of the rescaled spike and slab variable selection technique.

Variable Selection for Survival Data Analysis

◆ Runze Li, Pennsylvania State University; Jianwen Cai, University of North Carolina, Chapel Hill; Jianqing Fan, Princeton University; Haibo Zhou, University of North Carolina, Chapel Hill

Pennsylvania State University, Dept. of Statistics, University Park, PA 16802-2111

rli@stat.psu.edu

Key Words: *penalized likelihood, SCAD*

In most model selection problems, the number of parameters may be large, and grows as the sample size increases. I propose a penalized-likelihood method for variable selection with survival data analysis with growing number of regression coefficients. Under

certain regularity conditions, we show the consistency and asymptotic normality of the penalized-likelihood estimators. We further demonstrate that, for certain penalty functions with proper choices of regularization parameters, the resulting estimate possesses an oracle property, namely, the resulting estimate can correctly identify the true model as if the true model (the subset of variables with nonvanishing coefficients) were known in advance. Using a simple approximation of the penalty function, the proposed method can be easily carried out with the Newton-Raphson algorithm. We conduct extensive Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures. We illustrate the proposed method by analyzing a real dataset.

The Large-sample Minimal-coverage Probability of Confidence Intervals in Regression after Model Selection

◆ Hannes Leeb, Yale University; Paul Kabaila, La Trobe University

Yale University, 24 Hillhouse Ave., Dept. of Statistics, New Haven, CT 06511

hannes.leeb@yale.edu

Key Words: *model selection, confidence intervals, minimal coverage probability, AIC, pre-testing*

We give a large-sample limit analysis of the minimal-coverage probability of confidence intervals for location parameters that are constructed based on the outcome of a “conservative” (or “overconsistent”) model selection procedure. We derive an upper bound for the large-sample limit minimal coverage probability of confidence intervals constructed after model selection that applies to a large class of model selection procedures, including AIC as well as various general-to-specific and specific-to-general pre-testing procedures. To our knowledge, this is the first analytical result on the large-sample-limit minimal-coverage probability of confidence sets constructed after model selection by AIC. This upper bound can be used as a monitor to identify situations where the actual coverage probability can be far below the nominal level.

Can One Estimate the Distribution of Post-model-selection Estimators?

◆ Benedikt M. Pötscher, University of Vienna; Hannes Leeb, Yale University

University of Vienna, Universitaetsstrasse 5, Dept. of Statistics, Vienna, A-1010 Austria

benedikt.poetscher@univie.ac.at

Key Words: *inference, model selection, model uncertainty, AIC, uniform consistency*

We consider the problem of estimating the distribution of a post-model-selection estimator. The notion of a post-model-selection estimator here refers to the combined procedure resulting from first selecting a model (e.g., by a model selection criterion like AIC or by a hypothesis-testing procedure) and second, estimating the parameters in the selected model (e.g., by least-squares or maximum likelihood), all based on the same dataset. We show that it is impossible to estimate this distribution with reasonable accuracy even asymptotically. In particular, we show that no estimator for this distribution can be uniformly consistent (not even locally). This follows as a corollary to (local) minimax lower bounds on the performance of estimators for this distribution. Similar impossibility results are also obtained for the distribution of linear functions (e.g., predictors) of the post-model-selection estimator.

176 Interdisciplinary Funding Opportunities at NSF ⌘

National Science Foundation

Tuesday, August 10, 8:30 am–10:20 am

Interdisciplinary Funding Opportunities at NSF

◆ Roger L. Berger, National Science Foundation; ◆ Shulamith Gross, National Science Foundation; ◆ Xuming He, National Science Foundation; ◆ Keith N. Crank, National Science Foundation; ◆ Cheryl Eavey, National Science Foundation

National Science Foundation, 4201 Wilson Blvd., Room 995, Arlington, VA 22230

ceavey@nsf.gov

Key Words: *NSF, interdisciplinary, opportunities*

We will describe funding opportunities in statistics and probability at the National Science Foundation with an emphasis on interdisciplinary funding opportunities. Many of the funds in budget increases for the Division of Mathematical Sciences at the National Science Foundation are to be used for interdisciplinary research with all the other divisions and directorates at NSF. These are called Mathematical Sciences Priority Area funds. Statistics and probability, as tools used by all other sciences, should be major recipients of these funds. Statistics and probability are not only unified disciplines but also unifying disciplines. To successfully compete for these funds statisticians and probabilists must be aware of these opportunities. The panel will describe these opportunities and give advice about successful proposal submission.

177 Causal Inference in Health Services Research: Taking Account of Hidden Bias

Section on Health Policy Statistics

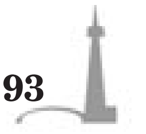
Tuesday, August 10, 8:30 am–10:20 am

Instrumental Variable Validation and Covariate Balance

◆ Thomas E. Love, Case Western Reserve University; Randall D. Cebul, Case Western Reserve University; Charles Thomas, Case Western Reserve University; Neal V. Dawson, Case Western Reserve University

Case Western Reserve University, Center for Health Care Res. and Policy, 2500 MetroHealth Dr., Cleveland, OH 44122

thomaslove@case.edu



Key Words: *observational studies, health services research, instrumental variables, hidden bias, selection bias, health policy*

This paper describes the importance of assumptions related to balancing covariates in instrumental variables analyses to infer causality in observational studies. In a previously published analysis of the effect of right heart catheterization (RHC) among seriously ill patients (SUPPORT study), RHC use significantly increased 30-day mortality rates. Bivariate comparisons suggest a candidate instrument: hospitals had different levels of RHC exposure but similar 30-day mortality rates. If hospital site is a valid instrument, stratification by site permits estimation of a local average treatment effect of RHC on mortality in marginal patients (who would get RHC only at high-RHC hospitals). While bivariate comparisons of the candidate instrument to exposure and outcome are one step in validating site as an instrument, we must also explore whether patients at the two sites differ in terms of observed (and unobserved) health characteristics. Researchers selecting instruments to estimate causal effects need to demonstrate balance in key covariates related to outcome. In particular, bivariate comparisons of instrument to exposure and outcome are insufficient to assess validity.

Use of Area-level Exposures as Instruments to Study the Effects of Treatment on AMI Mortality

◆ Therese Stukel, Institute for Clinical Evaluative Sciences; David Wennberg, Maine Medical Center; Lee Lucas, Maine Medical Center

Institute for Clinical Evaluative Sciences, G106 - 2075 Bayview Ave., Toronto, ON M4N 3M5 Canada

stukel@ices.on.ca

Key Words: *instrumental variable, observational study, selection bias, health services, propensity score*

We compared survival of heart attack patients exposed to different levels of medical and invasive management in their region of residence. Studies comparing these therapies need to account for selection bias as to who receives treatment. We undertook a national retrospective cohort study of 158,831 Medicare patients from the Cooperative Cardiovascular Project (CCP) database. Region of residence met the criteria for an instrumental variable (IV). We compared treatment effects using propensity score models (PSM) with patient-level exposures; instrumental variable (IV) techniques adapted to generalized linear models; and standard risk-adjustment using area-level exposures. PSM and adapted IV did not remove patient selection bias. Analysis using area-level exposures appeared to be equivalent to IV in the non-OLS setting, and to remove unmeasured confounding. Increased regional intensity of either treatment improved survival; no increased survival was seen with more invasive management beyond that seen with medical management. We interpret our findings in the context of observational studies with binary, censored outcomes in the presence of selection bias.

Propensity Score Matching with Time-dependent Covariates

◆ Bo Lu, Brown University

Brown University, 167 Angell St., CSS, Providence, RI 02912

bolu@stat.brown.edu

Key Words: *observational study, proportional hazard model, propensity score, matching*

In the observational study with a time-dependent treatment and time-dependent covariates, it is desirable to balance the distribution of the time-dependent covariates at every time point. A time-dependent propensity score based on Cox's proportional hazards model is proposed and used in risk-set-matching. Matching on this propensity score is shown to achieve a balanced distribution of the covariates in both treated and control groups. Optimal matching with various designs is conducted and compared in a study of a treatment, cystoscopy and hydrodistention, given in response to a chronic bladder disease, interstitial cystitis. Patients received this treatment at varied times, and the goal is to compare patients who were similar up to the time of treatment.

Causal Inference for Morbidity Outcomes in the Presence of Death

◆ Brian L. Egleston, Johns Hopkins University; Daniel O. Scharfstein, Johns Hopkins University; Ellen Freeman, Johns Hopkins University; Sheila West, Johns Hopkins University

Johns Hopkins University, Dept. of Biostatistics/Johns Hopkins Public Health, 615 N. Wolfe St., Room E3001/E3527, Baltimore, MD 21205-2179

beglesto@jhsph.edu

Key Words: *causal inference, competing risk, sensitivity analysis, visual impairment*

Evaluation of the causal effect of an exposure on a morbidity outcome is often complicated by the presence of death as a competing risk. In this setting, the causal effect is only well-defined for the principal stratum of subjects who would live whatever be the exposure. Motivated by aging researchers interested in understanding the causal effect of visual impairment on depression in a population with a high mortality rate, we introduce a set of scientifically driven assumptions to identify the causal effect among those who would live both with and without visual impairment. To evaluate the robustness of our analysis to nonidentifiable assumptions, we propose a method for performing a sensitivity analysis. We apply our method using the first three rounds of survey data from the Salisbury Eye Evaluation, a population-based cohort study of 2,520 older adults.

Unsupervised Propensity Scoring: NN and IV Plots

◆ Robert L. Obenchain, Eli Lilly and Company

Eli Lilly and Company, Corporate Center, Indianapolis, IN 46285

ochain@lilly.com

Key Words: *causal inference, instrumental variables, nonrandomized study, clustering, nearest neighbors, treatment effects*

The traditional propensity scoring (PS) approach to adjustment for treatment selection bias in nonrandomized studies is to fit a logistic regression model predictive of treatment choice and to then match or subclassify patients using either the resulting linear functional or actual (nonlinear) PS. While patients with identical PS predictions must lie on the same hyperplane (linear subspace) of covariate X-space, they are not necessarily close to each other within that hyperplane. In contrast, clustering of patients in X-space can assure this sort of closeness, at least when clusters are both compact (due to complete patient linkage) and numerous. Because clustering is an unsupervised method, any number of treatments and health outcomes can then be compared using a single, hierarchical set of clusters. Here, we stress use of nearest neighbor (NN) and instrumental variable (IV) plotting approaches because they not only reveal underlying assumptions about covariates but also depict the sensitivity of estimated treatment effects to choice of tuning parameters ...such as the number of clusters, choice of clustering algorithm, and patient dissimilarity metric.

178 Changing Methodology in Seasonal Adjustment ▲ ⌘

Business and Economics Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Changing to X-12-ARIMA at Statistics Canada

◆ Norma Chhab-Alperin, Statistics Canada; Benoit Quenneville, Statistics Canada

Statistics Canada, 120 Parkdale Ave., JT10 - Service Industries, Ottawa, ON K1A 0T6 Canada

norma.chhab-alperin@statcan.ca

Key Words: *X11-ARIMA, seasonal adjustment, official statistics*

At Statistics Canada, the X11-ARIMA method or its successors X11-ARIMA/88 and X11-ARIMA version 2000 have been the official seasonal adjustment methods since 1980. Developed at Statistics Canada, the X11-ARIMA improved over the widely used Census Bureau's X11 method by extending the series with ARIMA forecasts that reduce the revisions of the estimates at the end of the series. X11-ARIMA/88 added new features such as the estimation of the Easter effect, the removal of the calendar effects before the ARIMA modeling of the series and increased accuracy of the asymmetric filters. X11-ARIMA version 2000 was introduced to be Y2K compliant. X-12-ARIMA represents a technical advance of X11-ARIMA including all the capabilities of the latest version of X11-ARIMA and retaining the X11 decomposition as the core method of seasonal adjustment. The improvements of X-12-ARIMA over X11-ARIMA can be grouped in three broad categories: new modeling capabilities, new X11 adjustment options, new diagnostics, and a graphical tool implemented in SAS.

Experience of Bank of England in Switching to X-12-Arima for Seasonal Adjustment of Monetary Statistics

◆ Peter B. Kenny, PBK Research; Susan D. Docker, Bank of England

PBK Research, 53, Lindal Crescent, Enfield, EN2 7RP England

pkenny@globalnet.co.uk

Key Words: *seasonal adjustment, monetary and financial, change of method*

In January 2004, following an 18-month project, the Bank of England adopted X-12-ARIMA for the seasonal adjustment of the UK's monetary statistics, replacing an in-house method of seasonal adjustment (General Linear Abstraction of Seasonality, GLAS). X-12-ARIMA was chosen as the replacement for GLAS for a number of reasons: its technical refinement; its wide-ranging statistical diagnostics; its support and maintenance by the U.S. Census Bureau; and the fact that its future development seemed assured. This paper examines the technical issues that arose surrounding the use of X-12-ARIMA, as well as the practical aspects of changing the seasonal adjustment methodology applied to a multidimensional dataset. The points considered include: preserving accounting constraints among a set of series versus optimal adjustment of each series, direct versus indirect adjustment of aggregates (including the problem of residuals if direct is used), manual versus automatic treatment of outliers, prior adjustment for breaks using information about their effects on levels and flows versus the usual level shift treatment, adjustment of the same data at monthly and quarterly frequencies.

Variance Estimation for X11 via Model-based Spectral Approach under Additive and Multiplicative Decompositions

◆ Zhao-Guo Chen, Statistics Canada

Statistics Canada, 17-H, R.H. Coats Bldg., Ottawa, ON K1A 0T6 Canada

chenzha@statcan.ca

Key Words: *seasonal adjustment, component of time series, spectral density, autocovariance*

The X-11 seasonal adjustment procedures and its enhancements, X-11-ARIMA and X-12-ARIMA, have been widely used throughout the world. A long-standing problem is how to estimate the variances of the errors of the output components. Statisticians have proposed some solutions, but they are lack of reliability in real applications. Following Pfeffermann (1994), Pfeffermann and Scott (1997), and Pfeffermann, Morry, and Wong (1995), who provided with solutions via the moment approach, we proposed some procedures for estimating these variances via the model-based spectrum approach under the additive and the multiplicative decompositions. Many advantages of the spectrum approach are demonstrated by simulation and the procedures are tested by using some Labour Force Survey series and Wholesale Trade series.

Variance Measures for Seasonally Adjusted Employment and Employment Change

◆ Stuart Scott, Bureau of Labor Statistics; Danny Pfeffermann, Hebrew University and University of Southampton; Michael Sverchkov, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 1950, Washington, DC 20212

scott.stuart@bls.gov

Key Words: *labor force statistics, X-11, sampling error*

The Pfeffermann method of variance estimation for X-11 seasonal adjustment is applied to series from the U.S. Bureau of Labor Statistics Current Employment Statistics (CES) program and the Current Population Survey (CPS). The method builds from the linear approximation to X-11. Change across one or more months also has a linear approximation, so the method extends easily to a variance measure for seasonally adjusted change. Employment exhibits very high correlation across time in CES, a huge monthly establishment survey. Ratio estimates are made and tied to benchmark values, which become available with roughly a nine-month lag. Balanced repeated replication provides estimates of sampling error variances and autocovariances. The treatment of sampling error for this survey should have wide applicability, due to the similarity of its estimator to typical index series estimators. Results are also obtained for CPS household employment series. Sampling error variances are estimated from generalized variance functions and autocorrelations from balanced repeated replication estimates.

Analysis of Raking on Seasonally Adjusted Household Gross Flows Data

◆ Thomas Evans, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 4985, Washington, DC 20212

evans.thomas@bls.gov

Key Words: *time series, seasonal factors*

For many years, it has been well known that gross flows can provide useful information about month-to-month changes in labor force data from the Current Population Survey (CPS). The X-12-ARIMA program is used to seasonally adjust the flows as seasonal movements can confound the analysis of gross flows. However, after the flows are seasonally adjusted, they are raked to insure additivity to official CPS monthly seasonally adjusted employment and unemployment series. Because of the possibility that a raking procedure can introduce instability in the seasonally adjusted estimates, the final gross flows are reevaluated for residual seasonality and stable seasonal factors. Graphical and spectral tools are used in the analysis.

179 Models for Estimating the Size of Population of a Biological Community When Detection Varies among Individuals or Species ▲

Section on Bayesian Statistical Science

Tuesday, August 10, 8:30 am–10:20 am

Bayesian Models for Estimation of Population Size in Presence of Heterogeneity

◆ Sanjib Basu, Northern Illinois University

Northern Illinois University, Division of Statistics, De Kalb, IL 60115

basu@niu.edu

Key Words: *capture-recapture, closed population*

Capture-recapture methods are used to estimate the unknown size of a population. They are applied to estimate the sizes of biological populations, to estimate the number of undetected errors remaining in a software and to estimate undercount in census and epidemiological registries. I will consider Bayesian parametric and semiparametric models for estimation of population size in presence of heterogeneity in the population.

Hierarchical Bayes for Capture-recapture Data

◆ James S. Clark, Duke University

Duke University, Biology/Nicholas School of the Environment, Durham, NC 27708

jimclark@duke.edu

Key Words: *MCMC, demography, capture-recapture, stage structure*

Understanding population dynamics requires inference that admits the complexity of natural populations and the data ecologists obtain from them. Populations possess structure, which may be defined as "fixed" stages through which individuals pass, and superimposed variability among individuals and groups. Data contain missing values and inaccurate censuses. I extend the "missing value" framework for Bayesian analysis of structured populations to admit the heterogeneity in demography and in data that is typical of ecological populations. This hierarchical treatment of capture-recapture data allows inference on demographic rates and variability, together with simultaneous inference on population structure. Predictive distributions demonstrate profound impacts of population and data complexity on inference pertaining to life history schedules and growth. I demonstrate the method in simulation and apply it to field datasets from natural populations.

Estimating Size and Composition of Biological Communities by Modeling the Occurrence of Species

◆ Robert M. Dorazio, U.S. Geological Survey; Andrew Royle, U.S. Fish & Wildlife Service

U.S. Geological Survey, 7920 NW 71 St., Gainesville, FL 32653

bdorazio@usgs.gov

Key Words: *biodiversity, species richness, species composition*

We develop a model that uses repeated observations of a biological community to estimate the number and composition of species in the community. Estimators of community-level attributes are constructed from model-based estimators of occurrence of individual species that incorporate imperfect detection of individuals. Data from the North American Breeding Bird Survey are analyzed to illustrate the variety of ecologically-important quantities that are easily constructed and estimated using our model-based estimators of species occurrence. We compute site-specific estimates of species richness that honor classical notions of species-area relationships. We suggest extensions of our model to estimate maps of occurrence of individual species and to compute inferences related to the temporal and spatial dynamics of biological communities.

Hierarchical Spatial Modeling for Estimation of Population Size

◆ Jarrett J. Barber, Duke University; Alan E. Gelfand, Duke University

Duke University, Campus Box 90251, Institute of Statistics and Decision Sciences, Durham, NC 27705-0251

jarrett@stat.duke.edu

Key Words: *coregionalization, generalized linear model, hierarchical model, log-linear model, model-based geostatistics, multivariate spatial random effects*

Estimation of population size has traditionally been viewed from a finite population sampling perspective. Typically, the objective is to obtain an estimate of the total population count of individuals within some region. Often, some stratification scheme is used to estimate counts on subregions, whereby the total count is obtained by aggregation with weights proportional to the areas of the subregions. We offer an alternative to the finite population sampling approach for estimating population size. The method does not require that the subregions on which counts are available form a complete partition of the region of interest. In fact, we envision counts coming from areal units which are small relative to the entire study region and that the total area sampled is a very small proportion of the total study area. In extrapolating to the entire region, we might benefit from assuming there is spatial structure to the counts. We implement this by modeling the intensity surface as a realization from a spatially correlated random process. In the case of multiple population or species counts, we use the linear model of coregionalization to specify a multivariate process.

Nonidentifiability of Population Size from Capture-recapture Data with Heterogeneous Detection Probabilities

◆ William A. Link, U.S.G.S. Patuxent Wildlife Research Center

U.S.G.S. Patuxent Wildlife Research Center, 11510 American Holly Dr., Laurel, MD 20708

william_link@usgs.gov

Key Words: *mark-recapture, heterogeneity, detection, population, wildlife, ecology*

Heterogeneity in detection probabilities has long been recognized as problematic in mark-recapture studies, and numerous models developed to accommodate its effects. Individual heterogeneity is especially problematic, in that reasonable alternative models may predict essentially identical observations from populations of substantially different sizes. Thus, even with very large samples, the analyst will not be able to distinguish among reasonable models of heterogeneity, even though these yield quite distinct inferences about population size. I illustrate the problem using the simple closed population model M_h.

180 Ranked-set Sampling ⌘

General Methodology

Tuesday, August 10, 8:30 am–10:20 am

Unbalanced Ranked-set Sampling for Estimating a Population Proportion

◆ Haiying Chen, The Ohio State University; Elizabeth A. Stasny, The Ohio State University; Douglas Wolfe, The Ohio State University

The Ohio State University, 1958 Neil Ave., Columbus, OH 43210

chen@stat.ohio-state.edu

Key Words: *binary variable, Neyman allocation, perfect ranking, relative precision*

Application of ranked-set sampling (RSS) techniques to data from a dichotomous population has been an active research topic recently. It has been shown that balanced RSS leads to improvement in precision over simple random sampling (SRS) for estimation of a population proportion. Balanced RSS, however, is not optimal in terms of variance reduction for this setting. The objective of this paper is to investigate the application of unbalanced RSS to estimation of a population proportion under perfect ranking, where the probabilities of success for the order statistics are functions of the underlying population proportion. In particular, Neyman allocation, which assigns sample units for each order statistic proportionally to its standard deviation, is shown to be optimal in the sense that it leads to minimal variance within the class of RSS estimators that are unweighted averages of the order statistics. We also use a substantial dataset, the NHANES III data to demonstrate the feasibility and benefits of Neyman allocation in RSS for binary variables.

Generalized Linear Model Analysis for Order Restricted and Ranked-set Sample Designs

◆ Omer Ozturk, The Ohio State University; Steven MacEachern, The Ohio State University

The Ohio State University, Dept. of Statistics, 1958 Neil Ave., Columbus, OH 43210

omer@stat.ohio-state.edu

Key Words: *efficiency, noncentrality parameter, judgment ranking, power comparison*

We develop a randomization technique to assign treatment combinations to experimental units to improve the statistical inference for contrast parameters in a designed experiment. This technique uses subjective judgments about experimental units in a given set to impose restrictions on the randomization. Treatment combinations are assigned to judgment ranked units within each set. Such an assignment creates positive dependence among units within sets and so yields a reduction in variation for treatment contrasts. For the proposed design, we provide an analysis based on the generalized linear model. Tests and confidence intervals are developed for effects in the model. It is shown that the new design performs better than its competitors—ranked set sample and completely randomized designs. Finite sample results are discussed.

Judgment Stratification with Imprecise Rankings

◆ Elizabeth A. Stasny, The Ohio State University; Steven MacEachern, The Ohio State University; Douglas Wolfe, The Ohio State University

The Ohio State University, Dept. of Statistics, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210

eas@stat.ohio-state.edu

Key Words: *random allocation, ranked-set sampling, stratification, variance reduction*

Judgement post-stratification, which is based on ideas similar to those in ranked set sampling, relies on the ability of a ranker to forecast the ranks of potential observations on a set of units. In practice, the authors sometimes find it difficult to assign these ranks. This note shows how one can borrow techniques from the literature on finite population sampling to allow a probabilistic ranking of the units in a set, thus facilitating use of these sampling plans and improving estimation. The same techniques provide one approach to estimation with a judgement post-stratified sample with multiple rankers. The technique is illustrated on allometric data relating brain weight to body weight in different species of mammals, and on a study of student performance in graduate school.

Estimation of the Finite Population Mean by Judgment Post-stratification

◆ Xiaobai Li, The Ohio State University

The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210

li.478@osu.edu

Key Words: *ranked set sampling, stratification, post-stratification, judgment, finite population, variance reduction*

It is well known that stratified sampling plan and ranked-set sampling plan both give better estimator of population mean than simple random sample. We need to know the fraction of the population that falls into each stratum and specify the rules for stratification before the data is analyzed. And we need to make the judgment before the sample is taken or before quantification. Judgment post-stratification is a new method that keeps the benefits of post-stratification and ranked-set sampling while overcomes their shortcomings. It is based on the collection of a simple random sample of units to be measured, while creating the strata by post-stratification and comparing the units in the finite population. The comparison can be based on subjective judgment or objective concomitant variables. This method is most efficient when the ranking of the units in the population can be carried out relative easily and cheaply compared to the effort and expense required for measuring the units. Simulation results and real dataset analysis demonstrate that judgment post-stratification yield unbiased estimator of the population mean with smaller mean squared error.

Optimal Distribution-free Confidence Bands for a Continuous CDF

◆ Jesse Frey, The Ohio State University

The Ohio State University, Dept. of Statistics, 1958 Neil Ave., Columbus, OH 43210

frey@stat.ohio-state.edu

Key Words: *confidence bands, ranked-set sampling, Kolmogorov-Smirnov confidence bands*

Distribution-free confidence bands for a cumulative distribution function are typically found through inversion of a distribution-free hypothesis test. We propose an alternative strategy in which the upper and lower bounds of the confidence band are chosen to minimize a weighted width criterion. We derive necessary and sufficient conditions for a confidence band to be optimal with respect to such a criterion, and we demonstrate that optimal bands are unique in most cases of practical interest. We apply our alternative strategy to the construction of confidence bands both for the case of simple random sampling and for the case of ranked-set sampling. Our confidence bands in the ranked-set sampling case condition on the observed sequence of order statistics. For that case, the computation of coverage probabilities is made possible by a new algorithm which does not seem to have any competitors in the literature.

181 Setting Standards for Government Surveys

Section on Government Statistics, Social Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Updating Federal Standards for Statistical Surveys

◆ Alan R. Tupek, U.S. Census Bureau

U.S. Census Bureau, Rm 3705 FOB 3, Washington, DC 20233

alan.r.tupek@census.gov

Key Words: *statistical policy, fitness for use, quality standards*

Current standards for surveys conducted for and by federal government agencies were issued in 1978. Since that time, these standards have been used by the U.S. Office of Management and Budget in the clearance process for statistical data collections. In 2003, the Federal Committee on Statistical Methodology formed an interagency group to update these standards. The group's recommendations for new standards will be discussed in the other papers in this session. This paper will include a description of the current standards and the goals for the new standards. It will also describe the process of establishing and obtaining approval for the new standards, and the potential impact on the clearance process.

Proposed Standards for Planning, Designing, and Conducting Surveys

◆ Marilyn M. Seastrom, National Center for Education Statistics; Jennifer Madans, National Center for Health Statistics

National Center for Education Statistics, 1990 K St. NW, Room 9049, Washington, DC 20006

Marilyn.Seastrom@ed.gov

Key Words: *standards, statistical policy, survey planning, survey design, conducting surveys, documenting surveys*

In 2003 OMB created a subcommittee of the Federal Committee on Statistical Methodology (FCSM) to make recommendations for a major revision of OMB Directive 1: Standards for Federal Surveys, and Directive 2: Standards for Publication of Statistics. In response to this request, the subcommittee adopted the framework developed by federal statistical agencies in response to the OMB Information Quality Guidelines requirements of 2002, and used that framework to develop a set of recommended Standards for Federal Surveys. The draft document from the subcommittee includes some eight chapters. This paper will summarize the background and introduction and discuss the contents of the first three chapters of the draft document—Development of Concepts and Methods, Planning and Design, and Collection of Data. To that end, topics ranging from planning and pretesting to data collection methods and documentation will be covered, along with a consideration of the measurement of nonresponse, developing sampling frames, and maintaining data series over time.

Standards for Producing and Disseminating Survey Estimates in the U.S. Statistical System

◆ Stephen Cohen, Bureau of Labor Statistics; Kevin Cecco, Internal Revenue Service

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Suite 1950, Washington, DC 20212

Cohen.Steve@bls.gov

Key Words: *statistical survey standards, data analysis, data dissemination, data editing*

The United States statistical system is highly decentralized, consisting of more than seventy agencies and units that engage in some statistical activities. Statistical surveys are conducted for a variety of reasons including general-purpose statistics, program management and/or evaluation, research, and customer satisfaction measurement. The Office of Management and Budget (OMB) issues directives to ensure agencies uniformly apply sound statistical techniques and best practices when developing and overseeing government sponsored statistical surveys. As part of ongoing efforts to update and revise a number of statistical standards, the OMB formed an interagency group of members from eight agencies to develop recommendations for updating and revising standards for statistical surveys. This paper will discuss the recommendations for standards associated with processing, analyzing, producing, and disseminating survey estimates. Each standard has recommended guidelines of best practices that will be discussed.

What OMB Looks for When Agencies Request Approval to Conduct Surveys

◆ Brian A. Harris-Kojetin, U.S. Office of Management and Budget; Susan Schechter, Office of Management and Budget; Katherine K. Wallman, U.S. Office of Management and Budget

bharrisk@omb.eop.gov

Key Words: *best practices, government surveys, standards*

Federal agencies conduct and sponsor a wide variety of surveys to gather information from state, local, and tribal governments, businesses, and individuals. In addition, surveys frequently are used in program evaluations or research studies conducted or sponsored by federal agencies. Under the Paperwork Reduction Act (PRA), agencies must obtain approval from the U.S. Office of Management and Budget (OMB) prior to collecting any information from 10 or more respondents. To assist agencies in preparing these information collection requests (ICRs), OMB is developing a series of questions and answers that have commonly arisen in the OMB review of agency surveys. The Q&A's address issues about the OMB review process and timelines as well as methodological issues that frequently arise during OMB's review. The guidance is intended to help agencies better understand OMB's expectations for survey information collections and to assist them in identifying and documenting information to facilitate the review process. We provide an overview of the current draft guidance and discuss how it is intended to be used by agencies and by OMB desk officers.

Standards for Surveys at Statistics Canada

◆ Gordon Brackstone, Statistics Canada; Paul Johanis, Statistics Canada; John Kovar, Statistics Canada

Statistics Canada, 78 Charing Rd., Ottawa, ON K2G 4C9 Canada

bracgor@statcan.ca

Key Words: *surveys, standards*

In a statistical agency that carries out a wide variety of surveys on a broad range of social, economic, and environmental topics, standards serve two primary purposes. Firstly, they can define levels of practice considered appropriate for the various activities that make up the planning, design, implementation, and evaluation of a survey. Such standards serve to ensure that all survey activity sponsored by the agency meets or exceeds at least a minimum professional standard. Secondly, standards may identify particular content definitions or methods that must be used in a survey in the interests of ensuring coherence among the outputs of different survey programs across the agency. Though distinct, these two purposes are complementary, and some standards serve both purposes. This paper outlines the approach that Statistics Canada takes to the definition and application of survey standards, and describes the main elements of the standards it has adopted.

182 Statistical Applications in Medical Device Clinical Studies

Biopharmaceutical Section

Tuesday, August 10, 8:30 am–10:20 am

Bayesian Analysis of Mislabeling of Training Observations in Discriminant Analysis: Applications to Diagnostic Testing in the Medical Device Industry

◆ Laura A. Thompson, U.S. Food and Drug Administration

U.S. Food and Drug Administration, Center for Devices and Radiological Health, 1350 Piccard Dr., HFZ-542, Rockville, MD 20850

lxt@cdrh.fda.gov

Key Words: *classification, Bayesian, medical devices, mislabel*

The standard case of discriminant analysis constructs a classification rule using training observations that are assumed to be correctly labeled as belonging to their respective classes by some labeling mechanism. However, in practice, the labeling mechanism can be fallible, with an unknown probability of mislabeling training observations. The probability of mislabeling can depend on the class label (class-dependent mislabeling) or even on the value of the observation vector to be labeled (observation-dependent mislabeling). We use a Bayesian framework to develop models for both of these types of mislabeling for the k-class multivariate Gaussian case, and suggest alternatives for relabeling the mislabeled training observations using the posterior distribution of the class labels for the training observations. We also derive predictive classification probabilities for new observations. Our analyses are investigated on several simulated datasets and on example scenarios related to diagnostic testing in the medical device industry.

Bayesian Methods for Combining Medical Device Studies under FDA Review

◆ Gene A. Pennello, U.S. Food and Drug Administration

U.S. Food and Drug Administration, Division of Biostatistics, HFZ-542, Rockville, MD 20850

gxp@cdrh.fda.gov

Key Words: *hierarchical model, effective sample size, Dirichlet process prior, synthetic control, historical control, meta-analysis*

To support pre-market approval of a medical device under review by FDA, Bayesian methods are sometimes used that combine studies on the same or very similar devices. For example, Bayesian methods have been used to allow a study on the rate of target vessel failure for a new generation coronary stent to borrow strength from studies on previously approved, earlier generation stents. As another example, Bayesian methods have been used to summarize multiple studies of historical controls that are being compared with the investigational device. The typical approach has been to use a Bayesian hierarchical model to account for variation between studies. I will (1) review advantages, potential pitfalls, and useful summaries when using hierarchical models to combine medical device studies, (2) discuss various approaches to using hierarchical models to combine historical control studies, including the derivation of synthetic controls, and (3) compare hierarchical models with other Bayesian methods for combining studies (e.g,. using a Dirichlet process prior).

Practical Issues with the Application of Propensity Score Analysis to Nonrandomized Medical Device Clinical Studies

◆ Lilly Yue, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1350 Piccard Dr., HFZ-542, Rockville, MD 20850

lqy@cdrh.fda.gov

Key Words: *propensity score analysis, missing covariates, treatment group comparability, sample size estimation*

In nonrandomized medical device clinical studies, large differences often exist on observed baseline covariates between the treatment and control groups, which can result in biased treatment effect estimates. Propensity score analysis has been performed in some recent nonrandomized medical device submissions to control for multiple imbalanced baseline covariates in two treatment comparison. There are many practical issues encountered in the application of the methodology, such as handling missing covariates when estimating propensity scores, assessing the success of the propensity score estimation by checking the resulting balance of the distributions of covariates, evaluating treatment group comparability by the distributions of propensity scores, comparing treatment groups by adjusting for the propensity scores, and estimating sample size required for the treatment comparison incorporating the propensity scores. These issues will be discussed and illustrated from a regulatory statistical reviewer's perspective.

Survival Analysis in Medical Device Clinical Studies with Heterogeneous Samples

◆ Barbara Krasnicka, U.S. Food and Drug Administration; Chang S. Lao, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1350 Piccard Dr. HFZ-542, Rockville, MD 20850

bxk@cdrh.fda.gov

Key Words: *survival analysis, heterogeneous samples*

A number of methods of medical treatment effect evaluation, e.g., Kaplan-Meier product limit estimator, Cox proportional hazards regression model and the log rank test, have been well established in survival analysis. However, these methods have problems connected with the heterogeneity of individual patients in clinical studies, for instance, with misspecification of regression model, biases of treatment effect comparison, and loss of power. We intend to overview types of heterogeneity and to show how to handle such problems in practice. An additional difficult issue arises in survival analysis in the case of correlated data (e.g., multicenter studies, hip, dental, or ear devices). Some approaches, in particular frailty models, that help to overcome the correlated data problem will be discussed.

Clinical Trial Analyzed with Cox Proportional Hazards Models (Stratified over Covariables) for Progression-Free and Overall Survival of Patients with a Progressive Disease

◆ Harry F. Bushar, U.S. Food and Drug Administration

U.S. Food and Drug Administration, DHHS/PHS/FDA/CDRH/OSB/DBS/DX, 1350 Piccard Dr., Rockville, MD 20850

hfb@cdrh.fda.gov

Key Words: *Cox Proportional Hazards, progression-free survival, overall survival*

A multicenter, longitudinal clinical trial was conducted to determine whether trends in marker level correlate with Progression-Free Survival (PFS) or Overall Survival (OS). Patients with a specific progressive disease, who were about to start a new line of therapy were enrolled. Initially, three ad hoc methods for interpretation of longitudinal changes in marker level were applied, under the naïve assumption that time segments within a patient would be independent, and data were collapsed into 2x2 tables of marker level changes by progression. However, the independence assumption was difficult to justify. Therefore, following selection of the optimal marker cut-off level, three patient populations were defined and SAS PROC PHREG (Cox Proportional Hazards Regression) was used to predict PFS or OS from marker level, based on patients with available data, with PFS stratified over one covariable and with OS stratified over two covariables. The results of these analyses demonstrate that marker level successfully predicts both PFS and OS, with distinct results for each of the three patient populations, without requiring the naïve assumption that time segments within a patient would be independent.

183 Current Topics in RDD Surveys ◬ ⌘

Section on Survey Research Methods, Social Statistics Section, Section on Government Statistics

Tuesday, August 10, 8:30 am–10:20 am

Household Telephone Service and Usage Patterns in the United States in 2004: A Demographic Profile

◆ David Morganstein, Westat; Clyde Tucker, Bureau of Labor Statistics; J. Michael Brick, Westat; James Espisito, Bureau of Labor Statistics; Brian J. Meekins, Bureau of Labor Statistics

Westat, 1650 Research Blvd., Rockville, MD 20850

davidmorganstein@westat.com

Key Words: *cell phones, coverage bias*

Over the past decade, the U.S. telephone system has undergone rapid change, with changes in the size of the universe of numbers and how they are assigned. The rapid growth of cellular phone usage is another key development and is the subject of this paper. In February 2004, a supplement to the Current Population Survey on telephone service was conducted. Questions address the prevalence of different types of telephone service in households and how these services are used. This information is of critical importance to developing efficient telephone survey methodology in these changing times. This paper first discusses the questionnaire development process and the quality assessment activities associated with the supplement instrument. Estimates are then provided of the percentages of households with landline only, no landline but wireless service, both landline and wireless service, and no telephone service. The way households use these services is also examined. Demographic profiles of households with these characteristics are presented. This information can be used to assess the potential bias in estimates from surveys depending on the telephone sampling design used.

The Prevalence of Wireless Telephone Substitution

◆ Julian V. Luke, National Center for Health Statistics; Stephen J. Blumberg, National Center for Health Statistics; Marcie Cynamon, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 2214, Hyattsville, MD 20782

jluke@cdc.gov

Key Words: *survey response rates, cell phones, National Health Interview Survey*

The proliferation of cell phones within the United States has had a negative impact on response rates for random-digit-dial (RDD) household telephone surveys. Of particular concern are households that have “cut the cord,” substituting their residential landline with a wireless telephone. Current estimates of the prevalence of persons who have cut the cord may be obtained from the National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics of the Centers for Disease Control and

Prevention. The NHIS annually conducts in-person health-related interviews from a large sample of households representing the civilian noninstitutionalized household population of the United States. To better understand the implication of excluding wireless households on RDD telephone survey estimates, this presentation will examine the prevalence of wireless substitution. Household prevalence estimates will be presented by household size, family composition, income, home ownership, and region. Person-level prevalence estimates will be examined by age, sex, race/ethnicity, education, and employment status, as well as by household demographic categories.

Meeting the Challenge of the Telephone Consumer Protection Act and the FCC Telecommunications Act: Development of a Process to Identify Landline Telephone Numbers Ported to Cellular Phones

◆ Dale W. Kulp, Marketing Systems Group

Marketing Systems Group, 565 Virginia Dr., Fort Washington, PA 19034

dkulp@m-s-g.com

Key Words: *number portability, cellular, automated dialing prohibitions, TCPA*

The decision of the FCC to enforce a November 23, 2003, deadline that all telecommunications carriers in the TOP 100 MSAs must support the porting of landline to cellular service created a legal entanglement for many survey research organizations. Provisions in the TCPA make it illegal to call certain cellular numbers using any automated dialing device, which could be interpreted to include PC-based modem dialing, or even “speed dialing” on a basic phone. This has left researchers and many others in the position of either ignoring the current law, or dialing all sample telephone numbers manually. This paper details the history and background of the current laws and efforts to develop the first system for identifying telephone numbers ported to cellular service so that those numbers can be dialed manually in compliance with the law. Information on reliability testing, prevalence, and industry background will be presented.

A Dual Frame Sampling Design for an RDD Survey that Screens for a Rare Population

◆ Kadaba Srinath, Abt Associates Inc.; Michael P. Battaglia, Abt Associates Inc.; Meena Khare, National Center for Health Statistics

Abt Associates Inc., 4800 Montgomery Lane, Bethesda, MD 20814

kp_srinath@abtassoc.com

Key Words: *screening sample size, list frame, optimum allocation*

In a random-digit-dialing survey, the sample from a rare population is selected through screening a sample from the general population. Generally, a large screening sample size is required to identify an adequate sample from the rare population. The effort and cost of screening could be high if the required sample size from the rare population is large. Alternatively, if a partial list of members of the rare population is available, then a method of minimizing the screening cost is to use a dual frame sampling design with samples from both the RDD frame and the list frame. We investigate the use of a dual frame sampling design for the National Immunization Survey (NIS), in which the population of interest is children aged 19 to 35 months. We consider a Hartely-type ratio estimator for population proportions based on samples from both the frames. We determine the optimum allocation of the screening sample between the RDD frame and the list frame which minimizes the screening cost for a given sample size from the rare population. We compare the efficiency of the estimates from the RDD frame design and the dual frame desgin in the NIS.

Moving BRFSS from RDD to Multimode: A Web/Mail/Telephone Experiment

◆ Michael W. Link, Centers for Disease Control and Prevention; Ali Mokdad, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, Mailstop K-66, Atlanta, GA 30341-3717

MLink@cdc.gov

Key Words: *health survey, web survey, telephone survey, mail survey, response rates, mixed mode*

The precipitous decline in RDD response rates in recent years is forcing researchers to reconsider the design of RDD-based studies and surveillances. This presentation provides results from a set of experiments conducted in five states, administering the Behavioral Risk Factor Surveillance System (BRFSS) in three modes: web, mail, and CATI. Comparisons are made between the regular, monthly CATI-only results and self-administered web with CATI follow-up and self-administered mail with CATI follow-up alternatives. We examine differences in response rates, the types of sample members responding to each mode, and potential biases in the estimates obtained across modes.

184 Case Control Studies in Genetic Epidemiology ▲

Biometrics Section, Section on Statistics in Epidemiology

Tuesday, August 10, 8:30 am–10:20 am

Two-stage Designs for Population-based Association Studies

◆ Jaya M. Satagopan, Memorial Sloan-Kettering Cancer Center; E. S. Venkatraman, Memorial Sloan-Kettering Cancer Center; Colin B. Begg, Memorial Sloan-Kettering Cancer Center

Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology and Biostatistics, 307, East 63rd St., New York, NY 10021

satagopj@mskcc.org

Key Words: *order statistic, power, optimal design*

Gene-disease association studies based on case-control designs are used to identify candidate polymorphisms conferring disease risk. Here we propose a two-stage approach to identify disease susceptibility markers. In the first stage all markers are evaluated on a fraction of the available subjects. The most promising markers are further evaluated in Stage 2 on the remaining individuals. Using

simulations, we show that this approach provides a substantial reduction in the total number of marker evaluations for a minimal loss of power in comparison with a one-stage approach (where all the candidate markers are evaluated on all the individuals). As a general guideline, the simulations indicate that evaluating all the markers on 50% of the individuals in Stage 1 and evaluating the most promising 10% of the markers on the remaining individuals in Stage 2 provides near optimal power while resulting in a 45% decrease in the total number of marker evaluations.

Power of SNP-by-SNP vs. Haplotype-based Analysis of Case Control Association Studies

◆ Philip S. Rosenberg, National Cancer Institute; Anney Che, National Cancer Institute; Derek Gildea, National Cancer Institute; Monica Ter-Minassian, National Cancer Institute; Bingshu E. Chen, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd., EPS/7006, Biostatistics Branch, DCEG, Bethesda, MD 20852

rosenbep@mail.nih.gov

Key Words: *case control study, false discovery rate, genetic epidemiology, multiple comparisons procedures, single nucleotide polymorphisms*

The risk of developing a complex disease may depend on variants in several genes that each exerts a small effect. Case control studies can investigate SNPs or haplotypes in a candidate gene; we assess the power of SNP-by-SNP vs. haplotype analysis. We model scenarios where genotypes for all SNPs in a candidate gene, and haplotypes, are available. We incorporate empiric patterns of linkage disequilibrium of SNPs using a nine-gene panel: CASP8, CASP10, CFLAR, GAD2, H19, INS, SDF1, TCF8, and CTLA4. Disease susceptibility follows a prospective linear logistic model; we consider dominant, co-dominant, and recessive allelic effects. We study three procedures: SNP-by-SNP analysis with p values adjusted for multiple comparisons using Benjamini-Hochberg False Discovery Rate (FDR); Haplotype-based logistic regression; and Omnibus Permutation Test of SNP-by-SNP and haplotype analysis. When disease risk is conferred by a SNP, we find that SNP-by-SNP analysis with FDR is superior to haplotype analysis. Conversely, haplotype analysis can be superior when disease risk is conferred by a haplotype. The Omnibus Test appears to have excellent performance because it tracks the most powerful procedure.

Omnibus Tests in Case Control Association Studies of Candidate Genes

◆ Bingshu E. Chen, National Cancer Institute; Philip S. Rosenberg, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd. EPS/8047, Rockville, MD 20852

ChenEric@mail.nih.gov

Key Words: *case control study, genetic epidemiology, Omnibus Test, single nucleotide polymorphisms*

In case control studies, it is of interest to investigate the association between disease and candidate genes. In a candidate gene analysis, when disease susceptibility is conferred by a SNP, the SNP-by-SNP test with false discovery rate (FDR) controlling multiple comparison procedure often has higher power than the Haplotype Frequency Test (HFT). However, the HFT may have higher power when disease susceptibility is conferred by a haplotype. We develop an Omnibus Test to address this model selection problem. Let p1 be the minimum adjusted p value of SNP-by-SNP FDR, and let p2 be the p value of the HFT. The Omnibus Test is constructed from p3 = min(p1, p2). A challenge arises because the distribution of p3 is unknown. Two approaches are considered: Bonferroni correction and permutation test. Simulation studies show that both Omnibus Tests have reasonable Type I error; the permutation test is more powerful. The Omnibus Tests successfully address the underpower problem of each procedure when the true genetic model favors the other: the power obtained by the Omnibus Test is close to the more powerful of the component test statistics p1 and p2.

The Impact of Diagnosis Error on Testing Genetic Association in Case Control Studies

◆ Gang Zheng, National Heart, Lung, and Blood Institute; Xin Tian, National Heart, Lung, and Blood Institute

National Heart, Lung, and Blood Institute, Office of Biostatistics Research, DECA, 6701 Rockledge Dr., MSC 7938, Bethesda, MD 20892

zhengg@nhlbi.nih.gov

Key Words: *case control study, diagnosis error, power, sample size, sensitivity, specificity*

In case control studies, subjects in the case group may be recruited from suspected patients who are diagnosed positively with disease. While many statistical methods have been developed for measurement error or misclassification of exposure variables in epidemiological studies, no studies have been reported on the effect of errors in diagnosing disease on testing genetic association in case-control studies. We study the impact of using the original Cochran Armitage trend test assuming no diagnosis error when, in fact, cases and controls may be clinically diagnosed by an imperfect gold standard or a reference test. The Type I error, sample size, and asymptotic power of trend tests are examined under a family of genetic models in the presence of diagnosis error. The empirical powers of the trend tests are also compared by simulation studies under various genetic models.

185 Distance Education: How is it working?

Section on Statistical Education, Section on Quality and Productivity

Tuesday, August 10, 8:30 am–10:20 am

Distance Education: How is it Working?

◆ Joan Weinstein, Harvard University Extension School; ◆ Carl Lee, Central Michigan University; ◆ Robert Goldman, Simmons College; ◆ Kim Robinson, Clayton College and State University

Clayton College and State University, Morrow, GA 30260

kimrobinson@mail.clayton.edu

Key Words: *statistics education, distance education, assessment*

Many colleges and universities are jumping on the Distance Education bandwagon, using a variety of methods to “deliver” courses to the students. The presenters will each address a set of questions to describe how their distance statistics courses are structured, including a discussion of how the course is presented to the students, how assessment is managed, and how the instructor communicates individually with the students. In addition, each presenter will evaluate the experience, describing how students have responded, any unforeseen problems that were encountered, and any significant difference in the success of distance students vs. traditional students. Finally, each will draw conclusions about the success of the experience for them, and for their students.

186 Profiling ▲

Section on Statistics in Defense and National Security, Cmte on Statisticians in Defense and National Security

Tuesday, August 10, 8:30 am–10:20 am

Perspectives on Profiling

◆ Jeff Butler, Bureau of Transportation Statistics; ◆ Steven Smith, U.S. Dept. of Justice; muel E. Buttrey, Naval Postgraduate School

Naval Postgraduate School, Code OR/Sb, Dept. of Operations Research, Monterey, CA 93943

buttrey@nps.navy.mil

Key Words: *profiling, clustering, classification trees, fraud detection, data-mining*

“Profiling” individuals according to characteristic traits suggests infringement on civil liberties. This session will consider three different perspectives on profiling: Jeff Butler on transportation, Steven Smith on legal aspects: “Statistical and Legal Issues in Profiling and Related Activites.” This presentation will address legal and data collection issues and experiences related to generating estimates on racial profiling, primarily in law enforcement. Sam Buttrey on classification: Clustering techniques divide observations into groups but often depend on the scaling of the variables, are changed by monotonic transformations, and do not provide for selection of “important” variables. We fit a set of regression or c lassification trees with each variable acting in turn as the “response” variable. Trees with poor predictive power are discarded. Therefore, “noise” variables will often appear in none of the trees and have no effect on the clustering. The technique is unaffected by linear transformations of the continuous variables and is resistant to monotonic ones. Categorical variables are included automatically.

187 Multivariate Statistics

Section on Statistical Computing

Tuesday, August 10, 8:30 am–10:20 am

Bayesian Multivariate Item Response Theory Model and Its Software BMIRT

◆ Lihua Yao, CTB McGraw-Hill

CTB McGraw-Hill, 20 Ryan Ranch Rd., Monterey, CA 93940

lyao@ctb.com

Key Words: *item response theory, Bayesian, MCMC, multidimension, multigroup*

We propose a multivariate item response theory model for educational testing data (MIRT). This model is a generalization of the standard item response theory model. We use a Bayesian approach and MCMC for parameter estimation of this model. The estimation procedure is implemented in a computer program called BMIRT (Bayesian Multivariate Item Response Theory model). BMIRT has several important properties. First, the guessing parameters are estimated based on the data, not fixed in an ad hoc manner prior to fitting the model as in some existing software such as TESTFACT, and MICROFACT. Second, by combining different content areas for concurrent calibration, BMIRT provides more accurate objective scores. Third, BMIRT can determine the dimension loadings of the test items that are in different categories but are correlated, and as a result, more accurate parameter estimation and score reporting. Finally, with BMIRT, multigroup tests can be put on the same scale so that guidance and methods to measure longitudinal growth at the student level can be provided. We use real and simulated data sets to illustrate the applications of the model and BMIRT.

A Significance Test for Intervening Variable Effects

◆ Wei Pan, University of Cincinnati

University of Cincinnati, PO Box 210002, Cincinnati, OH 45221-0002

wei.pan@uc.edu

Key Words: *structural equation modeling, indirect effects, mediated effects, intervening variables, path analysis*

The literature has shown enormous interest in intervening variable effects which are often termed indirect effects or mediated effects. However, the distribution of the intervening variable effects had been ignored and these effects were estimated without standard errors until Sobel (1982) derived an asymptotic distribution of the indirect effects; after that, the Sobel’s test is commonplace in structural equation software and several modified Sobel’s tests emerged. Unfortunately, the Sobel’s and modified test techniques relied exclusively on large-sample approximations, which has been cause for some concern. The present study provides an alternative approach for establishing a significance test for the intervening variable effects. In this approach, the distribution of the intervening variable effects is more accurately derived through the application of Pan and Frank’s (2004) distribution of the

product of two dependent correlations. A simulation study is conducted to compare this alternative approach with the Sobel's and modified tests. This approach enables researchers to test the intervening variable effects with more confidence.

Estimating Intrinsic Dimensionality from the Nearest-neighbor Distances

◆ Elizaveta Levina, University of Michigan

University of Michigan, Dept. of Statistics, 439 West Hall, 550 E. University, Ann Arbor, MI 48109-1092

elevina@umich.edu

Key Words: *dimensionality reduction, intrinsic dimension, manifold projections, nearest-neighbor distances*

Recently developed nonlinear manifold projection methods such as Locally Linear Embedding and Isomap have become popular dimensionality reduction tools. Using these methods successfully for applications such as classification requires estimating intrinsic data dimension. The existing methods for intrinsic dimension estimation are mostly heuristic and very little is known about how they scale with the number of observations and the true dimension. We propose a new method for estimating intrinsic dimensionality, derived through a rigorous analysis of distributions of the nearest-neighbor distances. A simple closed-form maximum likelihood estimator of dimension is obtained by treating observations in a small sphere as a homogeneous Poisson process. We apply the method to some of the popular manifold datasets and study its behavior as the dimension and the sample size grow, analytically and through simulations. This is joint work with Peter Bickel.

Multivariate Rankings of Outlier Groupings Using Minimal Spanning Trees and Convex Hull-peeling

◆ Mark W. Lukens, George Mason University; James E. Gentle, George Mason University

George Mason University, 4061 Cardinal Crest Dr., Woodbridge, VA 22193

mlukens@gmu.edu

Key Words: *outliers, multivariate ranking, convex hull, minimal spanning trees*

Order statistics and the ranking of data are important techniques in data analysis. Ranking is unambiguous in one dimension where a simple sorting from lowest to highest will order the data. In the multivariate case, however, this natural concept of ordering is not so clearly defined. Ordering multivariate data and the concept of multivariate rankings become complicated in higher dimensions, and do not necessarily correspond to other characteristics of data, such as clusters. This paper will examine two techniques for ranking multivariate data, minimal spanning trees and convex hull peeling, in the context of data groupings. Both of these methods yield potentially undesirable results when data clusters are present. The rankings tend to bounce back and forth between the various groupings. It desirable to have similar rankings for groups of data that are otherwise similar. This grouping of the ranks could then assist in analysis of the data. This paper introduces changes to the multivariate ranking algorithms for both the minimal spanning tree and the convex hull-peeling so the multivariate rankings are grouped analogous to the data.

Path Analysis of the Visual Attention Network for fMRI Data

◆ Jieun Kim, SUNY, Stony Brook

jekim@ams.sunysb.edu

Key Words: *Dynamic Causal Modeling (DCM), fMRI, structural equation modeling (SEM), Principal Components Analysis (PCA), AR(p) model*

The ultimate goal for brain connectivity study is to propose, test, modify, and compare certain directional brain pathways. Path analysis or structural equation modeling is ideal for such studies. We report the path analysis of the visual attention network using fMRI data of 28 subjects (14 male,14 female). The path analysis was done in three approaches: (1) The PCA was performed to reduce the multisubject multivariate time series data to the group-level eigen-image, and path analysis was performed on this assuming independent observations. (2) Average the multivariate time series data across subjects and perform path analysis in two stages— a suitable AR(p) model is determined for each relevant brain region, and the model goodness of fit from path analysis is examined. This approach tend to produce a similar pathway as approach (1). However, the model goodness of fit would increase because the AR(p) procedure reflects the time dependent nature of the data. (3) Finally, the original multisubject, multivariate time series data was used to fit the visual attention pathway. We used Dynamic Causal Modeling (DCM) analysis for this approach, and the comparison with SEM analysis was performed.

Searching Across Markov Equivalence Classes of Maximal Ancestral Graphs

◆ Rebecca A. Ali, National University of Singapore; Thomas Richardson, University of Washington

National University of Singapore, Dept. of Statistics & Applied Probability, Block S16, Level 7, 6 Science Dr. 2, 117546 Singapore

staara@nus.edu.sg

Key Words: *maximal ancestral graphs, joined graphs, Markov equivalence, DAG models, latent variables*

Graphical Markov models have emerged as a powerful tool for representing diverse data-generating processes. Ancestral graphs enable one to encode the conditional independence relations that arise in directed acyclic graph (DAG) models with latent variables, without explicitly including latent variables in the model (graph). However, for any ancestral graph, there may be several other graphs to which it is Markov equivalent thus making an exhaustive search across ancestral graphs inefficient. We present a joint operation on ancestral graphs which provides a simple representation for their Markov equivalence classes, thereby facilitating the model search process for some given data. We also provide a proof of the global Markov property for joined ancestral graphs. Model searches across equivalence classes of ancestral graphs can be conducted in statistical packages such as gR (for Gaussian models). The ultimate goal of this work is to obtain a full characterization of the structure of Markov equivalence classes for maximal ancestral graphs, thereby extending analogous results for DAGs given by Verma and Pearl (1991), Chickering (1995), and Andersson et al. (1997).

An Expanded Pre-processing Step to Latent Semantic Indexing

◆ Igor Perisic, Entopia, Inc.

Entopia, Inc., 3200 Bridge Parkway, Suite 101, Redwood City, CA 94065

iperisic@entopia.com

Key Words: *text mining, neural networks, Singular Value Decomposition, computational statistics*

Singular Value Decomposition (SVD) has often been used to benefit search engines retrieval performance. Its integration and performance at the core of Latent Semantic Indexing (LSI) has been documented. What we propose is to supplement the traditional pre-processing stages and usual associated weighting schemes with a more complex engine. An engine based on covariances across terms and an objective optimization function. We show early results of this two step approach on standard datasets with regards to retrieval and compression.

188 Kernel Smoothing and Functional Data Analysis

Section on Nonparametric Statistics

Tuesday, August 10, 8:30 am–10:20 am

A Practical Description of Density Estimation Based on Kernel Methods

◆ Simon J. Sheather, University of New South Wales

University of New South Wales, AGSM, UNSW, Sydney, 2052 Australia

simonsh@agsm.edu.au

Key Words: *density estimation, bandwidth, data-sharpening*

This talk will provide a practical description of density estimation based on kernel methods. An important aim of the talk is to encourage practicing statisticians to apply these methods to data. As such reference is made to implementations of these methods in S-PLUS and SAS. Topics covered include bandwidth selection, local likelihood density estimation, and data-sharpening. In addition, we shall compare the performance of some of the methods that have been discussed using a new example involving data from the U.S. PGA tour.

On Geometric Conditional Quantiles

◆ Jan G. De Gooijer, University of Amsterdam

University of Amsterdam, Roetersstraat 11, Amsterdam, 1018 WB The Netherlands

j.g.degooijer@uva.nl

Key Words: *asymptotic normality, Bahadur representation, confidence ellipsoids, conditional geometric quantile, kernel function*

Motivated by Chaudhuri's work (1996) on unconditional geometric quantiles, we explore the asymptotic properties of sample conditional geometric quantiles, defined through kernel functions, in high dimensional spaces. First, we establish a Bahadur-type linear representation for the conditional quantile estimator. Also we obtain the convergence rate of the remainder term in this representation. Using these results, we prove that the estimated conditional geometric quantile is asymptotically normally distributed. Based on these results we propose confidence ellipsoids for multivariate conditional quantiles. These ellipsoids are shown to have asymptotically good coverage properties and to behave well in finite-sample situations. The methodology is illustrated via data analysis and a Monte Carlo study.

Adaptive Multivariate Kernel Mixture-model Cluster Analysis for Mixed Data Using Information Complexity

◆ Halima Bensmail, University of Tennessee; Hamparsum Bozdogan, University of Tennessee

University of Tennessee, 334 Stokely Management Bldg., Statistics Dept., College of Business Administration, Knoxville, TN 37996-0532

bensmail@utk.edu

Key Words: *Gaussian kernel distribution, Fisher information, cluster analysis, mixture models*

A nonparametric approach is developed herein to perform a cluster analysis for mixed data. The mixture-model cluster analysis is based on mixture of kernel density function. For robustness, a variety of covariance matrix estimators are proposed to overcome the singularity and limitation of the maximum likelihood estimator. While traditional criteria such as BIC and AIC overfit or underfit the number of clusters selection, we propose a complexity-based criteria ICOMP from the information theory as an alternative criteria to calibrate the penalty term. Applications to simulated and real data are well investigated and results are discussed.

A Smoothing-spline Approach to Linear Regression for Functional Data

◆ Yehua Li, Texas A&M University

Texas A&M University, Dept. of Statistics, 3143 TAMU, College Station, TX 77840

yehuali@stat.tamu.edu

Key Words: *functional linear model, reproducing kernel Hilbert space, smoothing spline*

Consider the linear regression model where the dependent variable is a scalar and the independent variable is a function observed on a finite interval. Ramsay and Silverman (1997) proposed a functional linear model, in which the mean of the response equals to an intercept plus the integral of the independent variable with a weight function \β(t). To estimate the weight function, a finite dimension subspace approximation using a given set of basis functions was proposed. As an alternative, we show this model could be adopted in a Reproducing Kernel Hilbert Space framework and a smoothing-spline estimator for the weight function is developed. Methods of selecting the smoothing parameter are discussed. Some asymptotic property is explored and serum lipoprotein data are used to illustrate the approach. Also, some simulations are conducted to make a comparison with the basis function approach.

Functional Semiparametric Quasi-likelihood Generalized Linear Models, When Time is Random

◆ Latonya Garner, University of Mississippi

University of Mississippi, PO Box 3958, University, MS 38677

lcgarne1@olemiss.edu

Key Words: *functional, semiparametric, quasi-likelihood, generalized linear models*

We will discuss the estimators of the parameters in the functional semiparametric quasi-likelihood generalized linear models. The estimators are given as the solutions of two estimating equations. The estimators are efficient if the quasi-likelihood function is a true likelihood when the time is random. The asymptotic distribution of the estimators is given. Comparison of efficiency is given through examples in an effort to investigate different models, in particular, when time is random and nonrandom.

Diffusion Smoothing on Manifold Data

◆ Moo K. Chung, University of Wisconsin, Madison

University of Wisconsin, Madison, Dept. of Statistics, 1210 West Dayton St., Madison, WI 53706

mchung@stat.wisc.edu

Key Words: *Gaussian kernel, Nadaraya-Watson estimator, Laplacian, manifolds, diffusion, kernel smoothing*

The Nadaraya-Watson (NW) kernel estimator has been widely used in nonparametric regression and image analysis, in part, due to its simplicity in the numerical implementation. The NW kernel is isotropic in Euclidian space so it assigns the same weights to observations equal distance apart. However, when we smooth data residing on a curved surface, it fails to be isotropic. On the curved surface, a straight line between two points is not the shortest distance so one may assign smaller weights to closer observations. For this reason smoothing data residing on a curved surface requires constructing a kernel that is isotropic along the geodesic curves. The direct analytical construction of such isotropic kernel in manifolds is nontrivial. An alternate approach is to find the kernel of an isotropic heat equation in manifolds. Such heat kernel is locally isotropic in the conformal coordinates and is a generalization of the NW kernel. We construct the diffusion estimator based on the heat kernel and present its properties in connection with the NW kernel estimator. As an illustration, we show the nonparametric diffusion estimation of the cortical thickness in a group of autistic children.

189 Reliability

Section on Physical and Engineering Sciences

Tuesday, August 10, 8:30 am–10:20 am

Predicting Neutron-induced Soft Errors

◆ Sarah E. Michalak, Los Alamos National Laboratory

Los Alamos National Laboratory, PO Box 1663, MS F600, Los Alamos, NM 87545

michalak@lanl.gov

Key Words: *inverse problem, soft error, supercomputer*

A "soft error" is a nondestructive error that causes a change in the state of an electronic component. Bit flips in computer memory are a frequent example. Such bit flips may be caused by neutrons that are produced by cosmic rays hitting nuclei in the atmosphere. Neutron-induced soft errors are a major concern for systems that contain many susceptible parts (such as supercomputers) and for devices housed at very high altitudes where the neutron flux is quite high (such as satellites and airplanes). In such systems, soft errors can result in system crashes or incorrect information being passed to a program. Estimating a device's susceptibility to neutron-induced soft errors is crucial for assessing its feasibility and reliability in particular environments. This talk describes current methods for using the Los Alamos Neutron Science Center to estimate the impact of soft errors in the deployed setting and suggests a framework for providing more accurate predictions of the expected number of neutron-induced soft errors and an associated standard error. The framework involves a carefully designed set of experiments and the solution of an inverse problem.

Prediction Bounds for the Number of Successes in a Sequence of Markov-dependent Bernoulli Trials

◆ Florian Rueck, University of Missouri, Rolla

University of Missouri, Rolla, 1805 Ashwood Dr. Apt. D, Rolla, MO 65401

fsrq54@umr.edu

Key Words: *Bernoulli trials, Markov-dependent, prediction intervals, maximum likelihood predictive density*

A generalization of a prediction interval procedure for the binomial distribution to the case of the binomial distribution with Markov-dependent trials is considered. Several different methods have been developed for the binomial distribution but only a method proposed by Lejeune and Faulkenberry is applied to the case of the binomial distribution with dependent trials. Given n observations from a binomial distribution with Markov-dependent trials, it is of interest to find an interval estimate such that the number of successes in a future set of trials is contained in the interval with a certain level of confidence. A simulation study was considered to investigate the coverage probabilities of the prediction bounds, which are based on the maximum likelihood predictive density. The use of estimators due to Klotz, Price, Kim, and Bai in place of the MLE, is considered because the MLEs cannot be expressed in a closed form.

Estimation for a Three-parameter Weibull Distribution with an Intuitive Parameterization

◆ Yi Han, Auburn University; Mark Carpenter, Auburn University; Saeed Maghsoodloo, Auburn University

Auburn University, 902 Linden Ave. Apt. 6, Dept. of Discrete and Statistical Sciences, Auburn, AL 36832

hanyi01@auburn.edu

Key Words: *Weibull distribution, location parameters, characteristic life, maximum likelihood, censoring, bootstrapping*

The literature is replete with theoretical and applied research on the Weibull distribution. As a result, this distribution has been proposed in a plethora of forms and parameterizations. We examine a three-parameter Weibull distribution with a shape parameter, a threshold/location parameter, and an uncommon scale parameterization. This parameterization yields a more intuitive model for the practitioner and may be more realistic in some reliability and survival applications. In application, the threshold parameter represents the minimum guaranteed lifetime, while our scale parameterization represents the characteristic life minus the threshold parameter. Estimators, including maximum likelihood estimators, are developed under complete and censored sampling assumptions. Asymptotic confidence intervals and bootstrapping algorithms are also proposed. Examples in engineering and medical research are included.

Sample Size Determination in Comparing Mechanical Reliabilities Based on Weibull Model

◆ Mark C.K. Yang, University of Florida

University of Florida, Dept. of Statistics, Gainesville, FL 32611

yang@stat.ufl.edu

Key Words: *Weibull modules, probability of failure, confidence interval, sample size*

The main purpose of this paper is to determine the sample size required to test whether new materials have reach a given reliability requirement under a specific stress level. Weibull model is assumed for the data variation. Since sample size determination relies on the method of data analysis, the best method to make conclusions from data is also surveyed. One sample and two sample data are considered. Although the results are obtained from simulation, the estimators are constructed so that they are almost independent of the parameters in a Weibull distribution.

Statistical Inference for the Modulated Power-law Process

◆ Ananda Sen, University of Michigan; Nibedita Bandyopadhyay, Oakland University

University of Michigan, 3550 Rackham, 915 E Washington, Ann Arbor, MI 48109

anandas@umich.edu

Key Words: *repairable systems, power-law process, maximum likelihood, asymptotics*

Failure data pertaining to a repairable system is commonly modeled by a nonhomogeneous Poisson process (NHPP). A modulated gamma process evolves as a generalization to a NHPP, where the observed failure epochs correspond to every successive kth event of the underlying Poisson process, k being an unknown parameter to be estimated from the data. We focus on a special class of modulated gamma process, called a modulated power-law process (MPLP) that assumes the Weibull form of the intensity function. The traditional power-law process is a popular stochastic formulation of certain empirical relationships between the time to failure and the cumulative number of failures, often observed in industrial experiments. The MPLP retains this underlying physical basis and provides a more flexible modeling environment potentially leading to a better fit to the failure data at hand. We investigate inference issues related to a MPLP. The maximum likelihood estimators (MLE's) of the model parameters are not in closed form and enjoy the curious property that they are asymptotically normal with a singular variance-covariance matrix.

Reliability Estimation of an Unknown Load Share Rule

◆ Paul H. Kvam, Georgia Institute of Technology; Hyoung T. Kim, Georgia Institute of Technology

Georgia Institute of Technology, 765 Ferst Dr., Industrial & Systems Engineering, Atlanta, GA 30332-0205

pkvam@isye.gatech.edu

Key Words: *maximum likelihood, software reliability, order-restricted inference, system dependence*

We consider a multicomponent load-sharing system in which the failure rate of a given component depends on the set of working components at any given time. Such systems can arise in software reliability models and in multivariate failure-time models in biostatistics, for example. A load-share rule dictates how stress or load is redistributed to the surviving components after a component fails within the system. We assume the load share rule is unknown and derive methods for statistical inference on load-share parameters based on maximum likelihood. Components with (individual) constant failure rates are observed in two environments: (1) the system load is distributed evenly among the working components, and (2) we assume only the load for each working component increases when other components in the system fail. Tests for these special load-share models are investigated.

190 Model Selection and Convolution Methods ⌘

General Methodology

Tuesday, August 10, 8:30 am–10:20 am

Computation Aspects on the Nonlinear Designs

◆ Christos Kitsos, Technological Educational Institute of Athens

Technological Educational Institute of Athens, Dept. of Mathematics, Palikaridi and Ag. Spyridonos Str, Egaleo, 12210 Athens, Greece

xkitsos@teiath.gr

Key Words: *Beale's measure, confidence region, calibration, curvature*

When fitted a nonlinear model, this is selected among various nonlinear models. One of the targets of this paper is to discuss the computation criteria, applied for the selection of the appropriate nonlinear models and the optimal design linked with the proposed model. An optimal design is the one that eventually concludes to the optimal estimation of the parameters. Various optimality criteria exist, the most well-known being the D-optimality: the one that minimizes the variances of all the included parameters of the assumed nonlinear model under investigation. Computationally the nonlinear problem is approached by a first order Taylor approximation, as a second order Taylor approximation might create more problems than those which try to solve. The nonlinearity influences the confidence regions and therefore Beale's measure of nonlinearity is adopted. The second target of this paper is to provide an easy approximation of this measure, useful for the experimenter. The method is applied to the calibration problem, which is faced as an optimal experimental design.

Controlling Variable Selection By Addition of Pseudo-Variables

◆ Yujun Wu, North Carolina State University; Dennis D. Boos, North Carolina State University; Leonard A. Stefanski, North Carolina State University

North Carolina State University, Dept. of Statistics, 3109 A Kings Ct., Raleigh, NC 27606

ywu5@unity.ncsu.edu

Key Words: *variable selection, dataset selection, best subsets, pseudo-nariables, forward selection, model error*

Many variable selection procedures have been developed in the literature for linear regression models. We propose a general approach by adding pseudo-variables to the dataset with the advantage of being applicable to a broader class of regression models; for example, binary regression, Poisson regression, etc. By controlling the proportion of pseudo-variables falsely selected in the model, we are able to tune variable selection procedures resulting in a good bias-variance trade-off. We focus on forward selection because it is applicable in the case where there are more variables than observations. Due to the difficulty of obtaining analytical results, we study our approach by Monte Carlo and compare it with a variety of commonly used procedures. The new method is illustrated on some real datasets.

Model Selection in a Regression Setting with Missing Covariate Data

◆ Niel Hens, Limburgs Universitair Centrum; Marc Aerts, Limburgs Universitair Centrum; Geert Molenberghs, Limburgs Universitair Centrum

Limburgs Universitair Centrum, Center for Statistics, Universitaire Campus, Diepenbeek, 3590 Belgium

niel.hens@luc.ac.be

Key Words: *Akaiki Information Criterion, missing data, model selection, weights*

In a regression setting, the Akaiki Information Criterion, AIC, is one of the most frequently used methods to select an optimal regression model. In case the missingness probability of the partially missing covariates depends on the response, regression estimates based on the complete cases are known to be biased. The use of selection criteria in the presence of missing covariate data can lead to poor and wrong models. We introduce a weighted version of AIC in analogy with the weighted Horvitz-Thompson estimates. The weights are thus proportional to the inverse of the missingness probabilities. In some settings these probabilities are known, in other settings we propose to use semiparametric estimates. Next to theoretical properties, an extensive simulation study and several data examples show that the weighted AIC criterion provides better model choices. This modification can be seen as an implicit imputation of missing observations. The extension towards the Bayesian Information Criterion, Mallows' Cp and other criteria is straightforward.

Exploring Intrinsic Dependence with CID

Tailen Hsing, Texas A&M University; ◆ Li-yu D. Liu, Texas A&M University; Marcel Brun, Texas A&M University; Edward R. Dougherty, Texas A&M University

Texas A&M University, 3143 TAMU, College Station, TX 77843-3143

daisy@stat.tamu.edu

Key Words: *dependence, prediction, classification*

We introduce a new measure of dependence, the coefficient of intrinsic dependence or CID, between two sets of variables. CID is model-free, invariant under monotone transformations of the variables, and is capable of fully differentiate different degrees of dependence. The estimation of CID can be carried out nonparametrically based on the ranks of the data, and the MSE of the estimator can be seen to decrease quickly with sample size. The applications that we discuss include testing for dependence, variable selection, parameter estimation in generalized linear models, and genetics data analysis.

Model Combining in Factorial Data Analysis

◆ Lihua Chen, Iowa State University; Yuhong Yang, Iowa State University

Iowa State University, 15 D Schilletter Village, Ames, IA 50010

lchen@iastate.edu

Key Words: *ANOVA, model selection instability, model combining*

Evaluating factor effects in the ANOVA framework by model selection based on hypothesis testing or other model selection criteria such as AIC or BIC often leads to unstable estimotors that have correspondingly poor performance. Measures of model selection instability are studied. As an alternative to model selection, we investigate an algorithm, ARM, which convexly combines the candidate models to estimate the cell means. Simulation and data examples demonstrate the advantage of combining over selection in many cases. A theoretical risk bound on the combined estimator is also obtained.

Bayesian Inference on Umbrella Orderings

◆ Chris Hans, Duke University; David B. Dunson, National Institute of Environmental Health Sciences

Duke University, ISDS, Box 90251, Durham, NC 27708-0251

hans@stat.duke.edu

Key Words: *order constraint, mixture prior, variable selection, downturn order, shape restriction, probit model*

In regression applications with categorical predictors, interest often focuses on comparing the null hypothesis of homogeneity to an ordered alternative. We propose a Bayesian approach for addressing this problem in the setting of normal linear and probit regression models. The regression coefficients are assigned a conditionally conjugate prior density consisting of mixtures of point masses at zero and truncated normal densities, with a (possibly unknown) changepoint parameter included to accommodate umbrella ordering. The investigator specifies the prior probability assigned to the global null hypothesis and to subhypotheses of no difference in specific groups, with default values suggested to account for multiple comparisons. A single Gibbs sampling chain can be used to obtain posterior probabilities for the different hypotheses and for estimation of regression coefficients and predictive quantities, either by model averaging or under the preferred hypothesis. The methods are applied to data from a carcinogenesis study.

Regression-assisted Deconvolution

◆ Jule P. McIntyre, Carnegie Mellon University; Leonard A. Stefanski, North Carolina State University

Carnegie Mellon University, 132 Baker Hall, 5000 Forbes Ave., Pittsburgh, PA 15213-3890

julie@stat.cmu.edu

Key Words: *measurement error, bandwidth, kernel density estimate, mean-variance function model*

We present a semiparametric deconvolution estimator for the density function of a random variable X that is measured with error. Traditional deconvolution estimators rely only on assumptions about the distribution of X and the error in its measurement, and ignore information available in auxiliary variables. Our method assumes the availability of a covariate vector statistically related to X by a mean-variance function regression model, where regression errors are normally distributed and independent of the measurement errors. Under common parametric assumptions on the conditional mean and variance, the estimator is root-n consistent for the true density of X, a substantial improvement over the logarithmic rates achieved by nonparametric deconvolution estimators. Simulations suggest that the estimator achieves a much lower integrated squared error than the observed-data kernel density estimator when models are correctly specified and the assumption of normal regression errors is met. We illustrate the method using anthropometric measurements of newborns to estimate the density function of newborn length.

191 Issues and Methods of Program and Course Development in Teaching Statistics for the Health Sciences ◬

Section on Teaching Statistics in the Health Sciences

Tuesday, August 10, 8:30 am–10:20 am

Meeting Challenges of Statistical Education in Graduate Dental Training

◆ Deborah V. Dawson, University of Iowa College of Dentistry

University of Iowa College of Dentistry, Dows Institute for Dental Research, 439E Dental Science Bldg. North, Iowa City, IA 52242-1010

deborah-dawson@uiowa.edu

Key Words: *statistical education, dental education, graduate education, curriculum, dentistry*

Statistical instruction of post-DDS graduate students is complicated by several constraints inherent in dental graduate education programs. These include the limited time in the curriculum available for statistical instruction, and the varying needs of multiple graduate programs. Trainees with minimal statistical training must receive some exposure to fairly complex methodologies if they are to become discriminating readers of the dental literature. Another complicating factor is the ubiquity of repeated measures in dental research. Three curricular approaches to these challenges will be discussed: (1) a first-semester introductory course which emphasizes applications and combines practical examples with conceptual approaches for more advanced topics, (2) a second-semester course that targets specific needs of programs with expectations of analytic competence via a workshop format, and (3) integration of statistical content into other graduate courses and the graduate research experience. All three utilize recurring vignettes, strong reliance on dental examples from consulting experiences within the dental college, and extensive use of graphics.

Lectures in Medical Statistics for Clinicians Using Internet Resources

◆ Todd G. Nick, University of Mississippi Medical Center

University of Mississippi Medical Center, 2500 North State St., Jackson, MS 39216

tnick@shrp.umsmed.edu

Key Words: *biostatistics education, evidence-based medicine, physician training, internet, clinical research*

A new course was started for fellows and physicians using internet resources to illustrate important concepts. The goal of this course was to improve the clinicians' statistical knowledge. The lectures were six 1.5-hour sessions held once a week. Each participant was provided a laptop with internet access. The topics covered included setting-up hypotheses, describing data, comparing means and percentages, assessing diagnostic accuracy, regression and correlation, and understanding simple and not-so-simple aspects of multivariable modeling. Each session used various internet sites to compute statistics. The internet was also used to retrieve medical research articles to illustrate concepts. Most clinicians do not have access to statistical software nor do they want it. However, the internet may be a feasible alternative since it has sites that can compute statistics and it's free and easy to use. Clinicians seemed very interested in learning about importing data to a site to compute descriptive and inferential statistics. Feedback was positive and many participants expressed interest in forming a working relationship for future research projects.

Development of a Web-supported Biostatistics Module that Includes Automatic Marking of Randomly Generated Datasets

◆ Paul Corey, University of Toronto; Malcolm Koo, University of Toronto; Andrew White, Source Level Computing; Raymond Tam, Source Level Computing

University of Toronto, First Floor McMurrich Bldg. Fac Medicine, Dept. Public Health Sciences, Torornto, ON M5S 1A8 Canada

paul.corey@utoronto.ca

Key Words: *web, education, biostatistics*

Language study includes vocabulary (words) and their rules (grammar) of association. Applied statistics is a foreign language. The ultimate education has student and teacher analyzing data, relating statistical practice to a scientific question and making decisions in face of uncertainty. Preparation for such enriched education involves learning the vocabulary and grammar of statistics. The vocabulary includes means and standard deviations. Grammar arranges these "words" into measures of statistical evidence, conventionally consisting of p values and confidence intervals. Driven by ASP and MySQL, a web module was developed to support a graduate medical statistics course. Students receive questions selected in real time from a database of datasets randomly generated using SAS and stored securely on the server. Correctness of the statistical analysis is examined using a multiple-choice format. Feedback from quizzes prepares the student for an online timed exam with a random dataset. Quizzes and exam marks are automatically made available to students. This module can be used to support a course in statistics or as a screen of statistical proficiency for a course or program.

Statistical (Mis)Issues in the Health Sciences: Statistical Methods May Not be the Problem

◆ T. Mark Beasley, University of Alabama, Birmingham; Sean Mulvenon, University of Arkansas

University of Alabama, Birmingham, Dept. of Biostatistics RPHB 327, 1665 University Blvd., Birmingham, AL 35294

mbeasley@uab.edu

Key Words: *teaching, statistics*

"Bad statistics live on, they take on lives of their own" (Best, 2001). The common misuse of statistical procedures in published research could, in part, be attributed to problems in the educational process. That is, the fact that researchers commonly misuse statistics is an accumulating problem due to miseducation. There are many themes that underlie this issue. They are not new. In fact, Hotelling (1940) discussed many of these same issues of statistics education even before the development of graduate programs in statistics. Therefore, the following topics are not meant to be comprehensive, novel, or exhaustive but they do focus on some key issues that have led to recent publicity over "Junk Science." Topics include: How Much of the Misuse of Statistics is Due to Our Own Malfeasance? Who Should Teach the Service Courses in Quantitative Methodology? What Can We Do as Specialists in Quantitative Methods to Improve Current Statistical Practice? What Can We Do as Statistics Educators to Improve Future Statistical Practice and Ensure that Our Students Conduct Research with Appropriate Quantitative Methods?

Restructuring of Biostatistics at UAB

◆ George Howard, University of Alabama, Birmingham; T. Mark Beasley, University of Alabama, Birmingham; Christopher S. Coffey, University of Alabama, Birmingham; Varghese George, University of Alabama, Birmingham

University of Alabama, Birmingham, Dept. of Biostatstics, School of Public Health, 1665 University Blvd., Ryals 327, Birmingham, AL 35294-0022

ghoward@uab.edu

Key Words: *genetics, clinical trials, teaching*

The Department of Biostatistics in the School of Public Health at the University of Alabama at Birmingham (UAB) has undergone a rapid expansion over the past five years. As part of this growth, the Section on Statistical Genetics (SSG) was inaugurated in March 2001. One of the SSG's major goals is to develop a high-caliber teaching program in statistical genetics. In this regard, the SSG has initiated a Statistical Genetics "track" within the Biostatistics PhD Program. The goal is to produce marketable doctoral students who earn a PhD in biostatistics with an emphasis in statistical genetics. The department is also developing a similar "track" for students interested in Clinical Trials. To fulfill these goals, we have carefully developed a cadre of investigators at the faculty, postdoctoral fellow, and graduate student levels. However, creating such a graduate teaching program has required a substantial restructuring of the PhD Program in biostatistics. This presentation will summarize the revised program and outline the educational, administrative, and political issues encountered in creating this program.

Some Classroom Roles for Historical Health Information

◆ Mark C. Fulcomer, Richard Stockton College of New Jersey; Marcia M. Sass, Richard Stockton College of New Jersey; Michael W. Holton, Richard Stockton College of New Jersey; Jenny L. Clayton, Emory University; S. David Kriska, Restat Systems, Inc.; Carla L. Hischmann, Richard Stockton College of New Jersey

Richard Stockton College of New Jersey, UMDNJ, 32 Rebecca Ct., West Trenton, NJ 08628

mcfulcomer@aol.com

Key Words: *health statistics, teaching, epidemiology, applied statistics, bibliographies, discriminant functions*

Utilizing three major sources developed earlier, this presentation highlights how historical data and bibliographies have been applied to the teaching of epidemiology, applied statistics, and other health-related courses. Two of the sources are large bibliographies, the first covering the literature related to discriminant functions up until 1978, now with over 1,000 distinct entries, and a second related primarily to infant mortality. An archive of annual health statistics reports for New Jersey that dates back to 1877 and includes over 40,000 page-images, the third source has also been incorporated in the classroom. The development of these sources provides several examples of data (e.g., time series) and disease outbreaks (e.g., smallpox, anthrax, and "consumption") directly relevant to teaching such topics as health statistics, planning, and policy. In addition, some of the examples, including the early use of colors in graphing and mapping, have applicability to other courses.

Professionalism for the Junior Statistician

◆ Lemuel A. Moye, University of Texas School of Public Health

University of Texas School of Public Health, E815 RAS Bldg., 1200 Herman Pressler, Houston, TX 77025

Lemuel.A.Moye@uth.tmc.edu

Key Words: *professionalism, character, productivity, leadership, principles*

The evolution of research has placed new challenges before junior scientists. Although junior scientists have fine educational backgrounds, they frequently do not yet have the poise, vision, or coping skills that they need to identify and sustain the optimum balance in their careers. However, junior scientists typically give little consideration to a set of guiding principles that would help them achieve this equilibrium. While productivity is a central part of the junior statistician, character development is itself a core value that is central to the growth of the professional. Character growth involves the development of assuredness, patience, administrative diligence, moral excellence, and collegiality. From these are derived the leadership principles of confidence, strength, vision, and charitable influence. In a turbulent working environment, the ability to be centered, and to, through a combination of tact, prestige, firmness, and understanding, appeal to the better nature of wavering people, is more essential then ever. Like the apples of gold in settings of silver, character and productivity must coexist in young professional statisticians.

192 Estimating Bayesian Marketing Models ◬

Section on Statistics and Marketing

Tuesday, August 10, 8:30 am–10:20 am

Bayesian Estimation of Bid Sequences in Internet Auctions Using a Generalized Record-breaking Model

◆ Eric T. Bradlow, Wharton School; Young-Hoon Park, Johnson Graduate School of Management

Wharton School, 3730 Walnut St., 761 JMHH, Philadelphia, PA 19104

ebradlow@wharton.upenn.edu

Key Words: *data augmentation, latent bidders, generalized record-breaking*

We generalize current research on Bayesian stochastic models for record-breaking problems by allowing for the situation in which the number of unsuccessful attempts between records is unobserved. That is, by viewing a record break as the maximum order statistic from an unseen number of attempts, we are able to infer the distribution from which they are derived. Our approach to addressing this problem is through data augmentation in which we stochastically draw both the number and magnitude of the unsuccessful attempts, and then model the observed record (maximal order statistic) conditional on the nonevents. The motivation for this research was derived from a currently active research topic in both marketing and economics: internet auctions, in which successive bids can be viewed as record-breaking events, but the latent number of bidders and their valuations are unseen. We apply our method to data obtained from an internet auction site and provide detailed inferences.

Reducing Respondent Burden in Ranking Tasks: Hierarchical Bayesian Analysis of Pairwise Comparisons with Covariates

◆ Well Howell, Harris Interactive

Harris Interactive, 5 Independence Way, Princeton, NJ 08540

whowell@harrisinteractive.com

Key Words: *hierarchical Bayesian analysis, Bradley-Terry Model, pairwise comparisons*

Hierarchical Bayesian (HB) analysis of pairwise comparison data with respondent covariates as suggested by Allenby and Ginter reduces the number of comparisons needed per respondent. Marketing practitioners seek respondent level information about which product attributes are more important in stimulating a first purchase or which features make a product more useful. The key issue is how attributes are ranked on some utility scale. However, self-reported utility measurements are notoriously unreliable. Pairs of attributes can be compared instead, but this quickly becomes an overwhelming task, as an 11-item list would produce 55 possible pairs. This task can be fractionalized so that each respondent sees only a portion of all pairs. The data are then analyzed with an HB model that includes respondent-level covariates in the prior. This paper shows that useful results can be produced on a test dataset with reduced respondent burden when covariates are included in the model, and that these results appear more accurate than those produced by alternative HB models or an aggregate-level analysis.

A Bayesian Switching Regression Model Where Independent Variables in Different Regimes Can Be Different

◆ Duncan K.H. Fong, Pennsylvania State University

Pennsylvania State University, Dept. of Marketing, 707 G Business Administration Bldg., University Park, PA 16802

i2v@psu.edu

Key Words: *hierarchial Bayes, multiple changepoints, variable selection, switching regressions*

Determination of changepoints in a regression setting has drawn considerable interest in the literature. A common assumption in existing work is that the same set of independent variables is used for all regimes. We propose a Bayesian approach to the problem but allowing different independent variables in different regimes. Thus, we can perform variable selection and changepoint determination simultaneously. A dataset on prescription sales for an ethical drug is used to illustrate the methodology.

Towards an Improved Prediction by Preceding a Traditional Classification Method by a Sequence Analysis Method

◆ Anita Prinzie, Ghent University; Dirk Van den Poel, Ghent University

Ghent University, Dept. of Marketing, Hoveniersberg 24, Gent, B-9000 Belgium

anita.prinzie@UGent.be

Key Words: *customer attrition analysis, classification methods, decision trees, Sequence Alignment Method, sequence analysis, financial services industry*

The inability to capture sequential patterns in the data is a typical drawback when using traditional predictive classification methods like logistic regression and decision trees. We overcome this caveat by clustering the observations on a sequential dimension using an element and position-sensitive Sequence Alignment Method (SAM) before estimating a traditional classification model for each of these clusters. Moreover, as we cluster on a sequential dimension assumed to influence the dependent variable, the sequential clustering improves the predictive performance of the overall classification model built on all observations. We illustrate this new procedure, which combines a sequence analysis method with a traditional classification method, by estimating a customer attrition model for a large International Financial Services Provider (IFSP). The results show that, by clustering the customers on their evolution in turnover at the IFSP, which is assumed to influence the customer attrition probability, the predictive accuracy of the overall churn model is improved.

Dynamic Generalized Linear Model for Correlated Binary Responses with Skewed Link

◆ Amitabha Bhaumik, Bristol-Myers Squibb Company; Ming-Hui Chen, University of Connecticut; Dipak K. Dey, University of Connecticut

Bristol-Myers Squibb Company, 1 Northwood R., Apt. #31, Storrs, CT 06268

amitabha.bhaumik@bms.com

Key Words: *Bayesian computation, generalized t distribution, latent variables, Markov chain Monte Carlo*

The logit, probit, and student t-link are used to model time series of correlated binary response data via dynamic generalized linear model. All of these link functions are symmetric. However in some applications the overall fit can be significantly improved by the use of asymmetric link function. A new skewed link model is introduced in this paper to model time series of binary responses through dynamic linear model. A new class of distributions which is more versatile than student's t distribution is used to model the data. Skewness is introduced by using a skewed distribution for the underlying latent variable. The new class of distribution is developed by using scale mixture of normal distribution with suitably chosen mixing density. This distribution overcome the limitation of student's t distribution arises due to its two boundary normal and Cauchy distributions. Analytic form of the distributions are obtained. Markov chain Monte Carlo method is used to simulate from posterior distributions. Bayesian inference, model diagnostics and model selections are considered. The proposed methodology is illustrated on a real dataset on house price from Dade county, Florida.

Stochastic Volatility Models for Ordinal-valued Time Series

◆ Claudia Czado, Munich University of Technology; Gernot Mueller, Munich University of Technology

Munich University of Technology, Center for Mathematical Sciences, Boltzmannstr. 3, D-85747 Garching, 85747 Germany

cczado@ma.tum.de

Key Words: *MCMC, stochastic volatility, time series, intraday financial data, transformation group*

Our aim is to model the intraday development of stock prices changes. We introduce a model for ordinal observations that are considered to be realizations of discretized versions of latent continuous variables. For these latent variables we assume a stochastic volatility (SV) model. MCMC methods for SV models have been considered in Chib, Nardari, and Shephard. To estimate the model parameters we consider an MCMC algorithm which, however, exhibits a bad convergence behavior. Therefore we develop a GM-MGMC (grouped move multigrid Monte Carlo) sampler, where additional grouped move (GM) steps (Liu and Sabatti 2000) are used. We show that these GM steps lead to a fundamental improvement in convergence. Finally, we fit the model to price changes of the IBM stock in January 2001.

193 Design Clinical Trials

Biopharmaceutical Section

Tuesday, August 10, 8:30 am–10:20 am

Maximum Imbalance in Treatment Assignments in a Clinical Trial that Uses the Minimization Allocation Procedure

◆ Olga M. Kuznetsova, Merck & Co., Inc.; John K. Troxell, Merck & Co., Inc.

Merck & Co., Inc., PO Box 2000, RY34-316, Rahway, NJ 07065-0900

olga_kuznetsova@merck.com

Key Words: *minimization, allocation in clinical trials, randomization, balance in treatment assignments*

Minimization is a covariate-adaptive allocation procedure designed to balance the treatment groups in a clinical trial over a large number of factors. Although the balancing properties of minimization have been studied extensively through computer simulations, little attention has been paid to the fact that in Taves' minimization the maximum possible imbalance in treatment assignments can be established by a complete enumeration of combinations of marginal imbalances. The maximum imbalances in treatment assignments will be established for Taves' minimization with the Friedman-White overall imbalance function. For example, for minimization that balances over four factors with two levels each, the imbalance in treatment assignments across a single level of any factor is shown never to exceed 11; the overall imbalance in treatment assignments will never exceed nine. Since minimization is defined through an algorithm that sets no limit to maximum allowed imbalance, eliminating a concern that an allocation sequence with any given large imbalance in treatment assignments might occur as a result of the minimization algorithm is important.

Application of Quality Control Statistics to Clinical Trials

◆ Terry L. Katz, PRA International; Hans Kroger, PRA International

PRA International, 4 Industrial Way West, Eatontown, NJ 07724

Katzterry@PRAIntl.com

Key Words: *precision, accuracy, biostatistics, quality control, clinical relevance*

By the time data from a clinical trial reaches a biostatistician for analysis it is commonly assumed to be of adequate quality. Onsite monitoring and double data entry seek to ensure the validity of data reported on the case report form, but inaccuracies, nonetheless, are commonplace. By adopting statistical quality control concepts from manufacturing that are used to ensure the precision and accuracy of measurement instruments, clinical data collection and review can be enhanced. Statistical process control methods, such as control charts, can detect outlier data values or trends, enabling early queries. These proactive methods would reduce the need to clean data at the analysis stage and the entire data management process would be improved. Examples of applications of these concepts to clinical data will include the accuracy and precision of vital sign measurements and routine (random) in-process data checks. The relevance of collected data to the patient's clinical condition, especially in relation to published expected normal ranges, will be considered.

Estimation of Treatment Effect for Clinical Trials with Interim Analyses

◆ Michael Lee, Merck & Co., Inc.; Hui Quan, Merck & Co., Inc.

michael_lee@merck.com

Key Words: *early stopping, sample size re-estimate, confidence interval, alpha-adjustment, estimation bias*

Interim analyses are commonly seen in clinical trials, especially in long-term studies. Results of an interim analysis can affect the study in several ways. The study can be stopped due to extremely favorable/unfavorable results, or sample size can be redetermined to achieve desired power. Research has been done largely under the framework of sequential and flexible study design. Many methods have been proposed for hypothesis tests that take the interim analysis into account and control the overall Type I error rate. However, methods for estimation of treatment effect are not available in many cases. Without adjustment for interim analyses, a point estimate is potentially biased if early stopping boundary is asymmetric. Furthermore, when the trial is stopped early, a confidence interval based on a small level assigned to the interim analysis is too conservative. We propose a method for estimation of treatment effect for use with interim analysis. The proposed method corrects potential bias and produces confidence intervals with adequate coverage rate. We use a parallel study design with interim analyses and simulations for illustration.

Statistical Models for Monitoring Enrollment and Projecting Patient Exposure in Ongoing Clinical Studies

◆ Guanghan F. Liu, Merck & Co., Inc.; Duane Snavely, Merck & Co., Inc.

Merck & Co., Inc., 10 Sentry Pkwy., BL 3-2, Blue Bell, PA 19422

guanghan_frank_liu@merck.com

Key Words: *enrollment, patient-years, Kaplan-Meier estimate, piecewise linear model*

During the enrollment period of a clinical study, the sponsor may monitor the patient enrollment process and estimate the time at which the study enrollment completes. Based on the estimated timeline, the sponsor can make preparations to close the patient screening and enrollment process at the optimal time point. The sponsor may also desire to "cut off" a study at a specific time for data analysis and reporting. The study may be cut off at the time when enough patient exposure data is accumulated to fulfill regulatory requirements. For this purpose, the study sponsor may estimate the total patient-years of exposure for a given future time point. We introduce a few statistical methods to project the cumulative patient enrollment curve and to project the patient-years of exposure at a given future time point. Survival models with piecewise linear regression are used to estimate the enrollment process and the probability of a patient staying in the study at given time point. Clinical trial data are used to illustrate that the methods have good flexibility to provide accurate estimation for the cumulative enrollment and for the total patient-years of exposure.

Creating a Pharmacogenomics Data Warehouse

◆ Surya Mohanty, Johnson & Johnson Pharmaceutical R&D, LLC

Johnson & Johnson Pharmaceutical R&D, LLC, 1125 Trenton-Harborton Rd., Titusville, NJ 08560

smohanty@prdus.jnj.com

Key Words: *pharmacogenomics, anonymization, data warehouse*

In clinical trials is has become customary to collect blood samples for the purpose of genotyping. The intent is to associate genotypic or phenotypic measures to clinical measures. The presentation will discuss about what paradigms could be used for collection and analyses of genes know to impact the disease as well as the situations where no such information may be available. In order to generate any hypothesis on the association, large number of patient information would be required. This information will typically span over multitude of trials for the same compound and the same indication. Therefore a systematic method of storing these data is necessary. This presentation will discuss such issues.

Other issues will also be considered: (1) Once the data has been stored, what type of tools is required to analyze such data? (2) How do we implement these information in designing future trials? (3) What is the regulatory guidance?

Clinical-data-warehousing in Pharmaceutical Research

◆ C. V. Damaraju, J&J Pharmaceutical Research & Development

J&J Pharmaceutical Research & Development, 1125 Trenton-Harbourton Rd., Titusville, NJ 08560

cdamara1@prdus.jnj.com

Key Words: *knowledge discovery, data warehouse, safety and efficacy, standards, regulatory guidance*

Safety and efficacy data constitute the bulk of the pharmaceutical research data repository. An efficient integration of these two components entails knowledge discovery in a timely fashion. We discuss the existing paradigms in data-warehousing for safety and efficacy data and highlight some of the problems faced in structuring a clinical data warehouse. We also note the role of relevant regulatory guidelines in creating and maintaining overall data standards in the pharmaceutical industry.

194 Emerging Technologies in Genetics and Molecular Biology ▲ ⌘

Biometrics Section

Tuesday, August 10, 8:30 am–10:20 am

Analysis of a Large Structure/Biological Activity Dataset Using Recursive Partitioning, Simulated Annealing, and Genetic Algorithm

◆ Ke Zhang, North Carolina State University; Jacqueline M. Hughes-Oliver, North Carolina State University; S. Stanley Young, NISS

North Carolina State University, 3109A, Kings Ct., Dept. of Statistics, Raleigh, NC 27606

kzhang3@stat.ncsu.edu

Key Words: *recursive partitioning, simulated annealing, genetic algorithm, SAR, HTS, drug discovery*

Large quantities of structure and biological activity data are quickly accumulated with the development of high-throughput screening (HTS) and combinatorial chemistry. Analysis of structure-activity relationships (SAR) from such large datasets is becoming challenging. We propose a method which implements Recursive Partitioning (RP), Simulated Annealing (SA), and Genetic Algorithm (GA) to produce stochastic RP trees. RP is a statistical method that can identify SAR rules for classes of compounds that are acting through different mechanisms in the same dataset. In the new algorithm a set of structural descriptors is extracted at each splitting node by using SA combined with GA as a stochastic optimization tool. For one dataset, results show that the new method is advantageous in predicting potency based on quantitative SAR model.

Methods for Evaluating T Cell Repertoire via CDR3 Spectratyping Analysis

◆ Susan M. Geyer, Mayo Clinic; Matthew Maurer, Mayo Clinic

Mayo Clinic, 200 First St. SW, Kahler 1A, Rochester, MN 55905

geyer@mayo.edu

Key Words: *immunology, spectratyping, T cell repertoire*

Analysis of T cell repertoires (TcR) has been the focus of recent studies of immunologic health. Diversity of this repertoire can effectively be measured through spectratyping, an RT-PCR-based method for estimating V gene representation and diversity of complementarity determining regions 3 (CDR3s). Primary focus has been on graphical analyses of chromatographs, where normal TcRs exhibit an array of variable CDR3 lengths for a single rearranged V gene. Chromatographs could be used for evaluating a few patients, but for larger studies is unfeasible. Thus, we developed methods to summarize key features of spectratyping data and methods for TcR analysis where three primary measures are used to quantify the TcR diversity and health. Our approach focuses on quantifying the total number of peaks representing individual CDR3 lengths, the proportion of peaks representing productively rearranged alpha and beta transcripts, and disturbances in the distribution of CDR3 lengths. Examples of these methods to identify patterns in T cell diversity across different age groups as well as between healthy subjects and cancer patients will be presented.

Expression Stability of Internal Controls: Comparison of 10 Common Markers

◆ Suddhasatta Acharyya, Brown University; Jyotirmoy Dey, Schering-Plough Research Institute

Brown University, 167 Angell St., 2nd Floor, Providence, RI 02912

acharyya@stat.brown.edu

Key Words: *internal control, gene expression, cancer, Pittman's test, SAGE, microarray*

Internals controls are markers that are believed to have "stable" expressions in multiple tissue and organ types and are hence useful in validating and normalizing experimental data on other markers that differentially express themselves across tissue types, for example, normal and cancer tissue. We investigate serial analysis of gene expression (SAGE) technique based expression of 10 commonly used internal control to assess their suitability as controls. The 10 markers selected were PPIA, GAPDH, beta-actin, G6PDH, Lamin B1, Cox 4, Aldolase A, Piruvate Kinase, 36B4, and RPL32. All available expression data for these 10 markers in five different organs—brain, breast, prostate, colon, and ovary—were obtained from the National Center for Biotechnology Information web site. For each organ we first compare the expressions of a given marker in bulk tissue (in-vivo) and cell lines (in-vitro). Average expressions for each marker in bulk tissue were then compared across organs using suitable analysis of variance procedures. We then formalize the notion of "stability of expression" of a marker, and for each organ, select the "most stable" among the internal controls on study.

Nonparametric and Parametric Approaches to Control the FDR in Gene Expression Studies

◆ Ernst C. Wit, University of Glasgow

University of Glasgow, 15 University Gardens, Statistics Dept., Glasgow, G12 8QW UK

ernst@stats.gla.ac.uk

Key Words: *FDR, microarray, multiple testing, gene expression, false discovery rate, mixture model*

The multiple testing procedure by Efron et al. (2001), which employs a mixture model in an empirical Bayes setting, has sparked interesting new directions in multiple testing. This procedure departs from most traditional approaches in that it attempts to borrow strength across variables to improve inference as well as that it claims to control the False Discovery Rate. The relative innovative nature of this technique warrants a careful study into its properties. This talk extends some work by Genovese and Wasserman on the properties of this multiple testing procedure. More can be found in “Statistics for Microarrays: Design, Analysis and Inference.”

Multiply Imputing Ordinal Variables Using Latent Continuous Variables in Tissue Microarray Data

◆ Jun Xing, University of California, Los Angeles; Thomas R. Belin, University of California, Los Angeles; Steve Horvath, University of California, Los Angeles

jxing@ucla.edu

Key Words: *multiple imputation, latent variable, joint modeling, tissue microarray*

Tissue microarrays (TMAs) are a new high-throughput tool for study of protein expression patterns that are increasingly being used to evaluate the diagnostic and prognostic importance of tumor biomarkers. Protein expression variables are typically measured on an ordinal scale and typically have non-normal distributions, frequently with substantial skewness. The fact that protein expression values may be missing creates challenges when looking for associations between tumor biomarkers and clinical outcomes. The motivating example for the present work comes from TMA data with six candidate biomarkers obtained from bladder cancer patients. We propose to impute missing values by simultaneously modeling of protein expression covariates, survival time, and other covariates, with interest focusing on the connection between protein expression covariates and disease risk. This presentation will contrast results from simple approaches such as median imputation and multivariate normal imputation and will outline a joint modeling approach that we anticipate will offer advantages over simpler techniques.

Statistical Methods for Analyzing ChIP-Chip High Density Oligonucleotide Array Data

◆ Sunduz Keles, University of California, Berkeley; Mark van der Laan, University of California, Berkeley; Siew L. Teng, University of California, Berkeley; Sandrine Dudoit, University of California, Berkeley

University of California, Berkeley, Division of Biostatistics, 140 Earl Warren Hall, #7360, Berkeley, CA 94720-7360

sunduz@stat.berkeley.edu

Key Words: *microarrays, chromatin immunoprecipitation, multiple testing, regulatory motif finding, ChIP on chip data*

Cawley et al. (2004) recently mapped the locations of binding sites for three transcription factors along human chromosomes 21 and 22 using chromatin immunoprecipitation (IP) of transcription factor bound DNA followed by high density oligonucleotide hybridization of the IP-enriched DNA. This so-called ChIP-Chip technology generates a new type of genomic data for the identification of transcription factor binding sites. We investigate this data structure and propose methods for analyzing it. The proposed methods involve testing for each probe whether it is part of a bound sequence using a scan statistic that takes into account the spatial structure of the data. Different multiple testing procedures are considered for controlling the familywise error rate and false discovery rate. A nested-Bonferroni adjustment that is slightly more powerful than the Bonferroni adjustment when the test statistics are dependent is provided. Simulation studies show that taking into account the spatial structure of the data improves the sensitivity of the multiple testing procedures. Application of the proposed methods to p53 ChIP-Chip data identified many potential target binding regions for p53.

Is Scientific Discovery of DNA Fingerprint by Chance or by Design?

◆ Harry Yang, MedImmune, Inc.; Iksung Cho, MedImmune Vaccines

MedImmune, Inc., 35 W. Watkins Mill Rd., Gaithersburg, MD 20878

yangh@medimmune.com

Key Words: *DNA fingerprint, restriction enzyme, partial digestion, PCR, probabilistic model, reproducibility*

DNA fingerprinting is a microbiological technique widely used to find a DNA sequence specific for a microbe. It involves slicing the genomes of the microbe into DNA fragments with manageable sizes, sorting the DNA pieces by length and finally identifying a DNA sequence unique to the microbe, using probe-based assays. We introduce a probabilistic model to estimate the chance of identifying a specific DNA sequence of any given length from the genome of a microbe when the DNA fingerprinting method is employed. We derive a closed-form functional relationship between the chance of finding a specific DNA sequence, the length of the sequence, cutting efficiency of restriction enzyme and number of copies of the microbe genomes used in the fingerprinting experiment. The model can be potentially used to guide experiments in maximizing the chance of finding a DNA fingerprint of interest. It can also be used to assess the reproducibility of a specific DNA fingerprint discovery. In a broader sense, we demonstrate that the discovery of a DNA fingerprint of a microbe is governed more by chance than by design.

195 Interviewer Issues ▲

Section on Survey Research Methods, Social Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Detecting Interviewer Data Falsification

◆ Javier Porras, NORC, University of Chicago; Edward English, NORC, University of Chicago

NORC, University of Chicago, 55 E Monroe, Suite 4800, Chicago, IL 60603

porras-javier@norc.net

Key Words: *falsification, fabrication, interviewer*

Interviewer data falsification reduces the quality of survey data. Even when applying industry-standard monitoring and validation techniques, it can be difficult to detect data fabricated by an interviewer. A goodness-of-fit test to the Benford distribution and other tests that compare patterns and distributions can more easily identify data falsification. We tested these methods by intentionally fabricating records, adding them to legitimate survey datasets, and checking whether the tests correctly flag the fabricated records as suspicious. This paper details the approach used and summarizes the results.

Identification and Impact of Faked Interviews in Surveys—An Analysis of Genuine Fakes by Means of Benford's Law and Robust Machine Learning Approach for Outlier Detection

◆ Gert G. Wagner, DIW Berlin; Christin Schaefer, Fraunhofer FIRST.IDA; Klaus-Robert Mueller, Frauenhofer FIRST.IDA and University of Potsdam

DIW Berlin, Koenigin Luise Strasse 5, Berlin, 12163 Germany

gwagner@diw.de

Key Words: *Benford's Law, faked interviews, robust machine learning, genuine fakes, outlier detection, SOEP*

Panel data provide a unique opportunity to identify data which are actually faked by interviewers. By comparing data of two waves, unequivocal fakes are easily identifiable. We use the fakes which we know t in the raw data of the German Socio-Economic Panel Study (SOEP) for an analysis of the potential impact of nondetected fakes on survey results. Because in most surveys there is no second wave as they are of a purely cross-sectional nature, we searched for methods which do not need two waves of data. We test (1) an unconventional benchmark called Benford's Law, which is used by numerous accountants to discover frauds, and we apply (2) a robust machine learning approach for outlier detection utilizing a resampling technique. Through a combination of both methods we can identify the majority of interviews (0.5% of all interviews of SOEP) which are unequivocally faked. However, the major result is that the faked and fraudulent records have almost no impact on the mean and the proportions of substantial results. Finally, one should note that, except for some fakes in the first two waves of sample E, faked data were never disseminated within the widely used SOEP.

Interviewer Experience, Interview Order, and Interviewer Behaviors and Attitudes

◆ Kristen Olson, University of Michigan; Andy Peytchev, University of Michigan

University of Michigan, Institute for Social Research, 426 Thompson St., Rm. 4026, Ann Arbor, MI 48104

olsok@umich.edu

Key Words: *interviewer effects, interviewer behavior, interview order, interviewer experience, interview length, hierarchical linear models*

Traditional statistical analyses of interviewer effects on survey data do not examine whether these effects change over a survey cycle. However, the nature of the survey interview is dynamic, interviewers' behaviors and perceptions may evolve as they gain experience, potentially affecting data quality. This paper looks at how interview length, number of reports to open-ended questions, and interviewer evaluations of respondents change over interviewers' workloads. Multilevel models with random interviewer effects are used to account for the clustering of cases within interviewers and individual interviewer characteristics in the 1984 and 1988 National Election Studies. The NES released sample in four replicates, minimizing the confound between order in an interviewers' workload and sample composition. We find that over the course of the studies, all three measures change significantly. Interviewer prior survey experience also was significantly negatively related to the length of the interview. These findings have implications for interviewer training prior to and during studies, as well as suggesting future research to reveal why these behaviors and perceptions change.

CAPI Versus Audio-CASI: Understanding the Role of Privacy

◆ Mick P. Couper, University of Michigan; Eleanor Singer, University of Michigan; Roger Tourangeau, University of Maryland

University of Michigan, Survey Research Center, PO Box 1248, Ann Arbor, MI 48109

mcouper@umich.edu

Key Words: *audio-CASI, sensitive questions, CAPI*

Audio-CASI has been widely adopted by the survey industry for administering questionnaires on sensitive topics. Several field experiments have demonstrated the advantages of audio-CASI over interviewer administration for sensitive questions, but there have been relatively few attempts to understand the mechanisms by which audio-CASI improves the veracity of self-reports. We report on a laboratory-based experiment to explore some of these issues in detail, specifically the possible role of privacy in affecting the answers provided in the two modes. We carried out a 2◊3 experiment that crossed the mode of administration (CAPI versus audio-CASI) with three different privacy conditions. In the high privacy condition, the subject completes the survey in the presence of the interviewer, but with no third party present. In the medium privacy condition, the subject is introduced to an observer at the outset, but there is no expectation of later interaction with that observer. In the low privacy condition, the subjects are introduced to the observer, but also told that the observer will debrief them at the end of the interview, raising expectations about later interaction.

Monitoring, Experience, and CATI Interviewer Performance in the American Community Survey

◆ Bryan D. Garrett, U.S. Census Bureau; Andre L. Williams, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., STOP 9200, Planning, Research, and Evaluation Division, Washington, DC 20233-9200

b.dale.garrett@census.gov

Key Words: *feedback, quality control*

The American Community Survey (ACS) collects important socioeconomic data continuously in three month cycles. ACS mails out questionnaires during month one, then follows nonrespondents with computer-assisted telephone interviewing (CATI) in the second month and personal interviewing in the third month. Quality control for CATI is performed by monitoring interviewers, and providing them with regular feedback on their performance. Interviewers do not know exactly when and how often they are monitored. Over time, the interviewer gains experience and becomes more consistent in performance. Coaches rate the monitored interview session and store records in electronic files. This paper examines distributions of measures of interviewer experience and interview performance for the 2001 and 2002 ACS and examines changes in interview performance between 2001 and 2002.

Action at a Distance: Interviewer Effort and Nonresponse in the SCF

◆ Arthur B. Kennickell, Federal Reserve Board

Federal Reserve Board, Mail Stop 153, Washington, DC 20551

Arthur.Kennickell@frb.gov

Key Words: *nonresponse, interviewer effects*

In field surveys, the behavior of interviewers is usually only partially observable, at best. Like most other people, interviewers will tend to choose the easiest path to satisfy the measurable requirements they face—most often, requirements to complete a number of interviews and to keep costs within some bounds. In "suiting themselves" otherwise, the interviewers may impose an additional behavioral filter on the information obtained in the survey. In particular, unit nonresponse may become a function not only of respondents' willingness to cooperate, but also of interviewers' application of effort. Evidence from the 2001 Survey of Consumer Finances (SCF) presented in Kennickell (2003) indicates that interviewers imposed a systematic pattern of selection in nonresponse through the differential application of effort in that survey. As a consequence of that research, a new case control method was developed for the 2004 SCF to impose a more equal level of treatment on all cases through the first part of a phased field effort. This paper presents a stylized version of that method, analyses nonreponse at the end of the first phase, and explores further possibilities.

196 Variance Estimation I ▲

Section on Survey Research Methods

Tuesday, August 10, 8:30 am–10:20 am

Zero-calibrated Variance Estimators

Alan H. Dorfman, Bureau of Labor Statistics; Lawrence R. Ernst, Bureau of Labor Statistics; Christopher J. Guciardo, Bureau of Labor Statistics; ◆ Michael Sverchkov, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 1950, Washington, DC 20212

Sverchkov.Michael@bls.gov

Key Words: *BRR estimator, calibration estimator, jackknife variance estimator, survey sampling, two-phase sampling*

We consider variance estimation in the case of a calibrated estimator of a finite population total of a variable of interest Y. An estimator is calibrated with respect to a particular design variable X when the estimator is such as to yield the actual population total of the X's when X is taken as the variable of interest (that is, when it is substituted for Y in the estimation formula), whatever the sample. In this situation, we say a variance estimator is zero-calibrated if, for Y = X, it takes the value 0. This is a desirable property, since otherwise we may suspect the variance estimator of introducing gratuitous noise. We consider the utility of this criterion in identifying undesirable estimators as well as in adjusting weights for replication variance estimators in two-phase sampling.

Replicate Variance Estimation for the National Survey of Parents and Youth Using Eigenmatching

◆ Louis Rizzo, Westat; David R. Judkins, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

rizzol1@westat.com

Key Words: *jackknife, BRR, finite population corrections, multistage sample*

This paper describes a case study in replicate variance estimation from the National Survey of Parents and Youth, a nationally representative sample of families with youths 9-17 years old. The replicate weights are derived using a method first proposed by Fay for providing consistent replicate variance estimators for complex sample designs. This method defines modified jackknife replicate weights to provide replicate variance estimators which exactly match the sample-design unbiased variance estimators for totals, by "eigenmatching": algebraically reproducing the eigenvalue-eigenvector structure of the sample-design unbiased variance estimator (including the exact finite population corrections at each level of sampling). These replicate variance estimators will then be consistent for sufficiently smooth functions of unbiased estimators of totals such as ratios, regression coefficients, etc. (as is achieved for with-replacement sampling using unmodified jackknife or BRR methodologies). The NSPY application is a multistage stratified sample design with primary sampling units as the first stage of selection.

Using the Fractional Imputation Methodology to Evaluate Variance Due to Hot-deck Imputation in Survey Data

◆ Adriana Perez, University of Texas Health Science Center, Houston; John M. Finamore, U.S. Census Bureau; Jeffrey H. Stratton, U.S. Census Bureau

University of Texas Health Science Center, Houston, 80 Fort Brown SPH RAHC Rm N.200, Brownsville, TX 78520

aperez@sph.uth.tmc.edu

Key Words: *ignorability, missing at random, item nonresponse, serpentine sorting, nearest neighbor, successive difference replication*

Limited variance estimation methods exist in the statistical literature for large complex sample surveys after the imputation of missing data. Since variance estimates depend on both the sample design and the imputation method, methods under design-based or model-assisted perspectives exists. This paper examines empirically the effect on the variance estimate due to the use of hot-deck imputation with a nearest neighbor donor in the 1999 Survey of Doctorate Recipients. This examination involves a comparison to results from the pairwise fractional hot-deck imputation methodology. For this evaluation, two distinct donors were selected for each missing value and imputation fractions for the donors were computed. One hundred sixty replicate weights were created to determine variance estimates. Each replicate included an adjustment to account for the imputation fractions. The successive difference replication variance estimation method was used. In addition, three stages of weighting adjustments were evaluated to determine the most appropriate stage to incorporate the fractional imputation weight adjustment.

Variance Estimation for the National Compensation Survey When PSUs are Clustered Prior to the Second Phase of Sampling

◆ Christopher J. Guciardo, Bureau of Labor Statistics; Alan H. Dorfman, Bureau of Labor Statistics; Lawrence R. Ernst, Bureau of Labor Statistics; Michael Sverchkov, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 3160, Washington, DC 20212

Guciardo.Chris@bls.gov

Key Words: *replication, jackknife, zero-calibrated, NCS, BRR*

The National Compensation Survey uses a three-phase sample: areas (PSUs), establishments, and jobs. For phase two, we initially planned to take independent, stratified samples within each PSU, by fixing a sample size for each PSU and then allocating these sizes to industry sampling strata, ensuring at least one establishment sampled per PSU x industry stratum. Yet for the smaller, noncertainty PSUs, desired sizes were often less than the number of industry strata; so 99 of these PSUs were collapsed into a single cluster before sampling. Variances were estimated using balanced repeated replication (BRR), with the collapsing of the 99 PSUs ignored in the calculation of the variance estimates. Wang, Dorfman, and Ernst (to appear) evaluated via a simulation study the accuracy of using standard BRR for this type of sampling and also investigated an alternative variance estimator combining model-based and design-based ideas. We investigate alternate variance estimators, including zero-calibrated BRR and jackknife, evaluating the accuracy of these variance estimates by drawing multiple samples from our frame.

Replicate Allocation and Variance in the Shelter Component of the CPI

◆ William E. Larson, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212

larson_w@bls.gov

Key Words: *housing, rent, REQ*

The relationship of replicate allocation within PSUs to variance in the housing component of the Consumer Price Index is examined. Fluctuations in index-area variance are important contributors to overall variance. Present methods used to allocate housing segments to replicates involve the pre-initiation assignment of segments to replicates, which can result in marked design imbalance and contribute to spikes in index-area variance. Simulations of the housing price relative calculation are performed using post-initiation replicate assignments. Overall national variance is similar, but extremes in index-area variance are attenuated.

Replicate Variance Estimation For Combined Data of Independent Surveys

◆ Donsig Jang, Mathematica Policy Research, Inc.; Amang Sukasih, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024

djang@mathematica-mpr.com

Key Words: *SESTAT, replication method, combined dataset*

Several survey data are often combined into one mega dataset to represent the population covered by those surveys. Survey estimates for the combined dataset can be calculated under the usual stratified sample design assumption in that each survey is regarded as the first-tier stratification variable. In many cases each survey component has its own replication method for variance estimation method which accounts for its sample design and other adjustment procedures. However, replication methods used for each component survey are not necessary same. Consequently, a way to combine replicate weights from each survey is desirable. The Scientist and Engineers Statistical Data System (SESTAT) has an integrated data system combining three independent surveys. We construct combined replicate weights for the SESTAT based on individual survey specific replicate weights and evaluated variance estimates based on the combined replicate weights. Empirical results show that in practice this approach is preferable over the approach that approximates the replicate weights by using unified replication method.

The Bootstrap Variance Estimator in a Nested Two-stage Sample Design with High Sampling Rates

◆ Steven Kaufman, NCES

steve.kaufman@ed.gov

Key Words: *collapse stratum, domain estimation*

When the sampling rates are high, it is important to reflect the finite population correction (FPC) in the variance estimator. With replication methodologies, this can be accomplished by multiplying the replicate weights by an appropriate factor. This is the same thing as multiplying the replication variance by the first-stage FPC. Since this is a simple multiplication, this factor is applied to all variance components. In a single stage sample design, this works quite well because there is only one variance component, the first-stage component, which needs to be multiplied by the first-stage FPC. In multiple stage designs, the second and subsequent stages variance components are correct without adjustment. So this adjustment, when applied, will necessarily introduce a bias in the overall variance estimate. With the bootstrap, it is easy to adjust the variance estimator to correct for this bias. However, in this process, it is usually assumed that there are at least two units selected within each stratum for all selection stages. This paper describes how to modify the bootstrap procedure to handle the situation where only one unit is selected within a second and/or subsequent stage.

197 Statistical Uses of Administrative Records Data

Social Statistics Section, Section on Government Statistics

Tuesday, August 10, 8:30 am–10:20 am

What's the Number? An Evaluation of Phone Number Lookup for a Special Population in a Telephone Survey

Wei Yen, University of California, Los Angeles; E. Richard Brown, University of California, Los Angeles; ◆ Charles DiSogra, University of California, Los Angeles; Sherman Edwards, Westat; Kathleen Call, University of Minnesota

University of California, Los Angeles, 10911 Weyburn Ave., Suite 300, Center for Health Policy Research, Los Angeles, CA 90024

cdisogra@ucla.edu

Key Words: *Medicaid administrative data, address and telephone matching, telephone number lookup*

Surveying the Medicaid population by telephone usually requires access to the Medicaid program administrative data to obtain contact information. What if the administrative data contain no telephone numbers? This is the case in a study of the Medicaid recipients in California. The state administrative database contains addresses of the Medicaid recipients, but no telephone numbers. The study protocol requires that the survey be conducted by telephone. An alternative solution is adopted in which the addresses of the Medicaid recipients are used to look up phone numbers from published and unpublished telephone directories. This paper discusses the process of the look-up, the verification of the phone numbers obtained through the look-up, the effectiveness of this approach for reaching the Medicaid population, and implications for other special populations with similar contact data limitations.

Development of the School District Analysis System

◆ Stephen R. Wenck, Synectics for Management Decisions, Inc.; Meyer Zitter, Synectics for Management Decisions, Inc.; Tai Phan, National Center for Education Statistics

Synectics for Management Decisions, Inc., 1901 N. Moore St., Suite 900, Arlington, VA 22209

SteveW@smdi.com

Key Words: *school district data, National Center for Education Statistics, cross-tabulations, education policy, 2000 decennial census data, web-based data delivery*

The Census School District Special Tabulation contains summary data from school districts, counties, states, and the nation using 2000 Census short- and long-form data. These data provide tabulations for parents with children, households with children, children's own characteristics, children's household characteristics, and children's parents' characteristics, as well as total population. These data are augmented by data from the National Center for Education Statistics in areas of analytical interest pertaining to children and K-12 education (particularly the Common Core of Data) to form the 2000 School District Analysis System (SDAS), an online cross-tabulation system using the 2000 decennial census in a different perspective to examine children in relation to their family and household characteristics. The SDAS presents counts of children or school districts by various enrollment and school district characteristics. This paper discusses the data preparation including reclassifying the data by enrollment status & selected school district characteristics. The paper then examines uses for the SDAS including selected results as well as the potential for comparisons to the 1990 product.

Assessing the Census Master Address File in Small Areas Using Administrative Records

◆ Dawn E. Rosser, U.S. Census Bureau

U.S. Census Bureau, PRED, Admin. Recs. Rsch., FB#2, Rm.#1600, Washington, DC 20233

Dawn.E.Rosser@Census.gov

WITHDRAWN

Key Words: *Statistical Administrative Records System, geocoding*

The Master Address File (MAF) is the Census Bureau's list of housing units. The quality of the MAF is important because it is the sampling frame for ongoing surveys such as the American Community Survey and will be the base for the 2010 Decennial Census address list. Administrative records have much potential benefit for improving the quality of the MAF by identifying addresses that are suspect for deletion from the MAF. The Statistical Administrative Records System (StARS) is a database of address and person records built entirely from administrative records. The scale of StARS is comparable to the MAF, and StARS is independent of the MAF. This paper discusses one method to assess MAF quality in small areas using StARS.

Analysis of ACS Food Stamp Program Participation Underestimate

◆ Dean Resnick, U.S. Census Bureau; Susan Love, U.S. Census Bureau; Cynthia Taeuber, University of Baltimore/Jacob France Center; Jane M. Stavely, University of Baltimore/Jacob France Center

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233

Dean.Michael.Resnick@census.gov

Key Words: *American Community Survey, food stamps, survey methodology, ACS*

The U.S. Census Bureau's American Community Survey (ACS) has a question that asks households whether they have received food stamps in the previous year. Based on responses to this question, the Census Bureau estimates the number of households who are participating in this program. However, when these estimates are compared to family participation counts generated from food stamp administrative records, they are found to be quite a good deal lower. To understand the limitations of Census surveys for estimating program participation, we investigated the sources of this discrepancy. We matched ACS records to Maryland food stamp records at the individual level, and then used these individual level match results to generate family match results. Analysis of these family match results suggests that the largest factor causing this discrepancy is families reporting that they do not receive food stamps when in fact they do. This misreporting seems to be largely of function of who in the household received the food stamps, for how long they were received, and the interval since they were last received. Family definition and survey coverage also contributed to the underestimation.

Statistics on Nonprofit Organizations: Fulfilling the Public Right to Evaluate Their Performance?

◆ Martin H. David, Urban Institute

Urban Institute, 425 8th St. NW #1144, Washington, DC 20004-2115

mdavid@ui.urban.org

Key Words: *business demography, tax-exempt organizations, coverage*

Benefits and costs characterize nonprofit organizations. They provide many services, especially education, health, and social welfare. They are exempt from many taxes and contributors reduce their own tax liabilities. Data on nonprofits are collected through censuses with cut-off samples, surveys, and administrative records. Births of nonprofits and their subsequent reporting of finances are examined. We identify births of nonprofits from rulings that sanction exempt status. We calculate the rate at which organizations ever file annual information returns for periods that include the birth year. Analysis reveals that less than 20% of new charitable organizations ever file; the rate varies widely for subpopulations. Reliability of the filing rate is examined by investigating coverage and error rates in the annual information files and the approval rulings. Robustness is examined by estimating birth and death rates from aggregated data and making corresponding adjustments to the ever-filed rate. Thresholds and categorical exemption of religious organizations reduce filing rates.

198 Bayesian Inference for Gene Expression Data ▲

Section on Bayesian Statistical Science

Tuesday, August 10, 8:30 am–10:20 am

Multiclass Cancer Diagnosis Using Bayesian Kernel Machine Models

◆ Sounak Chakraborty, University of Florida; Bani K. Mallick, Texas A&M University; Debashis Ghosh, University of Michigan; Malay Ghosh, University of Florida

University of Florida, 103 Griffin/Floyd Hall - PO Box 118545, Dept. of Statistics, Gainesville, FL 32611

schakrab@stat.ufl.edu

Key Words: *Gibbs sampling, Markov chain Monte Carlo, Metropolis-Hastings, microarrays, reproducing kernel Hilbert space, support vector machine*

Precise classification of tumors is critical for cancer diagnosis and treatment. Using gene expression microarray data to classify tumor types is a very promising tool in cancer diagnosis. In recent years, several works showed successful classification of pairs of tumor types using gene expression patterns. Usually the multicategory classification problems are solved by using a bunch of binary classifiers which may fail in variety of circumstances. We have proposed the Bayesian multicategory kernel machine for multicategory classification. We have compared our method with several other available methods. The methods are illustrated with two real-life microarray datasets.

Modeling Nonlinear Gene Interactions using Bayesian MARS

◆ Veerabhadran Baladandayuthapani, Texas A&M University; Bani K. Mallick, Texas A&M University; Raymond J. Carroll, Texas A&M University; Chris C. Holmes, Imperial College

Texas A&M University, 3143 TAMU, College Station, TX 77843

veera@stat.tamu.edu

Key Words: *MARS, MCMC, microarray, nonparametric, hierarchical modeling*

DNA microarray technology enables us to monitor the expression levels of thousands of genes simultaneously, and hence to obtain a better picture of the interactions between the genes. To understand the biological structure underlying these gene interactions, we present a statistical approach to model the functional relationship between genes and also between genes and disease status.We suggest a hierarchical Bayesian model based on Multivariate Adaptive Regression Splines (MARS) to model these complex nonlinear interaction functions. The novelty of the approach lies in the fact that we attempt to capture the complex nonlinear dependencies between the genes which otherwise would have been missed by linear approaches. Owing to the large number of genes (variables) and the complexity of the data, we use Markov chain Monte Carlo (MCMC)-based stochastic search algorithms to choose among models. The Bayesian model is flexible enough to identify significant

genes as well as model the functional relationships between them. The effectiveness of the proposed methodology is illustrated using two publicly available microarray datasets: leukemia and hereditary breast cancer.

Gene Selection Using a Two-level Hierarchical Bayesian Model

◆ Kyounghwa Bae, Texas A&M University; Bani K. Mallick, Texas A&M University

Texas A&M University, 1 Hensel Dr. #U3G, College Station, TX 77840

kbae@stat.tamu.edu

Key Words: *gene selection, Markov chain Monte Carlo, two-level hierarchical model, cDNA data*

The fundamental problem in gene selection via cDNA data is to identify which genes have different gene expression between two tissue types (e.g., normal and cancer). CDNA data has a large number of genes for a single sample while sample sizes are relatively small. Therefore, estimates which incorporate variable sparsity are desirable for this kind of problem. This paper proposes a two-level hierarchical Bayesian model for variable selection which assumes a prior that favors sparseness in parameters. We adopt a Markov chain Monte Carlo (MCMC) based computation technique to simulate the parameters from the posteriors. The method is applied to leukemia data from Golub et al. (1999) and a published dataset of Hendenfalk et al. (2001) on breast cancer.

The Geno-Pheno Problem: Predicting Fitness from Genotype Using the Exchangeable within Subsets Prior and Prior Model Selection

◆ Christina R. Kitchen, University of California, Los Angeles; Robert E. Weiss, University of California, Los Angeles; Gang Liu, University of California, Los Angeles

University of California, Los Angeles, 650 Charles E. Young Dr., Dept. of Biostatistics, Los Angeles, CA 90095-1772

cr@ucla.edu

Key Words: *HIV, viral fitness, hierarchical modeling, Bayes theorem*

The genotype-phenotype problem is the fundamental problem of genetics where one utilizes the genotype (genetic information) to predict phenotype (observable characteristic). Our phenotype is viral fitness defined as the capacity of the virus to replicate as compared to a standard. We construct models to predict viral fitness as a function of mutations from a standard wild-type virus. Data of this nature are difficult to analyze because there are potentially many more parameters than observations. We treat the problem as a regression problem with both a shrinkage component and a variable selection component. Using the scientific literature, we set up several informative priors that treat parameter subsets as exchangeable within subsets (the ESP prior). There are several scientific papers, each providing separate prior information. We then do prior model selection (PMS) to select the best prior. In this way we provide a solution for the genotype-phenotype problem. Viral fitness and genotypic data from 161 clade B HIV-1 viral sequences were obtained to predict the level of viral fitness as a function of mutation.

Statistical Models for Clustering of DNA Binding Motifs

◆ Shane Tyler Jensen, Harvard University; Jun S. Liu, Harvard University

Harvard University, One Oxford St., Dept. of Statistics, Cambridge, MA 02138

jensen@stat.harvard.edu

Key Words: *DNA binding motifs, Bayesian hierarchical modeling, nonparametric clustering, cross-species comparisons*

In order to elucidate detailed gene networks and mechanisms of cell regulation, biologists are interested in modeling the similarity between different motifs that represent the recognition sites of DNA binding proteins. The common structure between a set of motifs is investigated by a Bayesian nonparametric hierarchical model. This nonparametric model enables similar motifs to be partitioned together into clusters. This approach has substantial benefits over traditional clustering methods. The clustering inference gained can be built into more sophisticated search algorithms based on the posterior predictive distribution from our model in order to find binding motifs in large sequence datasets. Our strategy is also used to predict co-regulated genes on the basis of the clustering of motifs found in the regulatory region of genes that are conserved between different but closely related species.

Hierarchical Models for Detecting Positive Selection in DNA Sequences

◆ Raquel Prado, University of California, Santa Cruz; Daniel Merl, University of California, Santa Cruz

University of California, Santa Cruz, AMS, Baskin Engineering, Santa Cruz, CA 95064

raquel@ams.ucsc.edu

Key Words: *molecular adaptation, hierarchical GLMs, Bayesian modeling*

We present a model-based approach for detecting molecular adaptation in a large number of DNA protein coding sequences that are closely related in evolutionary time and for which very little or no phylogenetic information is available. Our statistical modeling is based on a class of hierarchical generalized linear models. A Bayesian approach that requires the implementation of efficient computational methods for parameter estimation is followed. The proposed class of models is motivated by the study of several DNA sequences encoding malaria antigens taken in various geographical locations in Asia, Africa, and South America. The models allow the incorporation of information that might be relevant to infer the pattern of substitutions in the sequences, such as geographical location of the sequences or pairwise evolutionary distances, if available. We assess the predictive performance of the new models via simulation studies for different kinds of data. We compare the results obtained with the new methodology to those obtained using traditional methods for detecting positive selection.

A Bayesian Approach for Identifying Multiple Regulatory Motifs Using Reversible Jump Markov Chain Monte Carlo

◆ Sierra M. Li, University of Washington; Jon C. Wakefield, University of Washington; Steve G. Self, Fred Hutchinson Cancer Research Center

sierrali@u.washington.edu

Key Words: *Bayesian inference, transcription factor binding site (motif), Dirichlet distribution, product multinomial, Metropolis algorithm, reversible jump Markov chain Monte Carlo*

Identification of transcription factor binding sites (regulatory motifs) plays an important role in gene discovery and characterization. Many methods have been proposed for motif detection, but none provide the flexibility to find repeated instances of multiple motif patterns of different widths. We have developed a unified Bayesian approach to simultaneous identification of multiple regulatory patterns based on a motif-specific probabilistic model. Unlike common approaches, our method does not require each motif pattern to occur in every sequence. The number of unique motifs is fixed and then the width, base composition, and occurrences (number and location) are evaluated for each motif within given ranges. We evaluate motif starting positions and their position-specific nucleotide distributions conditional on the number of motif pattern, motif widths and motif occurrences. Reversible jump Markov chain is used to handle the dimension change in the complex hierarchical parameterization. We show simulation results with sensitivity and specificity assessment. Finally, we use this method to identify regulatory motifs associated with genes involved in cell cycle regulation of yeast.

199 Technical Advances in the Analysis of Social Sciences Data ▲

Social Statistics Section

Tuesday, August 10, 8:30 am–10:20 am

Causal Inference in Spatial Hierarchical Setting

◆ Natalya Verbitsky, University of Michigan; Stephen W. Raudenbush, University of Michigan

University of Michigan, 439 West Hall, Dept. of Statistics, 550 East University, Ann Arbor, MI 48109

natalya@umich.edu

Key Words: *causal inference, potential outcomes, spatial statistics, hierarchical models, violation of SUTVA, social networks*

Does drug A or drug B alleviate headache faster? Does community policing decrease the annual crime rate of a city? Causal questions arise in many facets of daily life. Rubin's causal framework is based on estimating the differences between individual's potential outcomes to treatment A and treatment B and on Cox's "no interference between units" assumption (Stable Unit Treatment Value, SUTVA). However, this assumption is no longer valid when causal questions arise in social networks or geographical settings, where the treatment received by one's neighbor may affect one's potential outcome. We propose an extension of Rubin's framework where the potential outcomes of any unit in the population depend of the treatment assignment of all units in the population. We define causal effects and discuss assumptions that make this framework tractable. Since the data are hierarchical in nature (e.g., individuals nested in neighborhoods) we propose a two-level hierarchical linear model with spatial dependence at level two.

Mixed Effect Models and Quasi-F: For Testing Several Conditions in Multiple Random Effect ANOVA Design

◆ Olivier Renaud, University of Geneva; Paolo Ghisletta, University of Geneva

University of Geneva, 40, Bd du Pont d'Arve, Geneva 4, 1211 Switzerland

Olivier.Renaud@pse.unige.ch

Key Words: *by item/by subject, language-as-fixed-effect fallacy, hierarchical model*

Exact tests for ANOVA exist merely for designs that are limited to one random effect, while multiple fixed effects are allowed. In the social sciences, the random factor is usually the subject. This is especially true in psychology, where these designs are called "repeated measure ANOVA." However, very frequently a second random factor is present, crossed with subject (items presented, such as images, words, etc.). The substantive interest mainly resides in the significance of the fix effect(s) and possible additional covariates of this complex design, after controlling for the random effects. We will review different methods for this problem. In particular, the so-called "by subject/by item" (or F1/F2) method, which is overwhelmingly used in psycholinguistics. The quasi-F method and the mixed effect models methodology will be discussed and compared. The actual level and power of the tests for the different approaches are quite different. Additional discussed issues concern handling of missing values, availability of contrast computations and possibility to have covariates that are linked to one of the random effects.

The Use of Internal Pilots to Derive Powerful and Cost-efficient Designs for Studies with Multilevel Data

◆ Mirjam Moerbeek, Utrecht University

Utrecht University, PO Box 80140, Utrecht, 3508 TC The Netherlands

m.moerbeek@fss.uu.nl

Key Words: *optimal sample sizes, sample size adjustment, power, multilevel model, intraclass correlation, cost function*

Many data in social science research have a multilevel structure with persons nested within groups. To design a cost-efficient study to test the effect of a predictor on the outcome at a desired power level, the optimal number of groups and group size need to be calculated. Optimal sample size formulae depend on the intraclass correlation coefficient (ICC) and variance of the predictor. An initial estimate of these may be obtained from similar studies in the past or from subject-matter knowledge. Incorrect estimates, however, may result in an under- or overpowered study. This presentation illustrates the use of internal pilots for studies with costs associated with sampling and measuring groups and persons within groups. At the design stage the sample sizes and associated costs are calculated based on an initial estimate of the ICC and variance of the predictor. At an interim time point in the study these two parameters are estimated and the sample sizes and costs are adjusted. All data, including those of the pilot study, are used in the final analysis. Designs with different interim time points are compared on basis of their Type I error rates, power levels, and median costs.

Causal Inferences in Finite Mixture-based Trajectory Models, with Application to the Impact of Gang Membership on Violent Delinquency

◆ Amelia Haviland, Carnegie Mellon University; Daniel Nagin, Carnegie Mellon University

Carnegie Mellon University, Heinz School of Public Policy, 5000 Forbes Ave., Pittsburgh, PA 15213

amelia@andrew.cmu.edu

Key Words: *causal inference, finite mixture models, propensity scores, kernel estimation, gangs and violence*

A developmental trajectory describes the course of a behavior over age or time. A turning point describes some event which alters that trajectory. We build upon recent advances in identifying trajectory groups based on finite mixture modeling to devise improved methods for drawing causal inferences about the impact of turning point events. Based on the work of Rosenbaum and Rubin, we draw connections between trajectory groups and both balancing scores and kernel estimation. This allows us to identify the empirical conditions under which trajectory probabilities qualify as balancing scores. Where the trajectory probabilities do qualify as balancing scores, we show how they can be used as weights in a kernel matching procedure to obtain treatment effects. We demonstrate this approach with an analysis of the impact of gang membership on violent delinquency in a large Canadian longitudinal dataset.

Case-mix Adjustment for Comparing State Substance Treatment Outcomes

◆ Gary Huang, Synectics for Management Decisions, Inc.; Sameena M. Salvucci, Synectics for Management Decisions, Inc.; Dhirendra Ghosh, Synectics for Management Decisions, Inc.; Ping Yu, Synectics for Management Decisions, Inc.; Sue Becker, Substance Abuse and Mental Health Services Administration

Synectics for Management Decisions, Inc., 1901 N. Moore St., Suite 900, Arlington, VA 22209

Garyh@smdi.com

Key Words: *length of stay, retention, discharge rate*

Case-mix adjustment is used in healthcare cost studies and program evaluation to assure comparability across providers. The procedure, however, has not been widely used in substance abuse treatment research. We use discharge data collected in the 2000 Treatment Episode Data Set (TEDS) to explore risk variables that are critical for case-mix adjustment in comparing treatment outcomes (length of stay and planned discharge) across states. In addition to sex and age, the case-mix adjustment models specify and test a number of measures of risk factors, including type of substance used, years of using the substance, and co-occurring disorders. The case-mix adjusted average length of stay and planned discharge rates will be presented state by state. The results of the study will contribute to knowledge about measures of risk factors and statistical techniques specifically relevant to the case-mix procedures used in studying substance abuse treatment outcomes across organizations at levels above the local provider level.

Parametric and Nonparametric Bayesian Modeling for Ecological Inference in 2 x 2 Tables: A Data Augmentation Approach

◆ Kosuke Imai, Princeton University; Ying Lu, Princeton University

Princeton University, Corwin Hall, Dept. of Politics, Princeton, NJ 08544 United States

kimai@Princeton.Edu

Key Words: *ecological regression, racial voting, Dirichlet process prior, nonparametric methods, aggregate data, missing data*

The ecological inference problem arises when one wishes to make inferences about individual behavior from aggregate level data. Such a situation is frequently encountered in the social sciences and epidemiology. We propose a general framework for ecological inference based on data augmentation. Ecological inference in 2 x 2 tables is formulated as a missing data problem where only the weighted sum of the two unknown variables is observed. Following the existing literature, our first approach uses a parametric model to impute the missing data. We show how this model can be extended to incorporate covariate information as well as any hierarchical structure that may exist in the data. To overcome the limitation of parametric assumptions, we also develop a nonparametric model, which is based on a mixture of Dirichlet processes. This nonparametric model offers the advantage of greater flexibility to estimate the distribution of the unknown parameters. Finally, through simulation studies and empirical illustrations, we evaluate the performance of the parametric and nonparametric models in the presence of various degrees of aggregation bias.

August Zillmer's Contribution to Stable Theory in Demography

◆ Peter O. Pflaumer, FH-Kempten

FH-Kempten, Postfach 1680, Kempten, 87406 Germany

peter.pflaumer@fh-kempten.de

Key Words: *demography, history, population statistics*

August Zillmer¥s early contribution to demographic analysis is described and appreciated. His population model is recalculated and graphically completed with EXCEL. His conclusions and results are discussed and compared with modern demographic methods. August Zillmer (1831-1893) was a German life insurance actuary in Berlin. He is credited for one of the first German textbooks of actuarial mathematics. His name is associated with the Zillmer method of calculating life insurance reserves. Virtually unknown is his contribution to demography. In 1863 he published a discrete population model, and produced several age distributions by a given life table and different population growth rates. He showed that the resulting age distributions will eventually become stable. Although the stable model in demography can be traced back to Euler in 1760, Zillmer's model is the first dynamic analysis of the influence of changes of population growth rates on the age distribution and population parameters such as the mean age.

200 Regular Contributed Posters

ENAR, General Methodology, Biometrics Section, Section on Statistics in Epidemiology, Section on Bayesian Statistical Science, Section on Health Policy Statistics, Section on Risk Analysis

Tuesday, August 10, 9:00 am–10:50 am

Thick-tailed and Heteroscedastic Linear Models for the Analysis of cDNA Microarray Data

◆ Juan P. Steibel, Michigan State University; Guilherme J.M. Rosa, Michigan State University; Robert J. Tempelman, Michigan State University

Michigan State University, 1261 Anthony Hall, East Lansing, MI 48824

steibelj@msu.edu

Key Words: *microarray, heteroscedasticity, thick tails, MCMC*

A number of model specifications were compared for the analysis of a cDNA microarray. The dataset consisted of a dye swap experiment involving two treatments and two arrays. A total of 76 genes were spotted four times and two genes were spotted 16 times. A Bayesian approach was used to fit the following models: (1) linear model with homoscedastic residuals across genes, (2) linear model with independent heteroscedastic residuals across genes, (3) linear model with exchangeable heteroscedastic residuals across genes, (4) a linear model with homoscedastic student-t distributed residuals, (5) linear model with student-t exchangeable residuals. This approach allowed not only the estimation of treatment effects for each gene, but also the posterior distribution for the ranking of genes according to these effects. Results were compared on the basis of predicted treatment effects, size of the credibility intervals and rankings of genes. All models yielded similar point estimates for treatment effects, but with varying width of the credibility intervals. The most discordant results were obtained between the independent heteroscedastic and homoscedastic residual specifications.

Empirical Bayesian Estimation of the Disease Transmission Probability in Multiple-vector-transfer Designs

◆ Christopher R. Bilder, University of Nebraska, Lincoln; Joshua M. Tebbs, Kansas State University

University of Nebraska, Lincoln, Dept. of Statistics, East Campus, Lincoln, NE 68583

chris@chrisbilder.com

Key Words: *group testing, insect vectors, plant pathology, plant disease, proportion, virus transmission*

Plant disease is responsible for major losses in agriculture throughout the world. Diseases are often spread by insect organisms that transmit a bacterium, virus, or other pathogen. To assess disease epidemics, plant pathologists often use multiple-vector-transfers. In such contexts, $s>1$ insect vectors are moved from an infected source to each of n test plants. The purpose here is to present new estimators for p, the probability of pathogen transmission for an individual vector, motivated from an empirical Bayesian approach. In studying point estimate properties, one of our proposed estimators consistently results in a smaller bias and mean squared error than the maximum likelihood estimator (MLE) as proposed by Thompson and Swallow. This bias reduction is frequently fivefold or more in optimal settings for the MLE. Furthermore, these estimators are easier to compute than the classical Bayes estimators proposed by Chaubey and Li, and Chick. Finally our newly proposed empirical credible intervals possess the desirable property that lower bound will never be negative.

Meta-analysis Combines Affymetrix Microarray Results Across Laboratories

◆ John R. Stevens, Purdue University; Rebecca W. Doerge, Purdue University

Purdue University, Dept. of Statistics, 150 N. University St., West Lafayette, IN 47907-2067

jrsteven@stat.purdue.edu

Key Words: *meta-analysis, microarray, Affymetrix*

With the use of microarray technology becoming more prevalent, it is now not unusual to find several different laboratories employing the same microarray technology to identify genes related to the same condition in the same species. Although the experimental specifics are similar, a different list of significant genes may result from the data analysis in each laboratory. We propose a statistically based meta-analytic approach to microarray analysis for the purpose of systematically combining these results from the different laboratories. This approach provides a clearer view of which genes are significantly related to the condition of interest while simultaneously allowing for differences between laboratories. Of particular interest is the widely used Affymetrix oligonucleotide array, the results of which are naturally suited to a meta-analysis. A simulation model based on the Affymetrix platform is developed to examine the adaptive nature of the meta-analytic approach and to illustrate the usefulness of such an approach in combining microarray results across laboratories.

Statistical Issues of Protein Microarray Analysis

◆ Lianbo Yu, Purdue University; Rebecca W. Doerge, Purdue University

Purdue University, Dept. of Statistics, 150 N. University St., West Lafayette, IN 47907-2067

yulianbo@stat.purdue.edu

Key Words: *protein microarray*

DNA microarrays measure the intensities of mRNA and are used to make inferences about corresponding gene transcription. Proteins, on the other hand, are the final products of genes, and perform various functions in cells. Protein microarrays are emerging technologies that allow the complex properties of proteins to be studied in a manner that is unique from DNA microarrays. Miller et al. (2003) designed an antibody microarray to identify potential biomarkers for the identification of prostate cancer. We employ analysis of variance model to normalize antibody microarray data and explain the sources of variation associated with this technology. Both hypotheses and multiple comparison procedures are constructed to find potential biomarkers of prostate cancer.

Nonlinear Models for the Relation between Cardiovascular Risk Factors and Beverage-type Effects: Is the Effect Due to Wine, Beer, or Spirits?

◆ Nav Randhawa, CHEO Research Institute

CHEO Research Institute, Chalmers Research Group, 401 Smyth Rd., Ottawa, ON K1H 8L1 Canada

nrandhawa@cheo.on.ca

Key Words: *nonlinear models, Box-Tidwell power-linear model, cardiovascular disease, differential beverage-type effects*

An inverse association between moderate alcohol consumption and coronary heart disease has been shown in many epidemiological studies. However, it is not clear whether any specific type of alcoholic drink has a particular benefit. Nonlinear models are proposed for investigating the association between differential beverage type effects and risk factors for coronary heart disease (CHD). The parametric nonlinear models discussed are similar to those considered in Ambler et al. (2003) based on Box and Tidwell's power-linear model. The nonparametric generalized additive models provide a contrast to these models. The risk factors high-density lipoprotein cholesterol (HDL), systolic blood pressure (SBP), and fibrinogen are considered using data from the Whitehall II study of British civil servants. The results indicate that a substantial benefit is derived from the overall consumption of alcohol rather than specific beverage types, although increasing wine consumption is associated with slightly more favorable levels of all three cardiac risk factors. Findings also suggest that alcohol consumption relates positively with HDL and negatively with SBP and fibrinogen.

Imprinting Detection by Extending a Regression-based QTL Method: An Extensive Simulation Study

◆ Lei Lei, University of Texas M.D. Anderson Cancer Center; Olga Gorlova, University of Texas MD Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, Dept of Biological Sciences, PO Box 43131, Lubbock, TX 79409-3131

llei@mdanderson.org

Key Words: *linkage analysis, imprinting, regression-based QTL method, simulation*

We extended Sham et al.'s regression-based quantitative-trait linkage analysis method by regressing overall, paternal and maternal IBD sharing on traits separately. We also extended the method to calculate an imprinting index statistics. We suggest using a panel of statistics to detect imprinting, which includes overall T, both parental T statistics and the D index. After an extensive simulation study, we found that when using empirical percentiles (1) the method is extremely powerful by showing clear imprinting patterns with correct Type I error rate for the nonimprinted IBD, although the test of imprinting is too conservative; (2) the method is not sensitive to misspecification of heritability; (3) with low genetic heritability, power of the method decreases dramatically but the panel still performs well; (4) misspecification of the mean of trait decreases the power dramatically; and (5) missing parental marker increases Type I error of both linkage and imprinting test and decreases the power of imprinting test. The method is applied to a data set with six body size related measures and 23 loci on chromosome seven for 255 nuclear families. Linkage and imprinting are both detected.

Bayesian Meta-analysis of ROC with Normal Mixtures

◆ Minje Sung, Duke University Medical Center; Alaattin Erkanli, Duke University Medical Center; Refik Soyer, George Washington University

Duke University Medical Center, DUMC 3454, Durham, NC 27710

msung@psych.duhs.duke.edu

Key Words: *Dirichlet mixture model, receiver operating characteristics (ROC), Bayesian hierarchical model, Markov chain Monte Carlo simulation*

We develop a semiparametric Bayesian approach to multivariate receiver operating characteristics (ROC) using mixtures of Dirichlet Process (DP) priors. We address two challenging issues: modeling nonstandard distribution of screen scores and combining results from several different such ROC studies. In medical testing the results of a screening test usually does not have a Gaussian or a symmetric distribution. If this is the case, using a normal distribution will lead to incorrect choices of the diagnostic cut-offs and unreliable estimates of prevalence of a disease. The DP mixtures would provide a robust way of modeling nonstandard distributions. Hierarchical modeling framework is proposed to obtain study-specific ROCs as well as combined ROC. The model is implemented with real data from child psychiatric studies using Markov chain Monte Carlo technique.

Intuitive Nonparametric Effect Size Index for Treatment Trials

◆ Laura Acion, University of Iowa; Scott Temple, University of Iowa College of Medicine; Stephan Arndt, University of Iowa

University of Iowa, Dept. of Biostatistics, College of Public Health, 1-192 MEB Psychiatry Research, Iowa City, IA 52242

laura-acion@uiowa.edu

Key Words: *effect size, treatment trials, Mann-Whitney U statistic, Kendall's tau*

Effect size indices (ESI) tell the magnitude of the difference between treatments and, ideally, should tell clinicians how likely their patients will benefit from the treatment. The currently used ESIs are expressed in statistical rather than clinically useful terms and may not give clinicians the appropriate information. We restrict our discussion to a study of two groups: one with n patients receiving treatment X and the other with m receiving treatment Y. Cohen's d (d = (Xmean - Ymean) / SD) is a well known ESI for continuous outcomes. There is some intuitive value to d, but measuring improvement in SDs is a statistical concept that may not help a clinician. How much improved is a half SD? A more intuitive and simple ESI is the probability that a patient receiving X does better than a randomly chosen patient on Y, i.e., p(X>Y), where larger values mean better outcomes. This probability is related to the Mann-Whitney U statistic and Kendall's tau. We define criteria for good ESI, propose p(X>Y) as an alternative index, and show the relation of p(X>Y) with d under the uniform distribution. We discuss advantages and disadvantages of both ESIs and illustrate with clinical data.

Estimation of Age-Specific Sensitivity and Sojourn Time in Breast Cancer Screening

◆ Xiuyu Cong, Rice University; Yu Shen, University of Texas M. D. Anderson Cancer Center

Rice University, 6100 Main St. MS-138, Statistics Dept., Houston, TX 77005

xcong@stat.rice.edu

Key Words: *age-specific estimation, screening sensitivity, sojourn time distribution, breast cancer screening*

The main objective of this research is to quantitatively model the effect of age at detection on screening sensitivity and sojourn time distribution in breast cancer screening trials. Building on the modeling framework of Shen and Zelen (1999), we proposed parametric models to associate the age at detection with screening sensitivity and sojourn time distribution and obtained the corresponding parameter estimates using the maximum likelihood methods. The AIC criterion was used in the process of model selection. Extensive simulation studies show that the estimators have reasonable accuracy and the model selection criterion works well. The proposed methods are illustrated using data from two large breast cancer screening trials. The results based on these two trials show that the screening sensitivity increases with age at detection.

Estimating Disease Prevalence Based on Data from Population-based Registries

◆ Limin Clegg, National Cancer Institute; Mitchell H. Gail, National Cancer Institute; Mourad Touzani, Mathematica Policy Research, Inc.

National Cancer Institute, NIH, 6116 Executive Blvd., MSC 8316, Bethesda, MD 20892-8316

lin_clegg@nih.gov

We propose a new "Poisson" method to estimate the variance for prevalence estimates obtained by the counting method described by Gail et al. (1999) and to construct a confidence interval for the prevalence. We evaluate both the Poisson procedure and the procedure based on the bootstrap proposed by Gail et al. in simulated samples generated by resampling real data. These studies show that both variance estimators usually perform well and yield coverages of confidence intervals at nominal levels. When the number of disease survivors is very small, however, confidence intervals based on Poisson method have supra-nominal coverage, whereas those based on the procedure of Gail et al. tend to have below nominal coverage. For these reasons we recommend the Poisson method, which also reduces the computational burden considerably.

Modeling the Outcome of the Alberta Cervical Cancer Screening Program's 2003 Result Letter Survey

◆ Jacqueline Gregory, Alberta Cancer Board; Zeva Mah, Alberta Cancer Board; Carolyn Pim, Alberta Cancer Board

Alberta Cancer Board, Calgary, AB Canada

jacquieg@cancerboard.ab.ca

Key Words: *cancer, screening, survey, modeling, logistic regression*

Few cervical cancer screening programs send correspondence directly to women. The purpose of this survey was to measure the perceived helpfulness of five types of correspondence from the Alberta Cervical Cancer Screening Program. The survey was a pre-tested 20-item RDD telephone survey for women aged 18-69. 1,000 women responded (62% response rate). 86% had had a Pap test less than three years ago. Women who had hysterectomies were excluded. Younger women ($p< 0.01$) would find it more helpful if the program sent them Pap test and cervical cancer screening information or overdue followup letters. Reminders were seen as more helpful by women who were: younger ($p< 0.01$), without a regular doctor ($p=0.05$), and without a recent Pap test ($p< 0.01$). Normal and abnormal result letters were considered more helpful to women who were: younger ($p=0.02$, $p< 0.01$), less educated ($p=0.01$, NS), without a regular health center ($p=0.02$, $p=0.02$), and without a recent Pap test ($p< 0.01$, $p< 0.01$). Women perceived program correspondence as being helpful.

Predictive Values for Combinations of Diagnostic Tests with Multiple Outcomes

◆ Natalie Khutoryansky, Amersham Health; John R. Ventre, Amersham Health; Simon Roitt, Amersham Health

Amersham Health, 101 Carnegie Center, Princeton, NJ 08540

natakhut@hotmail.com

Key Words: *diagnostic tests, truth standard, Bayesian method, predictive value, sensitivity, specificity*

In many diagnostic efficacy studies the disease status of each patient is determined using a truth standard. However, sometimes a perfect truth (gold) standard test does not exist or may be invasive or expensive. We consider an approach alternative to a gold standard by a combination of imperfect diagnostic tests. Many authors examined this approach in the case of dichotomous diagnostic tests. We address the case when diagnostic tests can have multiple outcomes (>2). For each sequence of outcomes of consecutive diagnostic tests the total probabilities of disease presence or absence (predictive values) are calculated using a Bayes formula that includes the likelihood ratio of outcomes. The subsequent test can be prescribed if one of the predictive values does not reach a pre-specified level. The outcome likelihood ratios for dichotomous tests can be calculated based on their sensitivity and specificity. However, in the case of multiple outcomes the calculations are more complicated and include probabilities of outcomes under conditions of disease presence or absence. This approach is implemented to calculate predictive values for selected combinations of imaging diagnostic tests.

Techniques for fMRI Visual Field Diagram Analysis

◆ Raymond G. Hoffmann, Medical College of Wisconsin

Medical College of Wisconsin, PO Box 26509, Milwaukee, WI 53226

hoffmann@mcw.edu

Key Words: *fMRI, spatial statistics, areal models, geostatistical models*

A visual field diagram is a 2D projection of the regions of the brain activated by a visual stimulus. The projection is a time-dependent areal map of the activation levels of the voxels obtained from an fMRI scan. The projection is in a circular region corresponding to the visual field usually used for assessing retina sensitivity. Analysis of the visual field diagram has usually resulted from dividing the circle into 24 radial regions of pi/4 radians and three concentric rings about the center of the diagram. Statistical analysis, at best, has resulted in treating these 24 regions as a 24-point repeated measures analysis without any spatial structure. The current study uses simulated data with known effect structure and actual fMRI spatial correlation to compare the effectiveness of (1) a linear repeated measures with no assumed structure, (2) a two-factor repeated measures with eight angles and three distances from the origin, (3) an areal spatial model for the 24 segments, and (4) a geostatistical model for the over 663 projected voxels.

3D Wavelet Denoising of SPECT Images

◆ Liansheng Tang, Southern Methodist University

Southern Methodist University, 5657 Amesbury Dr., Apt. 2008, Dallas, TX 75206

ltang@mail.smu.edu

Key Words: *SPECT, wavelet, SPM, adaptive shrinkage*

The quality of SPECT (single photon emission computed tomography) three-dimensional (3D) human brain images is adversely affected by the noise inherent in photon counts. To improve the signal-to-noise ratio, 3D wavelet denoising may be preferred to Gaussian smoothing. A discrete wavelet transform via Mallat's pyramid algorithm is performed on the original SPECT image and denoising and enhancement are obtained by thresholding of the wavelet coefficients in the transform domain using adaptive shrinkage. The inverse wavelet transform of the thresholded coefficients is used to reconstruct the images. We compare these results with those of Gaussian-smoothed images for detecting areas of activation using the SPM brain-imaging software.

A Software to Automate Statistical Parametric Maps Analysis for PET Studies

◆ Manlong Rao, SUNY, Stony Brook; Wei Zhu, SUNY, Stony Brook; Yeming Ma, National Institutes of Health; Nora D. Volkow, National Institutes of Health; Gene-Jack Wang, Brookhaven National Laboratory

SUNY, Stony Brook, Dept of Applied Math and Statistics, Stony Brook, NY 11794

manlong2000@yahoo.com

Key Words: *SPM, PET, SPECT, fMRI, biostatistics, software*

Statistical parametric maps (SPM) is a customized statistical software for the analysis of brain image data (PET, SPECT, fMRI) developed by the Wellcome Department of Neuroscience in 1991. The SPM software is a suite of MatLab functions and subroutines built in a customized Windows environment. Images are first spatially normalized into a standard brain space and smoothed. Subsequently, the General Linear Model is applied to describe the variability in the data. However, to accomplish the analysis in the SPM interactive interface, one must enter tens and sometimes hundreds of parameters manually which is tedious, labor-intensive and extremely error-prone. To avoid these errors and facilitate the SPM analysis for, often repetitive, analysis of PET images, we have developed the SPM batch-mode software to accomplish the normalization and basic statistical analysis automatically. Required SPM input parameters are stored in one input Excel file and the output analysis results are stored in another excel file. The SPM batch-mode software also enabled us to implement a novel resampling method for the brain activation volumetric studies.

An Empirical Bayes Approach to Interval-mapping of Expression Trait Loci

◆ Meng Chen, University of Wisconsin, Madison; Christina Kendziorski, University of Wisconsin, Madison; Hong Lan, University of Wisconsin, Madison; Alan Attie, University of Wisconsin, Madison

University of Wisconsin, Madison, 1210 West Dayton St., Dept. of Statistics, Madison, WI 53706

mengchen@wisc.edu

Key Words: *empirical Bayes, mixture model, microarray, QTL mapping, expression trait loci (ETL) mapping*

Genetic linkage analysis has traditionally focused on mapping loci that affect one or more phenotypes. With microarray technology, we can apply such analysis to global patterns of gene expression, treating messenger RNA transcript abundances as quantitative traits and associating levels of gene expression with genotyping data. A number of groups have recently tackled this problem by applying QTL-mapping ideas. However, in mapping transcript expression levels, there are usually thousands of traits to be considered simultaneously, which greatly inhibits the efficiency of traditional QTL-mapping techniques. A model-based systematic framework is needed. We have developed an empirical Bayes approach to enable interval-mapping of ETL based on mixing over both unknown genotype and unknown expression levels. This formulation introduces computational challenges beyond those of standard QTL-mapping or differential expression analysis. We present a solution and assess the performance of the proposed method using data from an F2 mouse intercross in a study of diabetes.

Hidden Markov Models and Their Applications in Bioinformatics

◆ Sujay Datta, Northern Michigan University

Northern Michigan University, Dept. of Mmtcs, Stats & Ctr Sci, 1107 New Science Bldg., 1401 Presque Isle Ave., Marquette, MI 49855

sdatta@nmu.edu

A hidden Markov model (HMM) is a generalization of Markov chain models that allows for more modeling flexibility. The essential difference between a regular Markov chain and a hidden Markov model is that for the latter, there is no longer a deterministic one-to-one correspondence between the true state of the process and its observable manifestation. The true underlying states constitute a Markov chain but at every state, we observe a manifestation "probabilistically chosen" from a fixed and time-independent collection of possible manifestations. Over the years, a number of algorithms have been developed to efficiently compute the likelihood of an observed sequence of manifestations (given the parameters of the model), to estimate the true underlying state-sequence that maximizes the likelihood of a given sequence of manifestations, and so forth. These developments have been motivated by the rapidly expanding horizon of applications for these models in areas such as statistical genomics, proteomics and bioinformatics. A HMM is one of the most statistically sophisticated way of representing protein domain family sequences. It is useful for computing multiple sequence alignments.

An Interactive Web-based Program for Estimating Cancer Risk

◆ Edith A. Zang, Institute for Cancer Prevention; Shuquan Chen, Institute for Cancer Prevention; Brian Pittman, Institute for Cancer Prevention

Institute for Cancer Prevention, 390 Fifth Ave., 3rd Floor, New York, NY 10018

ezang@ifcp.us

Key Words: *cancer risk, logistic regression models, risk estimation, web-based program*

An interactive module for estimating cancer risk through the Institute for Cancer Prevention's (IFCP) web site was developed to educate users on the preventable causes of cancer. Estimates of individual risks for lung, colorectal, prostate, and breast cancer are based on the users' responses to a set of questions. The risk estimates, presented as a score of 0 through 1 and through graphic displays, represent the conditional likelihood of developing cancer based on a set of risk factors that are significant in either a biological or a statistical sense. The score is calculated using a multiple logistic regression equation, for which the coefficients were obtained by fitting logistic regression models to large sets of data. Source data for the regression models included IFCP's long-term case-control study on lung cancer, a large case-control study on colorectal cancer, an ongoing IFCP study on prostate cancer, IFCP's ongoing Women's Intervention Nutritional Study on breast cancer, and SEER data. Users of the module are also provided with basic facts about these cancers, a description of the risk factors, methods used to estimate the risk, and appropriate references.

201 ASA Stat Bowl Session I

ASA, ENAR, WNAR, IMS, SSC

Tuesday, August 10, 10:30 am–12:20 pm

Game 1

◆ Trecia Kippola, Oklahoma State University; ◆ Sounak Chakraborty, University of Florida; ◆ Michael Joner, Virginia Polytechnic Institute and State University

elaine@amstat.org

Game 2

◆ Abhyuday Mandal, Georgia Institute of Technology; ◆ Siuli Mukhopadhyay, University of Florida; ◆ Landon Sego, Virginia Polytechnic Institute and State University; ◆ Jesse Frey, The Ohio State University

The Ohio State University, Dept. of Statistics, 1958 Neil Ave., Columbus, OH 43210

frey@stat.ohio-state.edu

Game 3

◆ Tracy Morris, Oklahoma State University; ◆ Christopher J. Rigdon, Southern Illinois University, Edwardsville; ◆ Victor M. Johnson, Pennsylvania State University; ◆ Christy Spinka, Texas A&M University

Texas A&M University, 3130 E. 29th St., Apt. #B3, Bryan, TX 77802-2728

cspinka@stat.tamu.edu

Game 4

◆ Zhiguang Qian, Georgia Institute of Technology; ◆ Nusrat Jahan, Mississippi State University; ◆ Yuping Wu, University of South Carolina; ◆ Scott H. Holan, Texas A&M University

Texas A&M University, 3130 E. 29th St., Apt. #B3, Bryan, TX 77802-2728

sholan@stat.tamu.edu

202 Bayesian Modeling of Dependence

Section on Statistical Computing, Section on Bayesian Statistical Science

Tuesday, August 10, 10:30 am–12:20 pm

A General Approach for Constructing Priors for Covariance Selection Models

◆ Robert Kohn, University of New South Wales

University of New South Wales, Faculty of Commerce and Economics, UNSW, Sydney, 2052 Australia

r.kohn@unsw.edu.au

Key Words: *constrained Wishart distribution, element selection, graphical models, multivariate analysis*

We consider the Bayesian estimation of an inverse covariance matrix from Gaussian data. Methods are given to construct priors for covariance selection models where the marginal distribution of the model size has a simple form. The priors have normalizing constants for each possible model size, rather than for each possible model, which gives a tractable number of normalizing constants that we estimate using Markov chain Monte Carlo methods and store offline. Our priors do not require the restriction that the corresponding graphical models are decomposable. We illustrate our method using two examples. The first has a constrained Wishart distribution conditionally on the covariance selection model. The second example uses the prior where the inverse covariance matrix is decomposed into scale and partial correlation terms.

Bayesian Modeling of Several Covariance Matrices

◆ Michael Daniels, University of Florida

University of Florida, Dept. of Statistics, 207 Griffin-Floyd Hall, Gainesville, FL 32611

mdaniels@stat.ufl.edu

We explore simultaneous modeling of several covariance matrices using the spectral (eigenvalue) decomposition and modified Cholesky decomposition. We first assume the "dependence" matrices to be constant and the diagonal matrices of eigenvalues/innovation variances to be nonconstant, then log-linear models for the latter are used to reduce the number of free parameters to a manageable size. We explore the propriety of the posterior in such models with improper priors. We then explore an approach to weaken the assumption of constant dependence matrices in an automated fashion and describe how to compute Bayes Factors to test this hypothesis of constancy. The procedures are applied to data from two longitudinal clinical trials.

Objective Bayesian Analysis of Gaussian Markov Random Fields

◆ Victor De Oliveira, University of Arkansas

University of Arkansas, Dept. of Mathematical Sciences, Fayetteville, AR 72701

vdo@uark.edu

Key Words: *Jeffrey's prior, reference prior, spatial statistics*

Gaussian Markov random fields (GMRF) are important families of distributions for the modeling of spatial data that have been widely used in different areas of spatial statistics. GMRFs have been used for the modeling of spatial data, both as models for the prior of latent processes/random effects, and as models for the sampling distribution of the observed data. We consider in this work the latter use of GMRFs. We present an objective Bayesian analysis for the parameters of some classes of GMRFs. Specifically, we compute explicit expressions for the Jeffrey's and reference priors, and derive in each case conditions for posterior propriety of the model parameters. By way of a simulation experiment we study the frequentist properties of the Bayesian inferences about the model parameters resulting from the use of these priors, such as frequentist coverage of 95% credible intervals and the mean squared estimation errors. We illustrate the use of the proposed GMRF model for studying the spatial variability of gross domestic product (GDP) over the period 1980-1995 in 138 territorial units of 11 European countries.

203 Modeling Customer Choice

▲ ▦

Business and Economics Statistics Section, Section on Statistics and Marketing, Section on Bayesian Statistical Science

Tuesday, August 10, 10:30 am–12:20 pm

The Valuation of Price Experimentation on Profits

◆ Alan Montgomery, Carnegie Mellon University; Ting Zhu, Carnegie Mellon University

Carnegie Mellon University, Graduate School of Industrial Admin., 5000 Forbes Ave., Pittsburgh, PA 15213-3815

alan.montgomery@cmu.edu

Key Words: *experimentation, pricing, profit maximization, functional uncertainty*

Price experimentation is often advocated as an appropriate methodology to learn about the relationship between quantity and price. The basic premise underlying experimentation is that by introducing artificial variation into prices, the relationship between price and quantity can be estimated more precisely. Unfortunately, this learning may incur direct cost to changing prices, and even more importantly, there is an opportunity cost from experimenting since short-term profits will be less than they could have been. Economic research on experimentation considers this problem under very strict modeling assumptions, which looks quite different from classical statistical approaches to experimentation. This research attempts to bridge this gap and answer some of many unresolved practical questions for guiding managers to determine how much price experimentation is beneficial, when is an experiment warranted, and how should it be implemented optimally.

Using Check-all-that-apply Surveys to Support Product Development

◆ Mark Beltramo, General Motors Corporation; R. Jean Ruth, General Motors Corporation

General Motors Corporation, GM R&D Center, MC 480-106-359, 30500 Mound Rd., Warren, MI 48090-9055

mark.beltramo@gm.com

Key Words: *Markov chain Monte Carlo, censored data, consumer surveys, hierarchical Bayes*

Information for tracking purchasing behavior and consumers' product experiences is often collected using "check all that apply" survey batteries. Examples include sources of information, reasons for making (or not making) a purchase, and important product attributes. In product development, we are often interested in both the frequency with which each item is checked and the importance of each item to consumers. While estimating frequencies for the items is trivial, estimating each item's importance can be problematic if the list is long and the percentage of consumers checking any given item is relatively small. In addition, the importance of each item varies among consumers, so it is preferable to estimate the distribution of importance placed on each item. We describe an approach to this problem that uses two additional pieces of information from each consumer: which item is most important and a rating of how important it is (measured on an appropriate scale). We present an MCMC sampler based on a hierarchical Bayes formulation of the problem. The formulation allows the estimated importance of each item to borrow strength from data on different, but related, products.

Representing Brand and Category Personality with Circumplex Models

◆ Peter Lenk, University of Michigan; Rajeev Batra, University of Michigan Business School; Michel Wedel, University of Michigan Business School

University of Michigan, 701 Tappan, Ann Arbor, MI 48109-1217

plenk@umich.edu

Key Words: *Bayesian Inference, latent variables models, marketing applications*

Consumers often anthropomorphize brands by associating them with personality characteristics. This provides brands with durable and distinctive traits that differentiate brands within a category. The personality of a brand and its category are often confounded, caused by the transfer of perceptions across these two domains. This paper develops a hierarchical Bayes model for brand and category personality that separates the two. The model is based on a circumplex correlation structure for personality items. The circumplex model consists of two bipolar factors where the loadings are restricted to the unit circle. Personality items are ordered according to their angles. The model is applied to a survey of three categories (cars, jeans, and magazines) with 10 brands in each category. The empirical study detects substantial transference of category personality to the brands within the category and finds a strong relation between the subjects' feeling towards a brand and its perceived personality.

204 Successful Statistical Collaborations ⌘

Council of Chapters, Section on Government Statistics, Social Statistics Section, Section on Quality and Productivity, Section on Statistical Education, Section on Survey Research Methods

Tuesday, August 10, 10:30 am–12:20 pm

The Collaborative Nature of Statistics as a Discipline

◆ Larry D. Haugh, University of Vermont

University of Vermont, Statistics Program, 16 Colchester Ave., Burlington, VT 05401-1455

Larry.Haugh@uvm.edu

Key Words: *collaboration, statistical consulting, multidisciplinary studies*

Statistical methods are highly valued by others for the insights they can provide, as well as the power or efficiency they can bring to the process of learning for work, research and even recreation; and we wish even more people had that appreciation! The JSM 2004 theme is "Statistics as a Unified Discipline," and we value the commonality or similarity of the methods we statisticians use, despite the diversity of organizations that we work for and the wide range of types of data and application areas that we work with. If statistics is the discipline or science of methods for data collection, analysis, and inference, then is collaboration necessarily a core feature of statistics? Is ours a discipline that should naturally reach out to others? How do statistical consulting and collaboration differ? Are statisticians special in some way among professionals or scientists in general for the depth of their concern with such questions? A perspective on these questions will be given, based on my and others' experiences as statisticians.

Collaboration at the Joint Program in Survey Methodology

◆ J. Michael Brick, Westat; Roger Tourangeau, University of Maryland

Westat, 1650 Research Blvd., Rockville, MD 20850

mikebrick@westat.com

Key Words: *training, interdisciplinary, survey methods*

One of the main reasons for establishing the Joint Program in Survey Methodology (JPSM) at the University of Maryland in 1993 was to facilitate the education of survey statisticians and methodologists to support the needs of the U.S. Federal Statistical System. To accomplish this goal required an unprecedented collaboration of organizations, disciplines, and researchers because the theory and practice of survey methodology is so diverse. The blending of the statistical and social science disciplines was a key focus in the initial stages of development of curriculum at the JPSM. Another important ingredient was the development of a partnership of organizations that concentrated on academic objectives, the efficient conduct of surveys, and the federal policy needs. We discuss how the collaborations at these various levels have contributed to a broader perspective on survey methodology in theory and application. We examine some of the contributions to survey methodology that have resulted from the collaborations in the first 10 years of the program and provide some thoughts on what might be expected in the future.

The Still Unfinished Journey from Single to Multiple Imputation

◆ Fritz J. Scheuren, NORC, University of Chicago

NORC, University of Chicago, 1402 Ruffner Rd., Alex., VA 22302

scheuren@aol.com

Key Words: *variance estimation, missingness models, hot-deck, complex samples*

Hard applied problems sometimes lead to more than frustration. They can also lead, as in this case, to new theory and new opportunities for partial solutions. Single hot-deck imputation of income by the Census Bureau was shown to be inadequate for variance estimation in the early 1970s. What to do about it was the issue. The Wiley book on multiple imputation by Don Rubin was the eventual result. This paper tells the story of the early stages in that still continuing journey, reflecting on those experiences from the perspective of what has happened and not happened since then.

205 Design and Analysis for 96-well Plates ▲

Section on Physical and Engineering Sciences, Section on Quality and Productivity

Tuesday, August 10, 10:30 am–12:20 pm

Statistical Considerations for 96-well Plates

◆ Douglas Hawkins, University of Minnesota

University of Minnesota, School of Statistics, 313 Ford Hall, 224 Church St. SE, Minneapolis, MN 55455

doug@stat.umn.edu

Key Words: *spatial effects, clinical chemistry*

The 8x12 96-well plate has long been the workhorse for clinical chemistry assays. Now it, and larger variants, are playing a vital role in high throughput screening and factorial experiments. There can be substantial location effects in the different wells of the plate. Traditional layouts in which for example calibrators and controls always go into the same plate location confound these location effects with treatment differences. Methods from agricultural experimentation where similar (though generally nonsystematic) effects occur can be used profitably in a more thoughtful allocation to the wells.

Robotics to Support Experimental Design for Microtiter Plates

◆ John S. Snider, Beckman Coulter, Inc.; Tim Sherrill, Beckman Coulter, Inc.; Nate Terpstra, Beckman Coulter, Inc.; Jeremy Smith, Beckman Coulter, Inc.; Jacob Schlitz, Beckman Coulter, Inc.

Beckman Coulter, Inc., 7451 Winton Dr., Indianapolis, IN 46268

John.Snider@SAGIAN.com

Key Words: *DOE, automation, microtiter plates, liquid handling, assay development*

Researchers in life sciences are challenged to design robust assays where numerous factors influencing performance must be investigated. DOE allows researchers to study multiple factors simultaneously with relatively few tests. Although the potential benefits of DOE for assay development have been acknowledged for many years, it has been avoided or limited in scope because of the lack of tools to cope with the cumbersome liquid handling. We have developed practical tools to make DOE for assay development feasible. Large, complex designs are now possible, and we give examples. Our commercially available products use robotic liquid-handling workstations to automate experimentation. Liquid handling is no longer too tedious and time-consuming to perform on a large scale. Our products assign treatment conditions to microtiter plates, collect information needed to perform experiments, and create automated liquid handling methods. Experimental results can be collected via automated microtiter plate readers and reported in several formats for analysis. The above process, required tools, technical challenges related to creating the tools, and benefits of this approach are described.

Highly Automated DOE for Complex Biological Experiments

Perry D. Haaland, Becton Dickinson Technologies; ◆ Matthew Mitchell, BD Technologies; Dylan Wilson, BD Technologies; Bryce Chaney, BD Technologies; Charles Schmitt, BD Technologies; Mohammad Heidaran, BD Technologies

BD Technologies, PO Box 12016, Research Triangle Park, NC 27709

matthew_mitchell@bd.com

Key Words: *DOE, automated analysis, cell differentiation, data-mining, biology*

We have created an automated, high-throughput system for conducting complex biological experiments. This system is currently in use to develop environments for cells that lead to controlled differentiation and/or proliferation. Experiments are typically carried out in multiwell plates and often include as many as 30 experimental factors. Because of the complexity of the designs, they are impossible to pipette directly by a human being. Consequently, we provided many tools that automate the creation of experiments. All information about the design is stored in a database, which facilitates automated statistical analysis and data-mining across multiple experiments. This system has allowed us to attack significant biological problems for which complexity has limited the success of previous efforts.

206 Decennial Census Coverage Measurement Program

Social Statistics Section, Section on Government Statistics, Section on Survey Research Methods

Tuesday, August 10, 10:30 am–12:20 pm

Measuring the Quality of Census Counts—The Australian Experience

◆ Paul S. Williams, Australian Bureau of Statistics

Australian Bureau of Statistics, Census, Demography, and Geography Branch, Locked Bag 10, Beconnen ACT 2616, Australia

paul.williams@abs.gov.au

Key Words: *estimate, census, coverage error*

Since 1971, Australia has employed a variety of techniques to measure the quality of census counts and has applied these measures to adjust the census counts to derive the official population. The Australian Bureau of Statistics used the census coverage as one of the important measures to judge the quality of census counts. The agency conducted a post-evaluation survey to estimate the total population which became the basis to determine the census coverage error. The survey used dual system estimator to produce total population estimate. The paper will outline these techniques, their application, and areas for future research.

Under-enumeration—the Census of England and Wales

◆ James J. Brown, University of Southampton; Ian Diamond, Polaris House

University of Southampton, Highfield, Southampton, Hants SO17 1BJ England

jjb1@socsci.soton.ac.uk

This paper provides a review of strategies for estimating the under-enumeration in the Census of England and Wales over the past three censuses. The paper first reviews the use of dependent coverage surveys which attempted to estimate under-enumeration and to judge that quality of the census. This approach was judged successful in 1981, but "'failed" to measure the extent of under-enumeration in 1991. Second, it discusses the deterministic demographic approach used in 1991 to adjust the Census for under-enumeration. Third, it describes the integrated post-enumeration survey/demographic adjustment model used in 2001, which estimated under-enumeration for each of 400 local authorities and which was then used to impute under-enumeration to very low levels of aggregate. The paper finishes with a broad evaluation of the 2001 approach and looks forward to possibilities for 2011.

Plans for Measuring Coverage of the 2010 U.S. Census

◆ Donna L. Kostanich, U.S. Census Bureau; David C. Whitford, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Suitland, MD 20746-7600

donna.l.kostanich@census.gov

Key Words: *measurement error, coverage error*

Coverage evaluation of the two previous United States censuses focused mainly on providing estimates of the net undercount. The initial coverage estimates of Census 2000, based on the Accuracy and Coverage Evaluation Survey, were unacceptable because the survey failed to measure significant numbers of erroneous census enumerations. Research efforts aimed at correcting for these measurement errors have provided groundbreaking information on the source and magnitude of coverage error components. These prompted us to expand the coverage measurement goals for 2010 beyond estimating net error to include estimating gross error and individual components of error. Based on our 2000 coverage measurement experience and our new goals for 2010 coverage measurement, we have identified major research issues. We believe these are the most important challenges we face to improve coverage measurement in 2010. This paper briefly summarizes the goals and objectives for coverage measurement in 2010, discusses the motivation for the major research topics and also provides a broad overview of its anticipated design features.

207 Outcome-dependent Sampling in Epidemiologic Studies

▲ ⊞

Section on Statistics in Epidemiology, Business and Economics Statistics Section, WNAR

Tuesday, August 10, 10:30 am–12:20 pm

Outcome Dependent Sampling, Biased Sampling, and Truncated Sampling: Connections and Applications in Epidemiologic Studies

◆ Jing Qin, Memorial Sloan-Kettering Cancer Center

Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology and Biostatistics, 307 East 63rd St., 3rd Floor, New York, NY 10021

qinj@biosta.mskcc.org

Key Words: *outcome-dependent sampling, biased sampling, truncated sampling*

Outcome dependent sampling is a sampling strategy that implies that the probability of being sampled depends directly on the value of the outcome variable or dependent vairable. It is a cost-effective way to enhance the efficiency in epidemiologic studies. Biased sampling is a convenient or economic sampling technique in the collection of positive-valued or lifetime data. Biased sampling occurs when healthier patients are more likely to participate a study. In survival analysis, failure time data sometimes are ascertained retrospectively subject to a certain criterion. Truncation sampling occurs when only those who have experienced the terminating event are qualified to be recruited. In this talk we will point out the connection between those sampling schemes and likelihood decompositions. Different approaches are discussed, including conditional inference, pseudo-likelihood based inference and empirical likelihood method.

Outcome-dependent Sampling in Genetic Studies of Dichotomous and Quantitative Traits

◆ Clarice R. Weinberg, National Institute of Environmental Health Sciences; Emily Kistner, National Institute of Environmental Health Sciences; David M. Umbach, National Institute of Environmental Health Sciences; Claire Infante-Rivard, McGill University

National Institute of Environmental Health Sciences, MD A3-03, PO Box 12233, Research Triangle Park, NC 27709

weinberg@niehs.nih.gov

Key Words: *transmission disequilibrium test, quantitative traits, logistic regression, genetics, family studies, linkage disequilibrium*

Studies that aim to identify alleles that are in linkage disequilibrium with a locus related to a dichotomous trait, such as a disease, or a quantitative trait, e.g., adiposity, can make use of genotype data from individuals and their parents. If the condition under study is dichotomous, a log-linear model can be employed to detect apparent departures from Mendelian transmission of a variant allele from parents to affected offspring. Such a design is outcome-dependent, because no population controls are required at all. The design offers robustness against bias due to genetic population stratification. When the outcome is quantitative, an extension of the log-linear model can be used, where the quantitative trait value is treated as a predictor in a polytomous logistic model. Selection of individuals with extreme values of the measured trait can enhance power of such a study. We provide results of simulations that reveal the enhancement that is possible with outcome-dependent sampling, and apply the method to data from a case-control study of intra-uterine growth retardation.

Semiparametric Likelihood for Two-stage ODS Scheme with Auxiliary Covariates and Random Effects

◆ Jinhong You, University of North Carolina; Haibo Zhou, University of North Carolina, Chapel Hill

University of North Carolina, Dept. of Biostatistics, Chapel Hill, NC 27599-7420

jyou@bios.unc.edu

Key Words: *auxiliary information, centered effect, two-stage sampling, outcome-dependent, semiparametric likelihood, asymptotic normality*

With the field of epidemiology expanding and evolving an increasing number of studies are conducted using the ODS design with a "continuous" outcome. Another complexity in practical studies often involves the cluster- or center-effects of the study subjects. In this paper we consider a two-stage outcome-dependent design in which center effects are involved, and the selection of the second-stage subsample depends on both the continuous outcome and a continuous auxiliary covariate of the true exposure variable. By nonparametrically estimating the conditional distribution of the exposure variable we propose a semiparametric likelihood method to handle the center-effects and the outcome-dependent nature of the subsampling components. The resulted estimator of the interested parameters is consistent and asymptotically normal. Simulation results show that the proposed estimator performs better than the competing methods that either use only the second-stage data or ignore the center effects. We illustrate the proposed method with a hypothetical dataset from an environmental epidemiologic study.

Design and Efficiency Issues in Two-phase Studies

◆ Jerald F. Lawless, University of Waterloo; Donald L. McLeish, University of Waterloo; Yang Zhao, University of Waterloo

University of Waterloo, Dept. of Statistics and Act. Sci., Waterloo, ON N2L 3G1 Canada

jlawless@uwaterloo.ca

Key Words: *incomplete data, response selection, restricted sampling*

Two-phase epidemiological studies in which the objective is to relate certain response variables to covariates are often conducted. At phase 1, variables that are easily or inexpensively measured are observed on a group of individuals. At phase 2 a random subset of these individuals is selected, and more expensive measurements taken. When the probability an individual is chosen for phase 2 depends on their phase 1 measurements, the combined sampling procedure is often outcome-dependent. The efficiency and cost of such studies depends on the sizes of the phase 1 and 2 samples, the phase 2 inclusion probabilities, and the true relationships between variables in the study. This talk examines these issues and the choice of good designs in "expensive covariate" and "expensive response" settings.

208 Methods of Approximation for Spatial Data

ENAR, Section on Statistics and the Environment, Section on Statistics in Epidemiology, Section on Bayesian Statistical Science, Section on Statistical Computing, WNAR

Tuesday, August 10, 10:30 am–12:20 pm

Methods to Approximate a Spatial Likelihood

◆ Montserrat Fuentes, North Carolina State University

North Carolina State University, Box 8203, NCSU, Raleigh, NC 27695

fuentes@stat.ncsu.edu

Key Words: *Whittle's likelihood, spectral domain, satellite data, irregular lattices, covariance, Fourier transform*

Likelihood approaches for large, irregularly spaced spatial datasets are often very difficult, if not infeasible, to use due to computational limitations. Even when we can assume normality, exact calculations of the likelihood for a Gaussian spatial process observed at n locations requires $O(n^3)$ operations. We present a version of Whittle's approximation to the Gaussian log likelihood for spatial lattices with missing values. This method requires $O(n^{2\log\_2} n)$ operations and does not involve calculating determinants. If the usual biased sample covariance is used in this approximated likelihood method, the estimated covariance parameters are efficient only in one dimension due to the edge effect. To remove this edge effect, we introduce a new data taper, a circular taper, that gives more tapering to the corner observations. We present simulations and theoretical results to show the benefits and the performance of proposed methods. We apply this likelihood method to sea surface temperature satellite data in the presence of clouds.

Intrinsic Autoregressions and the de Wijs Process

◆ Debashis Mondal, University of Washington, Seattle; Julian Besag, University of Washington

debashis@stat.washington.edu

Key Words: *intrinsic autoregressions, de Wijs process, approximation*

Intrinsic auto regressions are Markov random fields that have a rigorous interpretation as generalizations of random walks to higher dimensions. They have been used as prior distributions to represent spatial variation at different scales, from microscopy to agriculture to epidemiology to astronomy. Homogeneous Gaussian intrinsic auto regressions on rectangular arrays are nonstationary limiting versions of stationary conditional auto regressions. The most used first-order scheme induces well-defined distributions for all contrasts. We evaluate the associated variogram and show that this provides a close approximation to the variogram of the corresponding conformally invariant integrated de Wijs process in the plane, provided the cells of the array are not too elongated. This suggests that the more easily used auto regressions display an approximate invariance to changes of scale, which has also been our experience in agricultural applications.

Space-time Parameter Estimation Using Approximate Likelihoods

◆ Jonathan R. Stroud, Wharton School, University of Pennsylvania

Wharton School, University of Pennsylvania, Statistics Dept., 3730 Locust Walk, Suite 400, Philadephia, PA 19102

stroud@wharton.upenn.edu

Key Words: *likelihood approximation, Kalman filter, parameter estimation, model selection*

In a recent paper by Stein, Chi and Welty (2003), new methods were proposed for approximating the likelihood for a spatial Gaussian process. We extend their method to the spatio-temporal setting and show how this can be used to perform parameter estimation and model selection in a large class of linear state-space models. We illustrate the method through an example involving remote sensing.

Model Approximation in Complex Bayesian Analyses

◆ L. Mark Berliner, The Ohio State University

The Ohio State University, 1958 Neil Ave., Dept. of Statistics, Columbus, OH 43210

mb@stat.ohio-state.edu

Key Words: *environmental statistics, hierarchical modeling, Markov chain Monte Carlo, physical-statistical modeling*

While the development of modern MCMC techniques enable Bayesian analyses using realistic, hierarchical models in complex problems, some compromises in specifications of models are often necessary. The types of compromises made vary widely depending on the problem. The assessment of the impacts of model approximation is difficult, but important. Some basic notions of these issues are discussed and illustrated in selected environmental problems involving real data.

209 Proposed Guidelines for Teaching Statistics: K-12 and College

Section on Statistical Education, Advisory Cmte on Teacher Enhancement, Section on Quality and Productivity

Tuesday, August 10, 10:30 am–12:20 pm

Proposed Guidelines for Teaching Statistics: K-12 and College

◆ Christine A. Franklin, University of Georgia; ◆ Robin Lock, St. Lawrence University; ◆ James M. Landwehr, Avaya Labs Research; ◆ Patricia S. Wilson, University of Georgia

University of Georgia, 105 Aderhold Hall, Athens, GA 30602

pwilson@coe.uga.edu

Key Words: *guidelines, assessment, education*

This session reports on a recent project, funded by an ASA Strategic Initiatives grant, to develop ASA-endorsed guidelines for statistics education. The process of developing these guidelines involved two focus groups that had structured discussions over email during the past year. One group developed guidelines for K-12 instruction and the other group developed guidelines for the introductory college course. Two panelists will present reports and two will discuss and respond to the reports.

210 Comparing the Current Population Survey to the Census 2000 Using the CPS/Census Match

▲ ⊞

Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section

Tuesday, August 10, 10:30 am–12:20 pm

Issues in Weighting the Current Population Survey-Census Matched File

◆ Andrew A. Zbikowski, U.S. Census Bureau

U.S. Census Bureau, Washington, DC 20233-0001

andrew.a.zbikowski@census.gov

Key Words: *surveys, estimation, weighting*

The Census Bureau matched data from the monthly Current Population Survey (CPS) to Census 2000 data to create household-level and person-level files. These files have potential for a variety of research projects, including comparison of CPS and Census data on labor force status, race and ethnicity, and income. After the matching and prior to analysis of the data, we needed to weight the data to reflect national population estimates. This weighting presented several challenges. We used four months of CPS data, from February to May 2000, clustered around the date of Census 2000. A single monthly CPS sample includes eight rotation groups of sample households. We had 14 unique rotation groups in our matched file, since six of eight rotation groups are the same from month to month. Also, some CPS households and people did not match to Census 2000 data, but analysts wanted the option of including them in their analyses. We used basic CPS weighting as the platform from which to adjust the estimates. This paper discusses some matching results and unique aspects of weighting for the matched CPS-Census 2000 data.

The Effects of Employing Alternative Nonresponse Adjustment Cells on Labor Force Estimates in the Current Population Survey

◆ Kevin A. Farber, U.S. Census Bureau

U.S. Census Bureau, 103 A Walnut Ave., Baltimore, MD 21206

kevin.a.farber@census.gov

Key Words: *sample weighting adjustments, nonresponse, Current Population Survey*

To adjust for the effects of unit nonresponse bias on survey estimates, it has become common for surveys to use weighting adjustments to the respondent's sample weights. Regardless of which weighting adjustment method is employed, assumptions are made concerning the nonrespondents. These assumptions are difficult to verify, since data for nonrespondents are usually not available. Thus, one is not sure if using a weighting adjustment for nonresponse improves sample estimates. The focus of this paper is the nonresponse sample-weighting adjustment used in the Current Population Survey (CPS). Alternative weighting classes will be selected for performing this adjustment based on logistic regression models. Assumptions about nonrespondents will be investigated by matching CPS records to Census 2000. Mean square errors for major labor force estimates will be compared to judge the effects of using the different nonresponse cells.

Examining Differences in Labor Force Status in the Current Population Survey and the Census 2000

◆ Tamara S. Adams, U.S. Census Bureau

U.S. Census Bureau, 15007 Nebraska Lane, Bowie, MD 20716

tamara.s.adams@census.gov

Key Words: *CPS, labor force*

The Current Population Survey (CPS) is the source of the national monthly unemployment rate released by the Bureau of Labor Statistics. The Census 2000 also collects labor force status on the long form for approximately 1/6 of the population. However, the CPS and the Census 2000 produced different unemployment rates. The Census Bureau matched the CPS and the Census 2000, including those housing units receiving the Census 2000 long form. We will compare the labor force status as assigned by the CPS and the labor force status as assigned by the Census 2000 and various covariates that may be related to measurement error. Finally, we will model the differences in assignment of labor force status to attempt to determine the major covariates in the difference in the unemployment rate between the two data collection methods.

Using Census Match Data to Evaluate Models of Survey Nonresponse

◆ John Dixon, Bureau of Labor Statistics

Bureau of Labor Statistics, 10308 Meredith Ave., Kensington, MD 20895

dixon_j@bls.gov

Key Words: *survey, nonresponse, census*

Several methods are used to model nonresponse in surveys; panel information, item nonresponse, last 5%, and area characteristics. This study compares the methods to model nonresponse using match data from the census as a criterion. Recommendations based on the comparisons and limitations based on imperfect matching will be discussed.

211 Topics in Response Surface Designs I

Section on Quality and Productivity

Tuesday, August 10, 10:30 am–12:20 pm

Optimal Blocking with One Hard-to-change Factor

◆ Frank T. Anbari, George Washington University

George Washington University, Dept. of Management Science, 2115 G St., NW, Washington, DC 20052

anbarif@gwu.edu

Key Words: *hard-to-change factors, expensive-to-change factors, prediction variance, G-efficiency, cost function, fractional factorial designs*

This paper shows how to design, run and analyze 2**k and 2**k-p experimental designs for process improvement when there are hard-to-change or expensive-to-change factors. The paper discusses the different ways of running these experiments, and gives practical recommendations. It shows how to block designs to get a small prediction variance and low cost. It presents an algorithm to allow the selection of efficient confounding relation(s) in 2**k and 2**k-p designs, presents methods for calculating the prediction variance and G-efficiency when there are hard-to-change factors, and tabulates the results. The approach is extended to fractional factorial designs, and results are tabulated. The paper discusses the cost of running the experiment. It shows practical block patterns which have higher G-efficiencies and higher cost efficiencies than designs run in a random order. The paper shows that with one hard-to-change factor, a blocked design is both less expensive to run, because it requires fewer resets of the hard-to-change factor, and more precise as it gives a lower variance of prediction than a completely randomized design.

Blocking Strategies for Factorial Experiments with Hard-to-change Factors

◆ Derek F. Webb, Bemidji State University

Bemidji State University, HS-342, Box 23, 1500 Birchmont Dr., NE, Bemidji, MN 56601

dwebb@bemidjistate.edu

Key Words: *blocking, randomization, G-efficiency, hard-to-change factor*

This paper presents blocking strategies for factorial experiments that contain more than one hard-to-change factor. The strategies are based on an algorithm that generates efficient confounding relations. We present the blocking structures and give some examples that result in small prediction variances, are more cost-effective to run than randomized designs, and have higher G-efficiencies than completely randomized designs. We tabulate confounding relations, blocking structures and G-efficiencies for various factorial designs with more than one hard-to-change factor. We also, through example, discuss the differences between blocking with HTC factors and traditional blocking discussed in most books of experimental design.

Using a Genetic Algorithm to Generate Mixture Designs

◆ John J. Borkowski, Montana State University

Montana State University, Dept. of Mathematical Sciences, Bozeman, MT 59717

jobo@math.montana.edu

Key Words: *genetic algorithm, maximin criterion, D-optimality, response surface design*

A genetic algorithm (GA) is an evolutionary search strategy based on simplified rules of biological population genetics and theories of evolution. A GA maintains a population of candidate solutions for a problem, and then selects those candidates most fit to solve the problem. After the selection process, the most fit candidate solutions are combined and/or altered by reproduction operators to produce new solutions for the next generation. The process continues, with each generation evolving more fit solutions until an acceptable solution is evolved. A GA is developed to generate p-component mixture designs allowing for lower and upper bounds for each mixture component. Several criteria, including D-optimality and the maximin criteria, will be used as the fitness criterion in the GA.

Three-dimensional Variance Dispersion Graph for Designs with Split-plot Structures

◆ Li Liang, Virginia Polytechnic Institute and State University; Christine Anderson-Cook, Virginia Polytechnic Institute and State University; Timothy J. Robinson, University of Wyoming

Virginia Polytechnic Institute and State University, 823 Orchard St. Apt. 8, Blacksburg, VA 24060

lliang@vt.edu

Key Words: *restricted randomization, design assessment, scaled prediction variance, central composite design*

Complete randomization in many industrial and agricultural experiments is very difficult due to constraints in time and/or cost or existence of one or more hard-to-control factors. In these situations, restrictions on randomization lead to split-plot design (SPD), allowing certain factor levels to be randomly applied to the whole-plot units and remaining factor levels randomly to the sub-plot units. Random errors in whole-plot units as well as in subplot units are introduced due to the two separate randomizations in the experiment. The resulting compound symmetric error structure affects not only estimation and inference but also the choice of design. We present the three-dimensional variance dispersion graph (3D VDG) as a tool to evaluate the prediction ability of split-plot designs. Examples of central composite design (CCD) with split-plot structures are studied. From the distribution of scaled prediction variance (SPV) in 3D VDG, the optimal design is obtained by changing the factorial levels of CCD.

Split Plots: Modeling the Process Mean and Variance from a CCD

◆ Scott Kowalski, Minitab Inc.; Geoff Vining, Virginia Polytechnic Institute and State University; Douglas Montgomery, Arizona State University

skowalski@minitab.com

Key Words: *split plot, response surface methodology*

The response surface methodology (RSM) framework for designed experiments has been widely adopted in practice. Most of the literature on RSM is confined to completely randomized experimental designs (CRDs). Unfortunately, industrial experimenters often encounter situations where some experimental factors are hard to change or where there is significant discrepancy in the size of some experimental units. For both of these situations, the experimental units for some factors are observational units for other factors, which leads to split plot designs. An important question within industrial statistics is how to find operating conditions that achieve some goal for the mean of a characteristic of interest while simultaneously minimizing the characteristics process variance. This talk first establishes how one can modify the common central composite design to accommodate efficiently a split-plot structure. It then modifies this design to allow the estimation of separate models for the characteristic's mean and variance. The appropriate analysis of the experimental results will be discussed.

212 Advances and Applications in Causal Inference

Section on Bayesian Statistical Science

Tuesday, August 10, 10:30 am–12:20 pm

Propensity Score Estimation with Boosted Regression

◆ Daniel F. McCaffrey, RAND Corporation; Greg Ridgeway, RAND Corporation; Andrew Morral, RAND Corporation

RAND Corporation, 201 North Craig St., Suite 201, Pittsburgh, PA 15090

daniel_mccaffrey@rand.org

Key Words: *causal effect, bagging, generalized boosting, regression trees*

Causal effect modeling with naturalistic, rather than experimental, data is challenging. Variations in treatment exposure may be confounded with differences in pretreatment characteristics or treatment selection factors. Propensity score methods can theoretically eliminate these confounds for all observed covariates, but accurate estimation of propensity scores is impeded by large numbers of covariates, uncertain functional forms for their associations with treatment selection, and other problems. This paper demonstrates that boosting, a modern statistical technique, can overcome many of these obstacles. We use generalized boosting regression, which maximizes a likelihood function rather than minimizing prediction error. We report the mean-squared error of estimated treatment resulting from alternative stopping rules for the boosting algorithm including methods based on out-of-bag cross-validation and minimization of the imbalance of covariate distributions between groups. We illustrate this approach with a study of adolescent probationers in substance abuse treatment programs. Propensity score weights estimated using boosting eliminated all large pre-treatment group differences.

Bayesian Nonparametrics and Causal Inference

◆ Jennifer Hill, Columbia University; Robert E. McCulloch, University of Chicago

Columbia University, SIPA, 420 W. 118th St., 740 IAB, New York, NY 10027

jh1030@columbia.edu

Key Words: *Bayesian, nonparametrics, propensity scores, causal inference*

Propensity score approaches are becoming more popular for causal estimation in observational studies. Traditional estimation methods (logistic and probit regression), however, may not represent the best models for estimating propensity scores. In addition, no method for reliably estimating standard errors from resulting treatment effect estimates has been identified. A Bayesian nonparametric approach to propensity score estimation that addresses both of these issues will be illustrated.

Improving upon Intention-to-treat Analysis When Clinical Trials Become Open-label

◆ Samantha R. Cook, Harvard University; Donald B. Rubin, Harvard University

Harvard University, 1 Oxford St., Dept. of Statistics, Cambridge, MA 02138

cook@stat.harvard.edu

Key Words: *historical controls, missing data, multiple imputation, hierarchical models*

In a recent FDA trial, the drug under investigation became commercially available before the end of the trial. Patients in the trial therefore had the option of going off trial protocol and obtaining the commercially available active therapy. When patients randomized to placebo switch to commercially available therapy, they cease to be controls in the usual sense: all measurements after a placebo control switches to active therapy are missing. We propose a method to impute placebo controls' missing outcomes, as if they had stayed on placebo. This involves, first, fitting a Bayesian hierarchical regression model to data from untreated historical patients; second, incorporating information learned from the historical patients into a similar model for placebo controls; and third, using this model and observed on-protocol data to impute missing values for placebo controls who switched to active therapy. Using multiple imputation will allow for improved estimation of the treatment effect and a more powerful test of the drug's effectiveness.

Analysis of Treatment Response Data without the Joint Distribution of Counterfactuals

◆ Siddhartha Chib, Washington University, St. Louis

Washington University, St. Louis, Olin School of Business, CB 1133, 1 Brookings Dr., St. Louis, MO 63130

chib@wustl.edu

Key Words: *Bayesian inference, instrumental variable, Metropolis-Hastings algorithm, observational data, potential outcomes, treatment effect*

This talk is concerned with the Bayesian analysis of treatment response data when the (categorical) treatment is nonrandomly assigned but a binary instrumental variable is available, say from the design of the problem. We show that unlike previous work it is possible to analyze this problem without the modeling of the joint distribution of the counterfactuals. The modeling and prior-posterior analysis are developed for both binary and ordinal treatments under weak distributional assumptions. Estimation of these models is by Markov chain Monte Carlo methods, after rewriting the models in line with the framework of Albert and Chib (1993), and the comparison of the various models is by marginal likelihoods and Bayes factors, estimated by the method of Chib (1995). We discuss inferences for the treatment effects, outlining ways in which one can calculate the effect of the instrument on the outcome, both from the intrinsic structure of the model and the output of the MCMC simulations. Several illustrations of the methods are provided.

213 Retail and Trade Surveys: Accomplishments and Challenges

SSC, Section on Government Statistics, Section on Survey Research Methods

Tuesday, August 10, 10:30 am–12:20 pm

Expansion and Improvement of the U.S. Census Bureau's Retail and Wholesale Programs

◆ Ruth E. Detlefsen, U.S. Census Bureau

U.S. Census Bureau, Service Sector Statistics Division, Room 2641-3., Mail Stop 6500, Washington, DC 20233

Ruth.E.Detlefsen@census.gov

Key Words: *sample selection, time series, data collection integration*

The U.S. Census Bureau has conducted retail and wholesale trade censuses and sample surveys for many years. In keeping with the Bureau's long-standing commitment to produce quality data that is relevant, responsive to customer needs, and attentive to minimizing respondent burden, several projects are underway which will further expand and improve these retail and wholesale programs. This paper will discuss motivation, research, and implementation for: (1) expansion of the annual wholesale trade survey, (2) methodological and procedural improvements for the 2006 business sample revision, (3) restatement of historical time series to reflect the 2002 North American Industry Classification System, and (4) expansion of data collection integration for the 2007 Economic Census.

A New Estimator for the UK Retail Sales Inquiry

◆ Peter Brodie, Office for National Statistics

Office for National Statistics, Room D140, Government Bldgs., Cardiff Rd., Newport, NP10 8XG UK

pete.brodie@ons.gov.uk

Key Words: *retail sales, matched sample, Drift, trading day*

Estimation for the UK Retail Sales Inquiry/Index up until October 2003 was based on movements of the matched sample. The index was compared to a base value in January 1995 and consequently prone to "drift." This paper describes the analysis carried out by Methodology Group of ONS to investigate the accuracy of the index. The paper describes the issues that were addressed (choice of estimator, optimization of sample allocation, outlier detection and treatment) as well as the remaining issues (trading day weights), which were not, and outlines the recommendations made.

Monitoring Survey Processes of the Canadian Monthly Wholesale and Retail Trade Survey

◆ Mark J. Majkowski, Statistics Canada; Hélène Bérard, Statistics Canada; Catherine Dufour, Statistics Canada; Susie Fortier, Statistics Canada

Statistics Canada, Tunney's Pasture, 11-K, R.H. Coats Bldg., Ottawa, ON K1A 0T6 Canada

mark.majkowski@statcan.ca

Key Words: *monitoring, diagnostic tools, level estimates, trend estimates*

The Canadian Monthly Wholesale and Retail Trade Survey (MWRTS) produces monthly estimates for sales and inventories at various province and industry levels. The MWRTS has recently been redesigned in part to provide estimates for the new North American Industry Classification System (NAICS) and to take full advantage of administrative data from the Goods and Services Tax program. The redesign also addressed the need to maintain the estimate quality, reduce cost and respondent burden, update computer systems, and harmonize annual survey concepts. As part of the redesign, different tools were developed to ensure proper monitoring of survey steps. The first type of 'diagnostic' tool aims to assess the functionality of the modules in each survey step while the second type involves monthly descriptive statistics, such as number of live units in the sample, imputation method used, and recurrent top contributors. These statistics are studied longitudinally to detect changes that affect the estimates. Finally, a third set of diagnostic tools is used for analysis of level and trend estimates. The monitoring tools used monthly for this survey will be described.

The Redesigned Canadian Monthly Wholesale and Retail Trade Survey: A Post-mortem of the Implementation

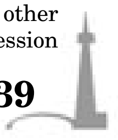
◆ Julie Trepanier, Statistics Canada

Statistics Canada, Tunney's Pasture, R.H. Coats Bldg., 11th floor, Ottawa, ON K1A 0T6 Canada

julie.trepanier@statcan.ca

Key Words: *Retail Trade Survey, Wholesale Trade Survey, parallel run, evaluation*

The development of the new MWRTS took place between 2000 and 2003. The sampling design was revised to incorporate the North American Industry Classification System, to expand the survey coverage to nonemployer businesses, to take advantage of the Goods and Services Tax data and to incorporate innovative solutions to maintain the quality of the survey over time. Some other changes to the survey were the inclusion of a multiple regression



imputation, the development of an outlier detection and treatment strategy to deal with influential units at estimation, and the addition of many monitoring tools at different stages of the survey. The first official release of estimates from the new MWRTS is for the May 2004 reference month. Prior to that, an extensive testing took place for a period of five months, when both the old and new surveys were in production. What were the major highlights during the testing? What were the main challenges when the new estimates were produced? How and why did they differ from the old survey estimates? What have we learned from this experience? This paper will provide answers to these questions while giving an overview of the old and new MWRTS.

214 Linear and Nonlinear Mean Prediction of Time Series

Business and Economics Statistics Section

Tuesday, August 10, 10:30 am–12:20 pm

A Linear Nonstationary Mean Predictor for Seasonally Adjusted Series

◆ Alessandra Luati, University of Bologna; Estela B. Dagum, University of Bologna

University of Bologna, via Belle Arti 41, Dept. of Statistics, Bologna, 40126 Italy

luati@stat.unibo.it

Key Words: *smoothing, system of weights, false turning points, gain function, 13-term Henderson filter*

Dagum developed in 1996 a nonlinear nonparametric estimator of the nonstationary mean (trend-cycle) of monthly seasonally adjusted time series characterized by several points of maxima and minima. Relative to the widely applied classical 13-term Henderson filter (H13), this estimator reduces significantly both the size of the revisions to most recent estimates and the number of false turning points (unwanted ripples) with the good property of identifying true turning points with very short time delays. The purpose of this study is to develop a linear approximation to the Dagum nonlinear filter. A linear approximation offers the advantage of preserving the relationship between original and adjusted data when dealing with seasonally adjusted aggregates, it is of easy application for it does not required complex software, and its statistical properties can be confronted with those of other potentially competitive linear filters. This new linear filter is compared to the classical H13 by means of the classical spectral analysis techniques and statistical measures of bias, variance, and mean square error based on the system of weights.

Modeling Dependence in Vector Stochastic Processes

◆ Paul Dagum, Rapt, Inc.

Rapt, Inc., 625 2nd St., 2nd Floor, San Francisco, CA 94107

paul.dagum@vecra.com

Key Words: *non-normal, stochastic process, copula, nonlinear*

The dependencies in a vector stochastic process arise from temporal dependencies within processes and dependencies across processes. Both sets of dependencies may express nonlinear switching behavior that is regime-dependent. Joint models for these processes typically invoke assumptions about linearity or normality of distributions. Among these are Brownian motion models that are instances of a Gaussian process. We propose a copula approach for separately modeling the marginal distributions and dependencies of vector stochastic processes. We allow for separate copulas to model the contemporaneous dependencies and the temporal dependencies including finite mixture copulas with categorical variables that permit model switching. By decoupling the specification of the marginal distribution from the dependence model, inferences on dependencies are robust to the specification of the marginal distribution. The approach admits non-normal and nonparametric marginal distributions. We discuss economic and business applications.

On the Model-based Interpretation of Trend-cycle Filters

◆ Tommaso Proietti, Dipartimento di Scienze Statistiche

Dipartimento di Scienze Statistiche, Via Treppo 18, Udine, 33100 Italy

proietti@dss.uniud.it

Key Words: *signal extraction, Kalman filter and smoother, revisions, reliability*

The paper is concerned with a class of trend cycle filters, encompassing the Hodrick-Prescott and rational square wave filters, that are derived using the Wiener-Kolmogorov signal extraction theory under maintained models that prove unrealistic in applied time series analysis. As the maintained model is misspecified, inference about the unobserved components, and in particular their posterior mean and variance, are not delivered by the Kalman filter and smoother or the Wiener-Kolmogorov filter for the maintained model. The paper proposes a model-based framework according to which the same class of filters is adapted to the particular time series under investigation; the idea rests on an “embedding principle” by which any linear time series can be decomposed into orthogonal components with a given representation; the resulting decomposition “objectifies” the filters, giving them autonomous justification; nevertheless, it is the interaction with the series and its time series properties than enables the components to be estimated and their uncertainty assessed. This embedding principle guarantees that correct inferences are standard and are provided in finite samples.

Deterministic and Stochastic Linear and Nonlinear Mean Prediction of Time Series

◆ Simone Giannerini, University of Bologna

University of Bologna, Via delle belle arti 41, Statistics Dept., Bologna, 40126 Italy

giannerini@stat.unibo.it

Key Words: *time series, detrend, prediction, inference, deterministic trend, stochastic trend*

Two common approaches are usually followed to render mean stationary an observed time series. The first one consists in removing a deterministic polynomial trend, so that the adjusted series is said to be trend stationary. The second approach makes use of the regular difference operator of a low order; in this instance, the detrended series is stochastic stationary. We compare the results of applying both the methods to a plethora of series whose trend component can be either deterministic or stochastic. The effect of every choice on the subsequent analysis to be performed on the series will be investigated. In particular, we will assess to what extent the use of the wrong detrender affects the true generating structure of the series under question.

215 Small-area Estimation III ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Tuesday, August 10, 10:30 am–12:20 pm

Hierarchical Bayes Inference to Small-area Census Undercoverage Estimation

◆ Yong You, Statistics Canada; Peter Dick, Statistics Canada

Statistics Canada, Household Survey Methods Division, R.H. Coats Bldg. - 16th Floor, Ottawa, ON K1A 0T6 Canada

yongyou@statcan.ca

Key Words: *benchmarking, census undercoverage, hierarchical Bayes, sampling variance, small area*

We will review and outline some hierarchical Bayes (HB) small-area models, the problem of sampling variances, and the calibration procedures for Canadian small-area census undercoverage estimation. The small-area models include the well-known Fay-Herriot model and the unmatched area level models. In general, sampling variances are assumed to be known in these models. Usually smoothed estimators of the sampling variances are used and then treated as known. We also consider the case of estimated sampling variances and a full HB model is constructed for the direct survey estimators and the direct estimators of sampling variances. Finally the HB estimates are calibrated to agree with national or provincial totals using the benchmarking HB method of You, Rao, and Dick (2002, 2004). Posterior me an squared error (PMSE) will be used as a measure of uncertainty for the benchmarked HB estimators. Examples in census undercoverage estimation will be provided.

Small-area Estimates of Diabetes and Smoking Prevalence for North Carolina Counties: 1996-2002 Behavioral Risk Factor Surveillance System-based Estimates

◆ Akhil Vaish, RTI International; Neeraja Sathe, RTI International; Ralph Folsom, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

avaish@rti.org

Key Words: *survey-weighted hierarchical Bayes, pseudo-score functions, MCMC NC county-level SAEs for diabetes and smoking*

We present North Carolina (NC) county-level Small Area Estimates (SAEs) for diabetes and smoking. The pooled 1996-2002 NC-Behavioral Risk Factor Surveillance System (BRFSS) data for diabetes and smoking is used to produce SAEs for all the 100 counties in NC. Estimates of change in diabetes and smoking prevalence rates from 1996-1999 to 2000-2002 are also obtained. The Survey Weighted Hierarchical Bayes (SWHB) methodology of Folsom, Shah, and Vaish (1999) is used to fit mixed logistic regression models. Multivariate age group specific random effects defined at two levels of hierarchy along with personal level covariates (race and gender) as well as county-level predictors obtained from various national data sources are used in the SWHB modeling. The SWHB methodology has several benefits over commercially available software such as MLwiN and BUGS. The SWHB methodology allows the use of personal-level predictors in the modeling and as a result SWHB-SAEs are internally consistent and more efficient than the Fay-Herriot (1979) aggregate level solution. The SWHB-SAEs are also design consistent and self-benchmarked to the robust survey-weighted estimates for large sample areas.

A Bayesian Prediction for Undecided Voters

◆ Jai W. Choi, Centers for Disease Control and Prevention; Balgobin Nandram, Worcester Polytechnic Institute

Centers for Disease Control and Prevention, 3311 Toledo Rd., Hyattsville, MD 20782

jwc7@cdc.gov

Key Words: *correlation, missing data, multinomial-Dirichlet, nonignorable, sampling-based, two-way categorical table*

Typically data from election polls or heath surveys are presented in two-way categorical tables. For example, in the 1998 Ohio polls for governor, the first category represents the candidates, Fisher, Talft, other, and undecided, and the second category represents the current status, i.e., likely to vote, not likely to vote and undecided, of the voters. There is a substantial number of undecided voters. Bayesian prediction allows to allocate the undecided voters to the three candidates and permits modeling different patterns of missingness, and a multinomial-Dirichlet model estimates the cell probabilities which can help to predict the winner. In 1998, Ohio Polls were conducted in January, April, and October. We extend our models to accommodate a correlation structure across these three polls. Markov chain Monte Carlo methods are used to fit the models, and a simulation study is used to assess the predictive power of the models.

Small-area Estimation Errors in SAIPE Using GLMM vs. FH Models

◆ Eric V. Slud, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., FB-4, Washintgon DC, DC 20233-9100

evs@math.umd.edu

Key Words: *small-area estimation, SAIPE, Fay-Herriot model, GLMM models, loss functions*

The Small Area Income and Poverty Estimation (SAIPE) program produces county-level estimates of child-poverty rates and counts of poor school-aged children, using census and administrative-records (CPS, IRS, food stamp) data. The author has for several years studied the properties—and discrepancies from both the “internal” standard of CPS direct estimates and the “external” standard of census-based estimates—of small-area estimators in this setting, based upon competing versions of the aggregate-level Fay-Herriot (FH) model used in SAIPE production or based upon Generalized Linear Mixed Model (GLMM) unit-level models. Recently, Slud (2003) has compared and assessed the fit of various FH models tracked over several years of SAIPE data since the program's original model choices and evaluations were made, and has found that augmentations to the SAIPE production model are needed for the 2000 data. In this paper, an overall summary and assessment is made of the relative strengths and weaknesses of the FH versus GLMM county child poverty rate estimators.

216 Joint Social, Government, and Survey Research Methods Statistics Sections Student Competition Winners

Social Statistics Section, Section on Government Statistics, Survey Methodology Editorial Board, Section on Survey Research Methods

Tuesday, August 10, 10:30 am–12:20 pm

Jointly Modeling Latent Trajectories and a Subsequent Outcome Variable: A Bayesian Approach

◆ Sujata Patil, University of Michigan; Trivellore E. Raghunathan, University of Michigan

University of Michigan, 2901 Baxter Rd., Ann Arbor, MI 48109-2150

spatil@umich.edu

Key Words: *Bayes, longitudinal data, trajectory*

We propose a general class of models that treat latent trajectory variables summarized in stage 1 of the model as predictors of a subsequent outcome in a second stage. This general class of models allows for a variety of trajectory shapes and can incorporate outcomes that come from the exponential family of distributions. We propose a fully Bayesian approach to fit this model. The Bayesian approach and two competing approaches are used to study trajectories of adolescent alcohol use as predictors of motor vehicle offenses incurred during young adulthood. Additionally, we describe results from a simulation study that show that Bayes estimates have less bias and root mean square error and have better confidence than two competing approaches.

Imputation by Propensity Matching

◆ Murthy N. Mittinty, University of Canterbury; Easaw Chacko, University of Canterbury

University of Canterbury, Dept. of Mathematics and Statistics, Ilam Rd., Christchurch, P.Bag 4800 New Zealand

nmi13@student.canterbury.ac.nz

Key Words: *propensity, matching, nearest neighbour, dissimilarity, missing at random, imputation*

Missing data is a common phenomenon. Many survey organizations use single imputation methods such as nearest neighbour imputation (NNI) to deal with missing data. Advantages of NNI are it makes use of covariate information and the point estimators obtained from the data imputed by NNI are less biased. In multivariate covariate situation finding NN is complicated because every variable needs to be matched. We investigate the use of propensity matching, introduced by Rosenbaum and Rubin (1983) for observational studies, to find the donor when data is multivariate. NNI by propensity score (NNPS) is investigated using simulations with data missing at random, both linear and convex. We use NNPS as it assures that the conditional distribution of the covariates given the propensity score is same for respondents and nonrespondents. We compare NNPS with NNI by dissimilarity matrix (NNDM) given by Murthy et al. (2003). We use NNDM as it preserves the marginal distributions. The results indicate that estimates from data with imputation by NNPS are often similar to those of NNDM and it reduces the curse of dimensionality.

Using Double Sampling to Correct Gross Flows for Misclassification Error: Moment-based Inference vs. Likelihood-based Inference

◆ Nikolaos Tzavidis, University of Southampton

University of Southampton, S3RI, Highfield, Southampton, SO17 1BJ UK

ntzav@socsci.soton.ac.uk

Key Words: *measurement error, missing data, panel surveys, EM algorithm, re-interview surveys, missing information principle*

Longitudinal surveys provide a key source of information for analyzing dynamic phenomena. Typical examples of longitudinal data are gross flows between a finite number of states. Sample surveys are, however, affected by nonsampling errors. We investigate the use of double sampling for correcting discrete longitudinal data for misclassification error. In a double sampling context, we assume that along with the main measurement device, which is affected by misclassification error, we can use a secondary measurement device, which is free of error but more expensive to apply. Inference is based on combining information from both measurement devices. Traditional moment-based inference is reviewed and contrasted, under alternative double sampling schemes, with a proposed likelihood-based method that works by simultaneously modeling the true transition process and the measurement error process within the context of a missing data problem. Variance estimation, under both approaches, is discussed. Monte Carlo simulation experiments indicate that the proposed likelihood-based method offers significant gains in efficiency over the traditional moment-based method.

A Bayesian Hierarchical Modeling Approach to Age-period-cohort Analyses of Repeated Cross-section Survey Data

◆ Yang Yang, Duke University

Duke University, Dept. of Sociology and Institute of Statistics and Decision Sciences, 268 Soc/Psych Bldg., Box 90088, Durham, NC 27708

yangy@soc.duke.edu

Key Words: *age-period-cohort analysis, repeated cross-sections, hierarchical model, cross-classified random effects model, restricted maximum likelihood, Bayesian inference*

It is well known that the identification problem created by the exact linear dependency of age (A), period (P), and cohort (C) presents a challenge to conventional age-period-cohort (APC) analysis in demography and social science research. This study developed a new approach, namely, Bayesian hierarchical APC models of micro datasets in the form of repeated cross-section survey designs. It examined the impact of small sample sizes of birth cohorts and time periods and unbalanced data on statistical inference based on the usual REML-EB estimation through Monte Carlo simulations. A full Bayesian analysis using Gibbs sampling and MCMC estimation was conducted to strengthen the inference by accounting for this extra uncertainty associated with parameter estimates when the numbers of higher-level units are small. For a substantive illustration, it applied cross-classified random effects models to the vocabulary data from the General Social Survey from 1974 to 2000. The results shed new light on the controversy on APC trends in verbal ability in the U.S. for the past three decades.

Imputing Missing Income Data and Weighting Data with Imputed Income

◆ Bidisha Mandal, The Ohio State University; Elizabeth A. Stasny, The Ohio State University

The Ohio State University, 101 Curl Dr., 215 Jones Tower, Columbus, OH 43210

mandal.7@osu.edu

Key Words: *hot-deck imputation, mean imputation, regression imputation, RDD survey*

Income is an important demographic variable in social science research. But income data is more difficult to obtain compared to other demographic information, since it is typically considered to be private and personal. The focus of this research is to try several imputation schemes to account for missing income data and then to use income for weighting the survey data. Specifically, hot-deck imputation, mean imputation within adjustment cells, and regression imputation for missing income values are explored. Then weighting using each imputation scheme is carried out, and the resulting estimates for a key survey variable are compared. Weighting is also carried out without income as one of the weighting variables. The purpose is to see how successful imputation has been in replacing missing income data, and how good imputed income is as a weighting variable.

217 Bayesian Modeling in Wildlife, Fisheries, Environment Research

Section on Bayesian Statistical Science

Tuesday, August 10, 10:30 am–12:20 pm

Prediction of Riverine Sediment Categories from a Multinomial Reverse Dirichlet Spatial Mixture Model

◆ Kyoji Furukawa, Iowa State University; Mark Kaiser, Iowa State University

Iowa State University, 315 A Snedecor Hall, Ames, IA 50011

kyoji@iastate.edu

Key Words: *spatial mixture model, Bayes prediction, riverine sediment*

The characteristics of sediment structure in the river bed are known to be important predictors for many biotic responses that are indicators of the overall productivity of an aquatic ecosystem. A statistical model that allows prediction of sediment types is expected to enhance our ability to understand and model the levels of such responses. For this purpose, we develop a spatial mixture model that consists of a conditionally independent multinomial obseration process mixed over a conditionally specified spatial process that possesses Dirichlet-like behavior. The sediment type at an unobserved location is predicted by a Bayes predictor under the squared error loss and the predictive ability of the proposed model is assessed by cross-validation.

Bayesian Estimation in Spatially Inhomogeneous Pairwise Interacting Point Processes

◆ Matthew Bognar, University of Iowa

University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242-1409

mbognar@stat.uiowa.edu

Key Words: *spatial point pattern, reversible jump MCMC, importance sampling, pairwise interacting point process, Bayesian inference, Metropolis-Hastings algorithm*

Spatial point patterns are frequently modeled with pairwise interacting point processes. However, inference is complicated by the presence of an intractable function of the parameters in the likelihood. While various frequentist inferential techniques have been developed, Bognar and Cowles (2004) suggested the use of Bayesian methodology for inference. Because the Metropolis-Hastings acceptance probability contains a ratio of two likelihoods evaluated at differing parameter values, the resulting intractable ratio complicates the required application of MCMC. Within each iteration of the sampler, Bognar and Cowles estimate this intractable ratio with importance sampling. While the aforementioned inferential techniques have assumed spatial homogeneity in the density of points, we will examine a generalization of the Bayesian inferential framework of Bognar and Cowles which allows the modeling of spatial inhomogeneity in the density of points (the literature is absent of such an extension in a frequentist framework). We conclude with an analysis of a glacial drumlin dataset using an inhomogeneous Strauss point process.

Spatial Variation in Soil Nitrate and Carbon Levels in Metropolitan Phoenix and the Surrounding Deserts

◆ Jacob J. Oleson, Arizona State University

Arizona State University, Dept. of Mathematics & Statistics, Tempe, AZ 85287-1804

Jacob.Oleson@asu.edu

Key Words: *linear mixed model, MCMC, LTER, spatial correlation, Bayesian*

Cities provide unique opportunities for integrating humans into ecology. Using data from an integrated inventory of metropolitan Phoenix, Arizona, we explore the contribution of human-related variables to explaining observed variation in soil nitrate-N and total carbon across the city, agricultural fields, surrounding desert, and mixed regions. The spatial relationships for carbon and soil nitrate may be different in each of these regions. In estimating their levels, we take into account differing spatial patterns in each region and use additional covariate information as well. Soil nitrate shows markedly greater spatial autocorrelation in the desert regions, while the carbon levels show varying amounts of spatial relationships in the different regions.

Bias Adjustment in Bayesian Estimation of Bird Nest Age-specific Survival Rates

◆ Jing Cao, University of Missouri, Columbia; Chong He, University of Missouri, Columbia

University of Missouri, Columbia, 510 High St., Apt. 326, Columbia, MO 65201

caoj@stat.missouri.edu

Key Words: *age-specific survival, Bayesian estimation, simulation, iteration, hierarchical prior*

The populations of many North American landbirds are showing signs of declining. Gathering information on breeding productivity allows early detection of unhealthy populations and helps develop good habitat-management practices. We use simulation to study the performance of the Bayesian model for age-specific nest survival rates with irregular visits. We find that the estimates are satisfactory except for the age-one survival rate. Usually the more days skipped between two visits, the more serious the underestimation of the age-one survival rate. We investigated the problem and developed three approaches to adjust for the underestimation. The simulation results show that the three approaches can significantly improve the estimation of the survival rate at age one. Finally, the adjustment is illustrated through an analysis of the Missouri Dickcissel dataset.

218 Statistical Graphics II

Section on Statistical Graphics

Tuesday, August 10, 10:30 am–12:20 pm

A Mixture-based Approach to Latent Class Discovery

◆ Jeffrey L. Solka, Naval Surface Warfare Center Dahlgren Division

Naval Surface Warfare Center Dahlgren Division, NSWCDD, Code B10, 17320 Dahlgren Rd., Dahlgren, VA 22448-5100

solkajl@nswc.navy.mil

Key Words: *Gaussian, mixtures, latent, class, discriminant, boundary*

Given a set of pre-categorized n-dimensional observations, one is often interested in discovering hidden categories within the pre-defined categories. This paper will present new work that takes a Gaussian-mixtures-based approach to the problem. This approach uses Gaussian mixture models to subcategorize observations based on their relationship to the mixture-induced descriminant boundary. The methodology will be illustrated using artificial and real world datasets. Real-world datasets will include examples from the gene expression analysis arena and from the text-based classification arena.

Graphical Methods for Binary Data

◆ Karen L. Price, Eli Lilly and Company; John W. Seaman, Baylor University

Eli Lilly and Company, Lilly Corporate Center—D/C 6026, Indianapolis, IN 46285

price_k@lilly.com

Key Words: *binary data, graphical representation, grayscale plots, Bayesian methods*

Graphical analysis is essential in most statistical problems. Boxplots, histograms, stem-and-leaf plots, and scatterplots are standard graphical tools. For investigating bivariate relationships among continuous variables scatterplots are particularly useful, but if the response is binary such plots are of dubious value. Recently, Eno and Terrell (1999) introduced an innovative method using grayscale graphics as a solution to this problem. We improve on Eno and Terrell's method to ease interpretation of graphical results. We also adapt the method for use with Bayesian models allowing the user to plot posterior distributions. We review some standard methods for plotting binary data, detail the grayscale plotting method introduced by Eno and Terrell, discuss improvements we have made to the grayscale plot for binary data, and adapt grayscale plots for use with Bayesian models.

Learning the Linear Filters from Sounds and Images

◆ Ray-Bing Chen, National University of Kaohsiung; Ming Zheng, University of California, Los Angeles

National University of Kaohsiung, Institute of Statistics, No. 700 Kaohsiung University Rd., Nan-Tzu District, Kaohsiung, 811 Taiwan, R.O.C.

rbchen@nuk.edu.tw

Key Words: *filter linearing, gradient method, Markov chain Monte Carlo, FRAME model*

We are interesting in learning the linear filters from sounds and images. The model of sounds and images we considered here is FRAME model proposed by Zhu, Wu and Mumford (1997). After selecting the filters, FRAME model has good results in image analysis and synthesis. However, we assume the filters are all unknown and are learned from the data. In order to learn the filters, we propose an iterative algorithm that contents the two steps. The first step is to update the filters by gradient methods. Then the second step is to synthesize the signals based on the updating filters by MCMC methods. Finally, we demonstrate several experiments on sounds and images to show the performance of our algorithm.

Repeated Use of the Pythagorean Theorem for Plots of Common Statistics

◆ Frank J. Matejcik, South Dakota School of Mines & Technology

South Dakota School of Mines & Technology, 501 E. St. Joseph St., Rapid City, SD 57701-3901

Frank.Matejcik@sdsmt.edu

Key Words: *graphics, plots, regression, ANOVA, contingency tables, ancient tools*

Sums of squares occur in many common statistics. The square root of the sum of squares can be computed graphically using by repeating the Pythagorean theorem. Accordingly, repeating the Pythagorean theorem can be used for plots related to many common statistics and related calculations: the empirical standard deviation, the standard deviation of a probability mass function, the correlation coefficient in regression, the F-test in ANOVA, chi-squared goodness-of-fit test, and chi-squared tests for contingency tables. Plots for each of these calculations are displayed and a brief discussion of their potential value of as diagnostic and instructional aids.

Spherical Distance Matrices and the Horseshoe Effect

◆ Michael W. Trosset, College of William & Mary

College of William & Mary, Dept. of Mathematics, PO Box 8795, Williamsburg, VA 23187-8795

trosset@math.wm.edu

Key Words: *seriation, multidimensional scaling, distance geometry*

One important use of multidimensional scaling is seriation, the ordering of a set of objects on the basis of dissimilarities between them. In many applications, it is difficult to distinguish degrees of large dissimilarity, resulting in a phenomenon that D.G. Kendall dubbed “the horseshoe effect.” Kendall (1971) offered an intuitive explanation of the horseshoe effect; I will attempt a more technical explanation that appeals to the theory of spherical distance matrices.

Visualizing Item Nonresponse for Large-scale Surveys for High-stakes Examinations (SEER Method)

◆ Chris Chiu, Law School Admission Council

Law School Admission Council, 661 Penn St., Newtown, PA 18940

cchiu@lsac.org

Key Words: *data visualization, item nonresponse, matrix algebra & spare matrix, statistical computing, Law School Admission Test (LSAT), national survey*

Item nonresponse is a threat to large-scale surveys. Results derived from survey questions with high item nonresponse rates can lead to biased and misleading conclusions about estimates of national demographic, housing, social, and economic characteristics as well as the labor force in the country. Despite the substantial efforts and resources used to reduce item nonresponse, such as the use of Computer Assisted Telephone Interviews (CATI), Computer Assisted Personal Interviews (CAPI), item-skipping prevention methods (e.g., cognitive methodology, use of good information and organization, use of respondent-expected-skipping sequence), it is impossible to totally avoid item nonrespnse in large-scale surveys because item nonresponse is an artifact that can be attributed to many factors. Some of the factors cannot be completely controlled by the survey administers (e.g., respondents refuse to respond, respondents do not know the answers, respondents incorrectly skip an item). The proposed research project sets out to develop a set of statistical and computing tools with an ultimate goal of empowering the Bureau and other agencies to become more proactive in this issue.

Displaying and Combining Small-sample Likelihoods from Generalized Linear Models

◆ Nicholas J. Barrowman, Children’s Hospital of Eastern Ontario Research Institute; Ransom Myers, Dalhousie University; Keith O’Rourke, Ottawa Health Research Institute

Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Rd., Chalmers Research Group, Ottawa, ON K1H 8L1 Canada

nbarrowman@cheo.on.ca

Key Words: *generalized linear models, likelihood plots, likelihood-based confidence intervals, profile likelihood, raindrop plot*

For large sample sizes, the asymptotic Wald approximation used in standard output from many statistical packages is usually adequate for inferences for generalized linear models. However, there are cases where these approximations are inaccurate and even inappropriate such as when zeros are present and only one-sided confidence intervals are sensible. For example, in a clinical trial if all patients experience an event, the log-likelihood is not approximately quadratic. In ecological studies, a similar situation occurs when during one time period animals of a particular kind are observed but during another none are observed. In such situations, the direct use of likelihood is both less opaque and more accurate than Wald approximations. The purpose of this paper is to provide accessible methods and guidance for directly using likelihood for plotting and construction of confidence intervals.

Profile likelihood is used to summarize information about single parameters of interest and to combine information from different sources. R and S-Plus code is provided to compute profile likelihoods, display raindrop plots, and obtain confidence intervals.

219 Issues in Clinical Trial Design

WNAR, Biometrics Section, ENAR

Tuesday, August 10, 10:30 am–12:20 pm

A Statistical Design of Phase II Cancer Trials Using Total and Complete Response Endpoints

◆ Ying Lu, University of California, San Francisco; Hua Jin, University of California, San Francisco; Kathleen R. Lamborn, University of California, San Francisco

University of California, San Francisco, Dept. of Radiology, Box 0946, San Francisco, CA 94143-0946

ying.lu@radiology.ucsf.edu

Key Words: *phase II cancer trial, optimal, two-stage design, trinomial*

Phase II clinical trials in oncology are used to initially evaluate the therapeutic efficacy of a new treatment regimen. Simon's two stage desig based on total response (TR) rate is commonly used for such trials. Several authors have proposed alternative strategies to jointly consider either the response and toxicity or the responses and early progression. Because TR consists of both partial response (PR) and complete response (CR), and these two types of responses may have a different impact on patient survival, Lin and Chen (2000) proposed a flexible design that is based on a weighted average score of PR and CR rates to take into account the relative importance of a CR vs. PR. Panageas and colleagues (2002), on the other hand, used a trinomial model to develop a rejection region based on the joint frequencies of PRs and CRs . We reformat their hypotheses to assess the treatment based on TR and CR frequencies. A new two-stage optimum phase II trial design based on TR and CR is developed. We provide guides on searching the stopping and rejecting regions and determination of sample size.

Optimal Designs for Contingent Response Models

◆ Huwaida S. Rabie, University of Missouri; Nancy Flournoy, University of Missouri

University of Missouri, Dept. of Statistics, 146 Middlebush, Columbia, MO 65211

hsr2mf@mizzou.edu

Key Words: *dose-response, continuation-ratio model, extreme value distribution, phase I/II clinical trails, canonical designs*

We study D- and C-optimal designs for the contingent response models of Li, Durham, and Flournoy. In the contingent response model, there are two types of failure. We call one failure type toxicity and the other disease failure. No toxicity and no disease failure is a success or cure. We assume disease failures are contingent on toxicity in that they are only observed in the absence of toxicity. We also assume the probability of toxicity increases with the dose, and the probability of disease failure given no toxicity decreases with dose. Interest is in finding C-optimal designs for estimating the dose that maximizes the cure probability. We find canonical optimal designs and show that other designs in the location/scale family can be obtained from a canonical design. We use the positive/negative extreme value contingent response model to provide a specific illustration of the D- and C-optimal designs.

Statistical Properties of Modified Algorithm-based Designs for Phase I Cancer Clinical Trials

◆ Yong Lin, University of Medicine and Dentistry of New Jersey; Weichung J. Shih, University of Medicine and Dentistry of New Jersey

University of Medicine and Dentistry of New Jersey, 195 Little Albany St., New Brunswick, NJ 08901

linyo@umdnj.edu

Key Words: *algorithm-based design, phase I cancer clinical trials, maximum tolerated dose*

Algorithm-based designs are still widely used in practice for phase I cancer clinical trials because of its simplicity in logistics for clinical investigators to carry out, compared to various model-based continual reassessment methods (CRMs). Lin and Shih discussed statistical properties of the basic algorithm-based designs. We introduce modified traditional algorithm-based designs, which we encountered in many of the phase I cancer clinical trials. Key statistical properties as well as clinical issues related to these designs are investigated, including: (i) the probability of a dose being chosen as MTD (maximum tolerated dose); (ii) the expected number of patients at each dose level; (iii) target toxicity level (expected dose limiting toxicity at MTD); (iv) expected number of toxicity at each dose level; (v) expected overall toxicity in a trial. The exact formulae for the corresponding statistical quantities will be derived. These statistical properties are important for clinicians who use these algorithm-based designs for a phase I cancer clinical trial to gain insights of the design before starting the trial.

Methods for Matching Subjects on Baseline Outcome Measures Prior to Randomization

◆ Misook Park, Virginia Commonwealth University; Robert E. Johnson, Virginia Commonwealth University

Virginia Commonwealth University, Dept. of Biostatistics, Box 980032, 1101 East Marshall St., Richmond, VA 23298-0032

mpark@vcu.edu

Key Words: *clustering, matching, baseline measures*

Matching subjects prior randomization may increase the power to detect differences between subgroups by reducing the variance of the differences. Common methods include pairing where two subjects are matched on one or more characteristics. Matched groups of size greater than two may be formed in a similar way. One approach to the latter is to use an algorithm for identifying clusters of observations. Once identified, subjects within each cluster are randomized to the study arms. It is necessary to place constraints on the clustering algorithm, namely, each cluster must have at least the same number of subjects as arms of the study. For example, a hierarchical method for identifying clusters may result in one or more clusters having less than the required number of subjects. These clusters must then be fused with a neighboring cluster. This fusion, though necessary, may result in a nonoptimal result. For example, variation may be improved by breaking apart the newly formed cluster. We will explore adaptations of known methods in the presence of such constraints. Hierarchical methods such as the centroid and average linkage will be considered in addition to an all-clusters approach.

Bayesian Adaptive Designs for Clinical Trials

◆ Yi Cheng, Indiana University, South Bend; Yu Shen, University of Texas M. D. Anderson Cancer Center

Indiana University, South Bend, 1700 Mishawaka Ave., South Bend, IN 46634

ycheng@iusb.edu

Key Words: *adaptive designs, decision theory, group sequential clinical trials, loss functions, one-step backward induction*

A Bayesian adaptive design is proposed for a comparative two-armed clinical trial using decision theoretical approaches. A loss function is specified to consider the cost for each patient, and the costs of making incorrect decisions at the end of a trial. At each interim analysis, the decision to terminate or to continue the trial is based on the expected loss function while concurrently incorporating efficacy, futility, and cost. The maximum number of interim analyses is not pre-fixed but decided adaptively by the observed data. We derive explicit connections between the loss function and the frequentist error rates, so that the desired frequentist properties can be maintained for regulatory settings. The operating characteristics of the design are able to be evaluated on frequentist grounds. Extensive simulations are carried out to compare the proposed design with existing ones. The design is general enough to accommodate both continuous and discrete types of data.

SCPRT: A Sequential Procedure Giving Another Reason to Stop a Trial Early

◆ Xiaoping Xiong, St. Jude Children’s Research Hospital; Ming T. Tan, University of Maryland and Medicine; James Boyett, St. Jude Children’s Research Hospital

St. Jude Children’s Research Hospital, 332 N. Lauderdale St., Biostatistics, Memphis, TN 38105

xiaoping.xiong@stjude.org

Key Words: *clinical trial, design, sequential test*

A sequential procedure for clinical trial is a test design that allows early stopping and satisfies given significance level and power. By this convention, many different sequential test designs are available for a same requirement. Reversely, a same set of observed data can be interpreted as outcomes of different sequential designs with different significance levels and powers. With a same set of observed data, one may conclude a test significant by one sequential design, and conclude the test insignificant by another sequential test design; this phenomenon may lead one to ask a question whether application of sequential tests is rational. Withstanding this challenge, the sequential conditional probability ratio test (SCPRT) is a class of sequential test designs with property that a conclusion made at early stopping is unlikely to be reversed if the trial had continued to the planned end. This property gives us another reason to stop a trial early: if the trial were not stopped as it should, then adding more data and continuing the trial (by the planned end) will not change the conclusion. We will introduce SCPRT procedures for different distributions and various applications.

On the Maximum Total Sample Size of a Group Sequential Test about Binomial Proportions

◆ Myron N. Chang, University of Florida; James L. Kepner, Roswell Park Cancer Institution

University of Florida, 104 N. Main St., Suite 600, Gainesville, FL 32601

mchang@cog.ufl.edu

Key Words: *type 1-3 design, power function, uniformly most powerful test, minimax design*

It is well known that the standard single-stage binomial test is uniformerly most powerful to detect an increase or decrease in a binomial proportion. The general perception is that, to achieve a fixed significance level and power, a group sequential test will require a larger maximum total sample size than required by the corresponding standard single-stage test because missing observations are possible under the group sequential test setting. In this paper, it is proved that, under mild conditions, there exist group sequential tests which achieve the pre-designated significance level and power with maximum total sample size bounded above by the sample size for the corresponding standard single-stage test.

220 Interim Analysis and Two-step Designs

Biopharmaceutical Section

Tuesday, August 10, 10:30 am–12:20 pm

Adjusting Significant Level for Interim Analysis Using Resampling Technique

◆ Yi-wen Ma, Johnson & Johnson Pharmaceutical R&D, LLC; Akiko Okamoto, Johnson & Johnson Pharmaceutical R&D, LLC; Surya Mohanty, Johnson & Johnson Pharmaceutical R&D, LLC

Johnson & Johnson Pharmaceutical R&D, LLC, 1125 Trenton-Harbourton Rd., Titusville, NJ 08560

yma@prdus.jnj.com

Key Words: *significant level, interim analysis, resampling*

Significance level adjustment for interim analysis in survival studies as well as power monitoring in survival studies utilize the asymptotic normality of log-rank type tests. The spending function methods of calculating nominal significance level relies on the possibility of constructing a Brownian motion from the test statistics stochastic process. Unlike linear models, there is a dependence relationship between interim subjects and later subjects. Unless this dependence is assumed to be ignorable, Brownian motion construction may not be feasible. Resampling methods will be utilized to assess the significance level. Comparison to traditional spending function methods also will be provided.

Sequential Methods with Recurrent Events

◆ Michael Shnaidman, Columbia University; Zhiliang Ying, Columbia University

Columbia University, 618 Mathematics Bldg., MC 4403, 2990 Broadway, New York, NY 10027

ms523@columbia.edu

Key Words: *survival analysis, sequential methods, interim analysis, recurrent events, staggered entry*

Sequential methods for interim analyses with recurrent event times outcome and staggered entry. Recurrent event times are frequently used endpoints for clinical trials of diseases such as asthma, epilepsy, and infectious diseases. We will focus on our development of new sequential tests for interim analyses, with the endpoint being recurrent event times. We will discuss issues such as conditional power calculation, sample size modification, nonindependent increments, and covariate ajustment among others.

A Method to Estimate the Variance of an Endpoint from an Ongoing Blinded Trial

◆ Jitendra Ganju, Chiron Corporation; Biao Xing, University of California, Berkeley

Chiron Corporation, 4560 Horton St., M/S U-140, Emeryville, CA 94608

jitendra_ganju@chiron.com

Key Words: *adaptive design, interim analysis, sample size reestimation*

We propose a new method for estimating the variance of an endpoint from an ongoing blinded trial. Our method makes no strict assumption on the density form of the data distribution. The proposed blinded variance estimator is unbiased and compares favorably, when the randomization block size is small, with its unblinded counterpart. We apply our method to data from two clinical trials to estimate the variance of an endpoint and compare it with the result obtained from an EM algorithm based approach.

Application of Statistical-Decision Theory to Strategic Trial Design

◆ A. Lawrence Gould, Merck & Co., Inc.; Julie G. Ma, Merck & Co., Inc.; Michael Nessly, Merck & Co., Inc.

Merck & Co., Inc., BL3-2, West Point, PA 19486

goulda@merck.com

Key Words: *group sequential, Bayesian decision*

A proof-of-concept trial will evaluate the potential of a new drug with a novel mechanism of action and potentially novel clinical effect. Interim analyses will be done to decide whether to terminate the trial early for futility, continue the current trial to the next scheduled evaluation and then decide on how to proceed, or design and implement further trials on an at-risk basis and continue the current trial to obtain pivotal trial design information. We describe the application of decision science principles to select among specific decision rule alternatives, evaluate their statistical properties, and choose among them on the basis of anticipated payoffs under assumptions of costs and benefits and various States of Nature. Evaluating the consequences of differing decision rules (actions) on payoffs under reasonable assumptions of States of Nature and assumed cost/benefit structures provides guidance about how aggressive one should be about go/no go decision rules. Assumptions about costs turn out to have substantial effects on the decision rule payoffs. Moderately aggressive go rules and conservative no go rules tend to produce nearly optimal payoffs for the cost structure.

Internal Pilot Designs with Compound Symmetric and Gaussian Repeated Measures

◆ Christopher S. Coffey, University of Alabama, Birmingham; Keith W. Muller, University of North Carolina, Chapel Hill; Matthew J. Gurka, University of Virginia

University of Alabama, Birmingham, 327M Ryals Public Health Bldg., Dept. of Biostatistics, 1665 University Blvd., Birmingham, AL 35294-0022

ccoffey@uab.edu

Key Words: *exchangeable correlation, cluster sample, univariate approach*

An internal pilot design uses an interim sample size analysis, without interim data analysis, to adjust the final total number of observations. The approach typically gives a sample size sufficiently large (to achieve the statistical power desired), but not too large (which would waste money and time). Research to assess age differences in cerebral vascular tortuosity (curvature in three dimensions) would benefit greatly from internal pilot designs. However, interest in repeated measures across brain regions precludes using existing univariate methods for internal pilots. We report new methods to meet the need for Gaussian data with constant variance and correlation (compound symmetry) across repeated measures, and no missing observations, as in the tortuosity research. We directly extend all known exact (small sample) univariate methods to repeated measures with compound symmetry and any between subject design with fixed predictors (not just t tests). The new methods allow avoiding sample sizes that are too small or too large, while still controlling false positives with compound symmetric data.

Statistical Methods for Pilot Bioequivalence Trials

◆ Guohua Pan, Johnson & Johnson Pharmaceutical R&D, LLC; Yibin Wang, Novartis Pharmaceuticals

Johnson & Johnson Pharmaceutical R&D, LLC, 1125 Trenton-Harbourton Rd., Titusville, NJ 08560

JPan3@PRDUS.JNJ.com

Key Words: *pilot trial, bioequivalence, two-stage process, sample size determination, power function, cross-over design*

In clinical development of a bioequivalent (BE) drug product, a two-step strategy is sometimes adopted. At the first step, a pilot BE trial is conducted to evaluate the acceptability of the test drug product as a candidate for further evaluation in a subsequent pivotal BE trial. At the second step, a full-scale pivotal BE trial is conducted to formally establish bioequivalence. The objective and criterion of a pilot BE trial are different from those of a pivotal BE trial. However, in practice, a pilot BE trial is often designed and analyzed based on the criterion and methods for a pivotal BE trial. This talk proposes and discusses some practical statistical methods that can be easily applied to address the specific objective of a pilot BE trial within the two-step framework.

Optimal Three-stage Study Design for Single-arm Phase II Clinical Trials

◆ Muhammad Jalaluddin, Pfizer Inc.; Enayet Talukder, Pfizer Inc.

Pfizer Inc., 50 Pequot Ave., MS 6025-A4168, New London, CT 06320

muhammad_jalaluddin@groton.pfizer.com

Key Words: *clinical trials, multistage clinical trials, exact binomial distribution, sample size*

Optimal two-stage study designs are commonly used in single-arm phase II clinical trials. One of the objectives of these designs is to minimize exposure of a new therapy when there is no beneficial treatment effect. It is achieved by choosing sample sizes and decision rules such that the expected sample size under null hypothesis is minimum. This expected sample size could still be large. This raises ethical questions because the efficacy level of a new therapy is usually unknown at the early stage of development and can potentially be less efficacious than standard therapy. We have addressed these issues by introducing an optimal three-stage study design. This design requires pre-specification of minimum and desired efficacy levels. We initiate subsequent stages of a study only if data from completed stage(s) suggest better efficacy of the new therapy compared to pre-specified minimum and desired efficacy levels. Therefore, this design allows early termination of a study when the new therapy performs poorly. We have demonstrated the application of this design and have compared it with the existing designs under various set-ups.

221 Pharmaceutical I

Biopharmaceutical Section

Tuesday, August 10, 10:30 am–12:20 pm

Looking for Synergism/Antagonism in In-vitro Drug Combination Data

◆ Steven Novick, GlaxoSmithKline

GlaxoSmithKline, Five Moore Dr. PO Box 13398, MAI.T0201, Research Triangle Park, NC 27709-3398

steven.j.novick@gsk.com

Key Words: *drug combinations, hill model, synergy*

A typical drug-interaction investigation starts with an in-vitro study on human cells. Synergism, antagonism, and additivity (the lack of interaction) can be defined by a generalization of Loewe additivity, called the Interaction index. Several methods based on the interaction index have been published, such as by Greco, et al. (1990, 1995), Chou and Talalay (1984), and Berenbaum (1977, 1985). The method that I present extends the ideas put forth by Berenbaum (1985) and the Berenbaum section in Greco, et al. (1995). The method first divides the data into two parts: that which must be Loewe additive and that which may deviate from additivity. Berenbaum modeled the additive portion of the data with an appropriate response surface and then, after fitting the model to the remaining data, informally scrutinized the residuals. At the 2003 JSM, I introduced a formal method of examining the residuals for the simple case of homoscedastic variance. The method has been generalized to the heteroscedastic variance case.

Pooling Stability Data for Estimating Drug Product Shelf Life

Harry Yang, MedImmune, Inc.; ◆ Ryan Yamagata, MedImmune Vaccines; Iksung Cho, MedImmune Vaccines

MedImmune Vaccines, 297 N. Bernardo Ave., Mountain View, CA 94043

yamagatar@medimmune.com

Key Words: *stability study, degradation rte, poolability testing, shelf life*

The regulatory guidelines mandate that the shelf life or expiration dating of a drug product be estimated based on a stability study that involves several lots of the product. Regression analysis is often used for this purpose. The shelf life is determined to be the time when the lower confidence limit of predicted mean value intercepts the lower specification limit, either based on pooled lots or data from the lot that results in the shortest shelf life. We discuss several issues regarding the pooling test. Graphical and analytical methods are proposed for the determination of shelf life, under various scenarios regarding whether or not data are poolable to estimate a common slope and/or common intercept. A closed-form solution to shelf life is provided. Applications of the two methods to various real-life problems are also presented.

Fieller's Theorem with Its Applications in Clinical Trials

◆ Pam Launt, PPD Development, Inc.; Chunqin Deng, PPD Development, Inc.

PPD Development, Inc., 192 Hestia Dr., Plainwell, MI 49080

pam.launt@rtp.ppdi.com

Key Words: *Fieller's theorem, confidence interval, relative potency*

There are a lot of situations under which a ratio needs to be calculated to represent a quantity. While the estimate for this ratio is relatively straightforward, the confidence interval for the estimate of the ratio may not be easy to compute and some alternative ways may need to be employed to obtain the confidence interval. Fieller's theorem is one of the alternatives that can provide confidence intervals for the ratio of mean values of two random variables having a bivariate normal distribution. According to Fieller's theorem, if the variance and covariance for numerator and denominator of the ratio are known, the confidence interval can be obtained by solving a quadratic inequality function. Fieller's theorem has been often used for constructing the confidence intervals for a ratio in clinical trials, bioassay, and environmental statistics. To illustrate the application of Fieller's theorem in clinical trial, three examples are given: confidence interval for the incremental cost-effectiveness ratio (ICER) in health economic area; confidence interval for relative potency in a phase III clinical trial; and the confidence interval used in a bioequivalence test.

Clustering Error Rate in Temporal Data

◆ Nicoleta Serban, Carnegie Mellon University; Atalanta Ghosh, Eli Lilly and Company

Carnegie Mellon University, 5000 Forbes Ave., Baker Hall 232, Dept. of Statistics, Pittsburgh, PA 15217

nserban@stat.cmu.edu

Key Words: *clustering, clustering error rate, temporal gene expression data, temporal data*

We propose a technique for estimation of the clustering error rate when clustering a large number of nonconstant curves.We estimate the error due to the fact that we are clustering the estimated rather than the true curves. We obtain an asymptotic confidence bound for the clustering estimation error based on estimated confidence sets of the nonconstant curves. The method we use for uniform confidence set estimation was introduced by Beran and Dumbgen (1998). The clustering error rate can be used to infer how much information is lost when a small number of time points is used and when the signal-to-noise ratio is low. Our motivating example is temporal gene expression data but the method is very general.

A New Approach for Simple One-compartment PK Models

◆ Donghui Zhang, Merck & Co., Inc.

Merck & Co., Inc., PO Box 2000, RY33-300, Rahway, NJ 07065

donghui_zhang@merck.com

Key Words: *PK model, mixed effects model*

Pharmacokinetic (PK) models, usually nonlinear, characterize the process of drug absorption, distribution, elimination using the inferred form of plasma concentration as a function of time. A particular mixed effect model approach has been developed for the simple one-compartment model. Unlike the current methods, we do not restrict the distribution of the random effects to be lognormal, and we do not use the usual linear approximation, thus removing the source of inconsistency when estimating population parameters. This new approach especially suits the routine population PK data where the number of observations from each indiviudal is small, while the total number of individuals can be sufficiently large.

Developing and Validating an In-vivo In-vitro Correlation Using Formal Hypothesis-testing

◆ James T. Kost, Merck & Co., Inc.; Kalyan Ghosh, Merck & Co., Inc.

Merck & Co., Inc., 10 Sentry Pkwy., BL3-2, Blue Bell, PA 19422

James_Kost@merck.com

Key Words: *bioequivalence, dissolution, IVIVC*

A model in drug development that relates the in-vitro dissolution profile of a formulation to its in-vivo plasma concentration profile is often referred to as an in-vitro-in-vivo correlation (IVIVC). The method proposed by the U.S. Food and Drug Administration for developing and evaluating potential IVIVC models possesses several well-known shortcomings, the foremost of which is its failure to incorporate any measure of variability into the validation process. No quantification of risk is avialable at the design stage and there is limited confidence in the ultimate decision to either accept or reject the potential model at study's end. A new method for developing an IVIVC model, with subsequent validation performed through formal testing, is proposed herein. This testing is similar to that usually performed in an in-vivo bioequivalence study, and is facilitated by the construction of a predicted plasma concentration curve for each in vitro dissolution run. The relative advantages and disadvantages of the proposed method are discussed and an example is given.

Comparison of Treatment Effects Using Empirical Bayes Method

◆ Wei Shen, Eli Lilly and Company; Chaofeng Liu, Eli Lilly and Company; Jun Xie, Purdue University

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

shen@lilly.com

Key Words: *empirical Bayes, borrowing information, clinical trial, treatment comparison*

Bayes and empirical Bayes (EB) methods are widely used in medical research with applications ranging from disease-mapping to risk assessment. The beauty of the Bayesian approach is its ability to structure complicated models, inferential goals, and analyses. By formalizing the relationship among components and "borrowing information" among them, Bayes and EB methods can produce more valid, efficient and informative statistical evaluations than those based on traditional methods. Consider comparison of treatment effects in clinical trials, where data arise from a two-stage compound sampling model. We applied EB methods to estimate individual treatment effect within each group and compare treatment effects between groups. Simulation studies were performed to evaluate the performance of the EB approach as well as standard procedures. The results demonstrated that the EB method was more accurate. We applied this method to clinical trial data evaluating new therapies.

222 Analytical Issues in Time Series

Biometrics Section, Section on Statistics in Epidemiology

Tuesday, August 10, 10:30 am–12:20 pm

Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities

◆ Roger Peng, Johns Hopkins University

Johns Hopkins University, Bloomberg School of Public Health, Dept. of Biostatistics, 615 N. Wolfe St., Baltimore, MD 21205

rpeng@jhsph.edu

Key Words: *particulate matter, effect modification*

The focus of this project is to examine changes across seasons of the short-term effects of particulate matter on daily mortality. We are developing models for time series data that allow for interactions between season, pollution, and possibly other confounders. In addition, we use Bayesian hierarchical models to examine how the seasonality of pollution effects varies across regions of the U.S. We apply our methods to the recently updated NMMAPS database which includes daily time series for the period 1987-2000 for the largest 100 cities in the U.S.

Repetition, Rise, and Fall in the Analysis of Biotic Time Series

◆ Minu K. Patel, University of Illinois, Chicago; Hector Sabelli, Chicago Center for Creative Development

University of Illinois, Chicago, Dept. of Medical Surgical Nursing, 845 S. Damen, Chicago, IL 60612

mkp@uic.edu

Key Words: *bipolar, bios, novelty, diversification, nonrandom complexity*

New methods for time series analysis serve to identify creative phenomena. Recursions of trigonometric functions that model bipolar (positive and negative) feedback generate a pattern (bios) with characteristics similar to those found in time series of natural creative processes, namely novelty, diversification, and nonrandom complexity. We introduce another technique that differentiates bios from chaos. Repetition carries information in a nonrepetitive environment, as change conveys information in a stable one. Empirical time series (heartbeat intervals, respiration, some economic and meteorological data), biotic series and stochastic noise display more consecutive repetition than their shuffled copies. Chaotic series show very few consecutive repetitions. In evolving time series, consecutive repetition peaks at transitions from one pattern to another.

Estimating a Smooth Common Transfer Function with a Panel of Time Series—Inflow of Larvae Cod as an Example

◆ Elizabeth A. Hansen, University of Iowa; Kung-sik Chan, University of Iowa; Nils C. Stenseth, University of Oslo

University of Iowa, Dept. of Statistics and Actuarial Science, Iowa City, IA 52242

elizabeth-hansen@uiowa.edu

Key Words: *cod-spawning-date distribution, (generalized) cross-validation, seemingly unrelated regression, multimodality*

We consider the problem of estimating a smooth common transfer function shared by a panel of short time series that are contemporaneously correlated. We propose an estimation scheme using a likelihood approach that penalizes the roughness of the common transfer function. For a known smoothness parameter, the estimation can be done via an iterative procedure. The method of cross-validation can be used to determine the smoothness parameter. We illustrate the proposed method with a biological example of indirectly estimating the spawning date distribution of North Sea cod. Some simulation results are reported on the empirical performance of the proposed method.

Conditional Growth Charts

◆ Ying Wei, University of Illinois, Urbana-Champaign

University of Illinois, Urbana-Champaign, Dept. of Statistics, 101 Illini Hall, 725 S. Wright St., Champaign, IL 61820

yingwei@uiuc.edu

Key Words: *growth charts, longitudinal model, quantile regression*

The reference growth charts are widely used to screen measurements of individual subjects in the context of population values. Most of the growth charts in use are generated from cross-sectional data and relied on some form of normality in the models. We propose to construct conditional growth charts based on longitudinal quantile regression models, which allows screening of individuals based on their prior growth paths and other useful covariates. No distributional assumptions need to be made with this approach. Examples will be given to illustrate the difference between conditional and unconditional growth charts.

A Simple and General Changepoint Detector

◆ Robert H. Riffenburgh, Naval Medical Center San Diego; Kevin M. Cummins, San Diego State University

Naval Medical Center San Diego, 3069 Award Row, San Diego, CA 92122

riff@sdsu.edu

Key Words: *changepoint, time series, longitudinal data, Moving F, CUSUM/EWMA, process control*

A Moving F statistic to detect and locate changes in longitudinal data (time series) is easier to use than other commonly employed change point detectors (process control, CUSUM, EWMA, data-based bandwidth selection) and is shown to work in several situations where these other detectors fail or lack methodological development. In not-necessarily-stationary series, the expected

222 Analytical Issues in Time Series

Biometrics Section, Section on Statistics in Epidemiology

Tuesday, August 10, 10:30 am–12:20 pm

Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities

◆ Roger Peng, Johns Hopkins University

Johns Hopkins University, Bloomberg School of Public Health, Dept. of Biostatistics, 615 N. Wolfe St., Baltimore, MD 21205

rpeng@jhsph.edu

Key Words: *particulate matter, effect modification*

The focus of this project is to examine changes across seasons of the short-term effects of particulate matter on daily mortality. We are developing models for time series data that allow for interactions between season, pollution, and possibly other confounders. In addition, we use Bayesian hierarchical models to examine how the seasonality of pollution effects varies across regions of the U.S. We apply our methods to the recently updated NMMAPS database which includes daily time series for the period 1987-2000 for the largest 100 cities in the U.S.

Repetition, Rise, and Fall in the Analysis of Biotic Time Series

◆ Minu K. Patel, University of Illinois, Chicago; Hector Sabelli, Chicago Center for Creative Development

University of Illinois, Chicago, Dept. of Medical Surgical Nursing, 845 S. Damen, Chicago, IL 60612

mkp@uic.edu

Key Words: *bipolar, bios, novelty, diversification, nonrandom complexity*

New methods for time series analysis serve to identify creative phenomena. Recursions of trigonometric functions that model bipolar (positive and negative) feedback generate a pattern (bios) with characteristics similar to those found in time series of natural creative processes, namely novelty, diversification, and nonrandom complexity. We introduce another technique that differentiates bios from chaos. Repetition carries information in a nonrepetitive environment, as change conveys information in a stable one. Empirical time series (heartbeat intervals, respiration, some economic and meteorological data), biotic series and stochastic noise display more consecutive repetition than their shuffled copies. Chaotic series show very few consecutive repetitions. In evolving time series, consecutive repetition peaks at transitions from one pattern to another.

Estimating a Smooth Common Transfer Function with a Panel of Time Series—Inflow of Larvae Cod as an Example

◆ Elizabeth A. Hansen, University of Iowa; Kung-sik Chan, University of Iowa; Nils C. Stenseth, University of Oslo

University of Iowa, Dept. of Statistics and Actuarial Science, Iowa City, IA 52242

elizabeth-hansen@uiowa.edu

Key Words: *cod-spawning-date distribution, (generalized) cross-validation, seemingly unrelated regression, multimodality*

We consider the problem of estimating a smooth common transfer function shared by a panel of short time series that are contemporaneously correlated. We propose an estimation scheme using a likelihood approach that penalizes the roughness of the common transfer function. For a known smoothness parameter, the estimation can be done via an iterative procedure. The method of cross-validation can be used to determine the smoothness parameter. We illustrate the proposed method with a biological example of indirectly estimating the spawning date distribution of North Sea cod. Some simulation results are reported on the empirical performance of the proposed method.

Conditional Growth Charts

◆ Ying Wei, University of Illinois, Urbana-Champaign

University of Illinois, Urbana-Champaign, Dept. of Statistics, 101 Illini Hall, 725 S. Wright St., Champaign, IL 61820

yingwei@uiuc.edu

Key Words: *growth charts, longitudinal model, quantile regression*

The reference growth charts are widely used to screen measurements of individual subjects in the context of population values. Most of the growth charts in use are generated from cross-sectional data and relied on some form of normality in the models. We propose to construct conditional growth charts based on longitudinal quantile regression models, which allows screening of individuals based on their prior growth paths and other useful covariates. No distributional assumptions need to be made with this approach. Examples will be given to illustrate the difference between conditional and unconditional growth charts.

A Simple and General Changepoint Detector

◆ Robert H. Riffenburgh, Naval Medical Center San Diego; Kevin M. Cummins, San Diego State University

Naval Medical Center San Diego, 3069 Award Row, San Diego, CA 92122

riff@sdsu.edu

Key Words: *changepoint, time series, longitudinal data, Moving F, CUSUM/EWMA, process control*

A Moving F statistic to detect and locate changes in longitudinal data (time series) is easier to use than other commonly employed change point detectors (process control, CUSUM, EWMA, data-based bandwidth selection) and is shown to work in several situations where these other detectors fail or lack methodological development. In not-necessarily-stationary series, the expected mean path may take on a variety of functional forms (constant, linear, quadratic, exponential, etc.), in which any one or any combination of parameters may change (jump, slope, curvature, etc.), and the residual variability about the path may also change (increase or decrease). Furthermore, a decrease in variability may mask a change in a parameter. The Moving F statistic detects a change in any parameter or simultaneous set of parameters and/or variability, and also uncovers a masked change. Patterns in the forces giving rise to the data may often be perceived. An example from monitored PSA level in a patient equivocal for prostate cancer shows that cancer could have been detected nearly two years earlier than by conventional means.

A Generalized Open-loop Threshold Model for Non-normal Nonlinear Time Series: Plague in Kazakhstan as in an Illustration

◆ Noelle I. Samia, University of Iowa; Kung-sik Chan, University of Iowa; Nils C. Stenseth, University of Oslo

University of Iowa, Dept. of Statistics and Actuarial Science, 241 Schaeffer Hall, Iowa City, IA 52242

nsamia@stat.uiowa.edu

Key Words: *binomial distribution, delay, epizootic events, exponential family, stochastic regression*

The Open-loop Threshold model is essentially a piecewise linear stochastic regression model, with the errors often assumed to be normal. We generalize the Open-loop Threshold model for time series regressions with non-normal errors. Specifically, it is assumed that the response variables conditionally belong to an exponential family, whose canonical parameters are linked to some piecewise linear stochastic regression function. We study the specific case where the response variables are conditionally Binomially distributed. Some probabilistic properties of the model and the large-sample properties of an estimation scheme will be discussed. Our approach is motivated by the need for modeling nonlinearity in serially correlated epizootic events. Data coming from monitoring conducted in a natural plague focus in Kazakhstan are used to illustrate this model by obtaining biologically meaningful conclusions regarding the threshold relationship between prevalence of plague and some covariates including past abundance of gerbils and other meteorological variables.

223 Genetics ▲ ⊞

Biometrics Section, ENAR

Tuesday, August 10, 10:30 am–12:20 pm

Optimizing Neural Networks for Proteins Structure Prediction

◆ Blythe Durbin, University of California, Berkeley

bpdurbin@stat.berkeley.edu

Key Words: *neural networks, protein structure prediction, bioinformatics*

Neural networks are a popular method for predicting the secondary structure of a protein from its amino acid sequence. However, overfitting poses a serious obstacle to effective use of neural networks for this and other problems. Due to the huge number of parameters in a typical neural network, one may obtain a network fit which perfectly predicts the training data yet fails to generalize to other datasets. Overfitting may be avoided by altering the network topology so that some connections are removed, thus reducing the total number of parameters. In the area of secondary structure prediction, work has focused on optimizing the network architecture by hand based on subject-matter knowledge. We propose instead a method for selecting an optimal network architecture in a data-adaptive fashion using the deletion/substitution/ addition algorithm introduced in Sinisi and van der Laan (2003) and Molinaro and van der Laan (2003), and demonstrate the application of this approach to protein secondary-structure prediction.

Statistical Method for Constructing Transcriptional Regulatory Networks Using Gene Expression and Sequence Data

◆ Biao Xing, University of California, Berkeley; Mark van der Laan, University of California, Berkeley

University of California, Berkeley, Division of Biostatistics, School of Public Health, 140 Warren Hall #7360, Berkeley, CA 94720

bxing@stat.berkeley.edu

Key Words: *transcriptional regulatory network, analysis of microarray data, transcription factor, model selection, false discovery rate, mixture model*

Transcriptional regulation is one of the most important means of gene regulation. Uncovering transcriptional regulatory network helps us to understand the complex cellular process. We describe a comprehensive statistical approach for constructing the transcriptional regulatory network using data on gene expression, promoter sequences, and transcription factor binding sites. Our simulation studies show that the overall and false positive error rates in the estimated transcriptional regulatory network are expected to be small if the systematic noise in the constructed feature matrix is small. Our analysis based on 658 microarray experiments on yeast gene expression programs and 46 yeast transcription factors suggests that the method is capable of identifying important transcriptional regulatory interactions and uncovering the corresponding regulatory network structures.

A Multivariate Analysis of Protein Polymorphism

◆ Eric A. Stone, Stanford University

Stanford University, 390 Serra Mall, Dept. of Statistics, Stanford, CA 94305

eastone@stanford.edu

Key Words: *cSNP, coding polymorphism, amino acid substitution*

Identifying amino acid substitutions that will impair protein function is of major consequence to researchers studying human disease. An alignment of homologous proteins across species is a sample from the space of functional proteins, with the importance of an amino acid residue typically correlated inversely with evolutionary variation. We decompose variation in terms of physicochemical properties salient to protein structure and function. We propose a transparent method that uses comparative sequence data to assess the specific variation due to physicochemical constraints at every position in a protein. The potential deleteriousness of an amino acid substitution is quantified by one score reflecting its agreement with the positional constraints, and under the null hypothesis of preserved function, we show that this score is distributed like Snedecor's F. Employing a threshold to classify the fate of variant proteins, we evaluate our predictions against the results of published mutagenesis experiments.

Permutation-based methods for One- and Two-sample Problems with High-dimensional Covariates and Multivariate Responses

◆ Greg DiRienzo, Harvard University

Harvard University, Dept. of Biostatistics, 655 Huntington Ave. Bldg. II, 4th Floor, Boston, MA 02115

dirienzo@hsph.harvard.edu

Key Words: *genotype sequence, joint confidence region, multiple testing, simultaneous inference, sparse data, variable confidence band*

Permutation-based resampling methods are proposed for the multiple-testing problem with P>=1 response variables. Simultaneous alpha-level variable confidence bands are obtained for an R-dimensional vector of test statistics. Inference is simultaneously conducted with respect to all P response variables. These methods are utilized to: reduce the dimension of a large set of covariates; test association between covariate patterns and response; test for a treatment group effect on response separately for each of several covariate patterns. We present a simulation study comparing our methods and those of Pollard and vander Laan (2003) in the gene expression setting and show that precise variance-covariance estimation enables our methods to outperform theirs for small sample sizes or skewed response data. Simultaneous P-dimensional confidence regions for statistics of interest are easily obtained for each setting considered. Examples include relating HIV-1 genotype sequence and resistance phenotypes for P drugs, and joint estimation of treatment effect on time to virologic failure and time to off-track within several strata of baseline covariates.

Improved Changepoint Methods for the Analysis of Array-based DNA Copy Number Data

◆ Adam B. Olshen, Memorial Sloan-Kettering Cancer Center; E. S. Venkatraman, Memorial Sloan-Kettering Cancer Center

Memorial Sloan-Kettering Cancer Center, 307 East 63rd St., 3rd Floor, New York, NY 10021

olshen@yahoo.com

Key Words: *changepoint, binary segmentation, CBS, DNA copy number, array CGH, ROMA*

We have previously presented circular binary segmentation (CBS), a method for analyzing DNA copy number data. We now present improvements to it. These include a second round of segmentation that splits the data relative to a reduced set of means. Also, we present a modification if the data are binary.

Acceptance Sampling Methods Extensions to Crop Seed Purity Testing for Transgenic Traits

◆ Kirk M. Remund, Monsanto; Jean-Louis Laffont, Pioneer GÈnÈtique; Deanne L. Wright, Pioneer Hi-Bred International Inc.

Monsanto, 800 North Lindbergh Blvd., St. Louis, MO 63167

kirk.m.remund@monsanto.com

Key Words: *acceptance sampling, GMO testing, assay errors, DNA, PCR*

Acceptance sampling methods have been used for decades to reject or accept product lots before release to the consumer. Many industries have even abandoned the use of acceptance sampling methods because end-product testing is no longer needed. However, in crop seed testing for transgenic trait purity or impurity, these methods have proven very useful in recent years. Purity tests include testing for protein expression via ELISA methods and DNA sequences via polymerase chain reaction methods (PCR). This paper presents some useful extensions to the standard acceptance sampling methods based on the binomial distribution for seed purity testing. These extensions include modifications to allow purity characteristics to be measured on pools of many seeds rather than individual seeds. An approach to adjust acceptance sampling plans using qualitative methods by the assay false positive and false negative rates is presented. For quantitative PCR assays, standard acceptance sampling methods extensions are made to allow method and subsampling variability estimates from a variance components model to be coupled with binomial sampling variability in the operating characteristic curve.

Are Gene Locations Clustered on Chromosomes?

◆ Naomi S. Altman, Pennsylvania State University; Eli Walters, Pennsylvania State University; Laura Elnitski, Pennsylvania State University

Pennsylvania State University, Dept. of Statistics, 326 Thomas Bldg., University Park, PA 16802

naomi@stat.psu.edu

Key Words: *gene clustering, goodness-of-fit, chi-squared test, Kolmogorov test, Cramer-von Mises test, bioinformatics*

Genes that are more closely spaced on the chromosome than expected by chance are said to be spatially clustered. The arrangement of genes along the chromosome is known to be associated with gene expression, so understanding spatial clustering is important. Standard tests of clustering vs. uniformity do not take into account two important features of genes: the high variability of gene length and the low probability that gene locations overlap (exclusion). We show by simulation that the standard null distributions which ignore length and exclusion do not appropriately approximate the true null distributions of standard tests such as the chi-squared test. We therefore recommend bootstrap sampling to estimate the null distributions. Simulations demonstrate that the chi-squared goodness-of-fit test is a more powerful test of clustering than two other commonly used tests—Kolmogorov and Cramer-von Mises—when the distribution of gene lengths and locations is modeled by a mixture of exponentials and there is a single cluster. The bootstrap method to test clustering is illustrated using data from human chromosome 22.

224 Hypothesis Testing and Model Selection △ ⊞

Biometrics Section, Section on Statistics in Defense and National Security, Section on Statistics in Epidemiology

Tuesday, August 10, 10:30 am–12:20 pm

Approximate Score Confidence Intervals for a Beta-binomial Proportion

◆ Michael E. Schuckers, St. Lawrence University

St. Lawrence University, 211 Valentine Hall, Dept. of Math, Computer Science and Statistics, Canton, NY 13617

schuckers@stlawu.edu

Key Words: *rate estimation, false accept, biometric identification, false reject*

We consider an adjustment to the Beta-binomial proportion similar to that made by Agresti and Coull for binomial proportions. Traditional methods for making confidence intervals for a Beta-binomial proportion tend to perform poorly when the intra-individual correlation is large. In addition, the likelihood of the product Beta-binomial is not tractable for making most classic confidence intervals. To overcome this, we use an approximate score test where we substitute the estimated variance based on the likelihood function with the estimated variance of the estimator assuming the null hypothesis holds. A confidence interval is then derived by solving the resulting quadratic equation. Both the original and a simplified version are considered. We then compare these approaches to traditional interval for a Beta-binomial proportion via Monte Carlo simulation. The application that motivates this work is the estimation of false accept and false reject error rates for biometric identification devices.

A New Approximation to the Distribution of the Empirical Score Statistic

◆ Brian G. Leroux, University of Washington; Lloyd A. Mancl, University of Washington; Elizabeth Thomas, University of Washington

University of Washington, Dept. of Biostatistics, Box 357475, Seattle, WA 98195

leroux@u.washington.edu

Key Words: *score test, correlated data, small sample*

The empirical (generalized) score test and the robust Wald test are commonly used procedures for hypothesis testing about marginal models with cluster-correlated data. Several researchers have shown that these tests do not have valid Type I error rates in small samples (small number of clusters). In particular, the Wald test has been shown to be anticonservative whereas the score test has been shown to be conservative. Small-sample corrections for these tests have been proposed that improve the performance but do not completely correct the problem. We have found an upper bound for the empirical score test that depends on sample size and helps to explain the small-sample conservativeness of the test. We propose a new distribution for the empirical score statistic that acknowledges this upper bound and provides an excellent approximation to the distribution of the statistic in small samples. Simulation studies show that use of the critical values from the new distribution yield a score test with valid Type I error probabilities in many situations even with very small sample sizes. The test will be illustrated through applications to dental datasets.

A Generalized test of the Magnitude of Parametric Assumption Departure for Linear Models in Longitudinal Studies

◆ David K. Williams, University of Arkansas for Medical Sciences

University of Arkansas for Medical Sciences, 4301 West Markham Slot 820, Little Rock, AR 72205

williamsdavidk@uams.edu

Key Words: *longitudinal, nonparametric, parametric assumptions*

Recently there has been a unification of theory for nonparametric models that can be applied to longitudinal studies. These methods have a few trivial assumptions to be met for their application. Generalized linear models (GLM) are also in common use to model longitudinal data. Associated with these GLM models is a set of assumptions about the data. These assumptions include properties about the form, distribution, and covariance structure of the observations. One question that naturally arises in consideration of nonparametric methods is: Does there exist departures from the parametric assumptions that affect the conclusions for the data at hand? We propose an approximate test that provides a statistical yardstick to provide indications of the magnitude that assumption departures may be playing on a particular dataset. This is done by side-by-side statistical test of corresponding parametric and nonparametric models.

Variable Selection for Marginal Longitudinal Generalized Linear Models

◆ Eva Cantoni, University of Geneva; Joanna Mills Flemming, University of Geneva; Elvezio Ronchetti, University of Geneva

University of Geneva, Econometrics Dept., Geneva, 1205 Switzerland

Eva.Cantoni@metri.unige.ch

Key Words: *Cp, generalized estimating equations (GEE), prediction error, robustness, variable selection*

Variable selection is an essential part of any statistical analysis and yet has been somewhat neglected in the context of longitudinal data analysis. We propose a generalized version of Mallows's Cp (GCp) suitable for use with both parametric and nonparametric models. GCp provides an estimate of a measure of adequacy of a model for prediction. We examine its performance with popular marginal longitudinal models (fitted using GEE) and contrast results with what is typically done in practice: variable selection based on Wald-type tests. An application to real data further demonstrates the merits of our approach while at the same time emphasizing some important robust features inherent to GCp.

Selecting the Best Linear Mixed Model: The State of the Art

◆ Lloyd J. Edwards, University of North Carolina, Chapel Hill; Keith W. Muller, University of North Carolina, Chapel Hill; Paul W. Stewart, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Dept. of Biostatistics, 3105H McGavran-Greenberg, CB#7420, Chapel Hill, NC 27599-7420

Lloyd_Edwards@unc.edu

Key Words: *longitudinal data, model selection, repeated measures*

The linear mixed model for the analysis of repeated measures can be viewed as the multivariate extension of the univariate linear model. However, unlike standard univariate regression, there is comparatively little literature on model selection for the linear mixed model. We review the state of the art in model selection techniques for comparing parametric linear mixed models. We identify areas for which there are no existing model selection methods and areas where either an existing method has been proven to have undesirable properties or there is a lack of evidence supporting its adoption. The overall conclusion made from this review is that there is much left to do to improve the availability of model selection techniques for linear mixed models.

Model Selection for the Linear Mixed Model in which the Response is Transformed

◆ Matthew J. Gurka, University of Virginia; Lloyd J. Edwards, University of North Carolina, Chapel Hill; Pranab K. Sen, University of North Carolina, Chapel Hill

University of Virginia, Division of Biostatistics and Epidemiology, 1522 Trail View Lane, Durham, NC 27713

gurka@email.unc.edu

Key Words: *model selection, linear mixed model, longitudinal data, transformations, AIC, residual likelihood*

Shi and Tsai (2002) proposed a residual likelihood criterion (RIC) for univariate linear model selection using a residual likelihood approach to estimation. The selection criterion is based on the expected Kullback-Leibler information of the residual log-likelihood of the fitted model. The Box-Cox transformation was treated as a special case of a generalized version of the univariate linear model. The methodology of the RIC is extended to the linear mixed model for longitudinal data in which the response is transformed using the Box-Cox approach. Consistency properties of the RIC proved theoretically for the univariate case are shown to hold true for the transformed mixed model as well. Extensive simulations are performed to assess the performance of the RIC in selecting the "best" transformed mixed model for both large and small samples. Specifically, the ability of the RIC to select the correct fixed effects structure for a given mixed model with a transformed response in both large and small sample settings is examined, and we compare its performance to those of existing criteria used in the linear mixed model, such as the AIC and BIC.

Global Assessment of Organic Contaminants in Farmed Salmon

◆ Steven J. Schwager, Cornell University; Ronald A. Hites, Indiana University; Jeffery A. Foran, Citizens for a Better Environment; David O. Carpenter, University at Albany; M. Coreen Hamilton, AXYS Analytical Services; Barbara A. Knuth, Cornell University

Cornell University, Dept. of BSCB, 434 Warren Hall, Ithaca, NY 14853

sjs5@cornell.edu

Key Words: *farmed salmon, organochlorine contaminants, risk analysis, analysis of variance*

The annual global production of farmed salmon has increased by a factor of 40 during the past two decades. Salmon from farms in northern Europe, North America, and Chile are now available widely year-round at relatively low prices. Salmon farms have been criticized for their ecological effects, but the potential human health risks of farmed salmon consumption have not been examined rigorously. Having analyzed over two metric tons of farmed and wild salmon from around the world for organochlorine contaminants, we show that concentrations of these contaminants are significantly higher in farmed salmon than in wild. European-raised salmon have significantly greater contaminant loads than those raised in North and South America, indicating the need for further investigation into the sources of contamination. Risk analysis indicates that consumption of farmed Atlantic salmon may pose health risks that detract from the beneficial effects of fish consumption.

225 Problems in Spatial and Cluster Analysis

Biometrics Section, Section on Statistics in Epidemiology

Tuesday, August 10, 10:30 am–12:20 pm

A Comparison of Frequentist and Bayesian Spatial Regression Approaches

◆ Sanjeev Sridharan, Westat; Robert M. Baskin, Agency for Healthcare Research and Quality

Westat, 1650 Research Blvd., Rockville, MD 20850

sanjeevsridharan@westat.com

Key Words: *spatial regression, maximum likelihood, Mlwin, WinBugs, multilevel models, Gibbs sampling*

From 1994 to 2000, both crime rates and unemployment rates declined across most of the United States. The cause of such a decline in crime rates remains unclear to most criminologists. Using spatial panel data models, we attempt to link the trajectories of unemployment rates to crime rates. Annual data from Virginia counties and cities between 1994 to 2000 are used to study this relationship.This paper focuses on a comparison of frequentist to Bayesian spatial regression models. Specifically, we compare a spatial lagged panel model developed using the following three methods: maximum likelihood techniques developed within a spatial econometric framework; spatial multilevel model using iterative generalized least squares procedures in Mlwin based on empirical Bayes methods; Markov chain Monte Carlo methods using Gibbs sampling based on Bayesian methods using the WinBUGS software. Implications of the results for spatial epidemiology are discussed.

Segment of Plasma Membrane Image and Spatial Statistical Analysis of PM Sterol Distribution

◆ Weimin Zhang, Texas A&M University; Avery L. McIntosh, Texas A&M University; Barbara Atshaves, Texas A&M University; Friedhelm Schroeder, Texas A&M University; Jyh-Charn Liu, Texas A&M University; Suojin Wang, Texas A&M University

Texas A&M University, Dept. of Statistics, 3143 TAMU, College Station, TX 77843

zhangwm@stat.tamu.edu

Key Words: *segmentation, rank statistic, union intersection test, completely spatial randomness test, Monte Carlo simulation*

Although plasma membrane (PM) cholesterol-rich and -poor domains have been isolated by subcellular fractionation, it has been unclear whether such domains exist in real-time in living cells. Therefore, dehydroergosterol (DHE), a naturally occurring fluorescent sterol, is incorporated into cultured L-cell fibroblast and visualized in real-time in PM of living cells by multiphoton laser scanning microscopy (MPLSM). Rank statistics are used for segmenting PM area from MPLSM image of living cells. The Miller's rank statistic is calculated for combined two areas and the union intersection test is used for multivariate (color) data comparisons. Spatial distributions of DHE within the PM are examined by the complete spatial randomness (CSR) test and a polynomial regression analysis is run for prediction intervals. A Monte Carlo simulation study shows that the distributions of DHE appear in spatial clusters. These results for the first time demonstrate that sterol distributes clusterly in the membranes of living cells.

Weighted Covariance Universal Kriging

◆ Sankar Bokka, University of Mississippi; Jake Olivier, University of Mississippi; Hanxiang Peng, University of Mississippi

University of Mississippi, PO Box 9282, University, MS 38677

sbokka@olemiss.edu

Key Words: *spatial, covariance kriging, universal kriging, mean squared error*

Of the spatial interpolation and prediction techniques, kriging is the optimal among linear procedures, as it is unbiased and has minimal variance of prediction techniques. We generalize the work by Olivier and Peng (2003) and propose weighted covariance universal kriging. Admitting different weights yields various weighted covariance universal kriging estimators. Weight equating zero results in universal kriging estimators; weight equating infinity gives covariance universal kriging estimators, and weights between zero and infinity yield different weighted covariance universal kriging estimators in accordance with different amount of weight in the mean squared errors and the squared differences of the covariances.

Investigating Standard and Nonstandard Models for Remotely Sensed Cotton Insect Data

◆ Robert Downer, Louisiana State University

Louisiana State University, Dept. Experimental Statistics, 161 Ag. Administration Bldg., LSU, Baton Rouge, LA 70803-5606

rdowner@lsu.edu

Key Words: *smoothing, zero-inflated, Poisson, counts, spatial*

A remotely sensed vegetation index (ndvi) has been paired with ground truth cotton insect counts. Poisson and zero-inflated Poisson (ZIP) regression models have been used to model the relationship. However, the spatial nature of the data had not yet been well incorporated into the analysis and geostatistical methods seemed inapproppriate. Adjustments to these Poisson models and other models involving some smoothing techniques are considered.

Penalized and Weighted K-means

◆ George C. Tseng, University of Pittsburgh

University of Pittsburgh, 130 De Soto St. Dept. of Biostatistics, Pittsburgh, PA 15261

ctseng@pitt.edu

Key Words: *cluster analysis, unsupervised learning, K-means, scattered point, microarray*

Scattered points, i.e., points not belonging to any of the clusters of interest, have been found critical in cluster analysis. Most current clustering methods become inapplicable in complex large data such as gene expression profiles due to the failure of managing scattered points. A class of cost-function to be minimized is proposed for cluster analysis. Traditional K-means and K-memoids are special cases in this class of method. The cost-function involves a penalty term for dealing with scattered points and a weighted procedure for avoiding selection of pre-specified taboo patterns during calculation of average dispersion. The penalized cost-function is shown to be equivalent to the classification likelihood of Gaussian mixture model with scattered points uniformly distributed. The weighted procedure also has a Bayesian interpretation for avoiding identification of taboo patterns. Some properties of this class of method are discussed. The method is then used to replace K-means as a building block for the resampling-based tight clustering method. Simulations and applications have shown that it improves the performance and efficiency of original tight clustering.

226 Agreement and Association in Binary Data

Biometrics Section, ENAR, Section on Statistics in Epidemiology

Tuesday, August 10, 10:30 am–12:20 pm

Analysis of Clustered Binary Responses Using GEE Estimation with Bias Corrected Covariance Estimator

◆ Sadia Mahmud, Aga Khan University; Naseem K. Afridi, Aga Kahn University; Juanita Hatcher, Aga Khan University; Debra Nanan, Pacific Health & Development Sciences

Aga Khan University, Dept. of Community Health Sciences, PO Box 3500, Karachi, 74800 Pakistan

sadia.mahmud@aku.edu

Key Words: *binary, correlated, bias corrected*

The method of Generalized Estimating Equations (GEE) is a regression approach for correlated responses. GEE uses a robust estimator to estimate the covariance matrix of the regression coefficients. However, the robust estimator may be biased when the number of clusters is small. Mancl and DeRouen (2001) proposed a bias-corrected robust covariance estimator for regression coefficients. A cross-sectional study was conducted in 2002 in Peshawar, Pakistan, to assess the tetanus toxoid vaccination status among women of reproductive age. The sampling technique employed was stratified two-stage cluster sampling. A total of 40 clusters were selected. We analyzed the clustered binary responses using GEE estimation with bias-corrected covariance estimator, implemented by a GEE SAS macro written by Mancl. The working correlation structure specified was "exchangeable." The multiple logistic regression model contained seven covariates. For all the regression coefficients the bias corrected standard errors were larger than the robust standard errors, though the former and the corresponding p values were not markedly different. For only two covariates the bias corrected estimator (in conjunction with using t-distribution to compute confidence intervals for odds ratio) led to a different inference as compared to that from the robust estimator.

Bias-corrected Maximum Likelihood and Double Extended quasi-likelihood Estimators of the Intraclass Correlation Parameter for Binary Data

◆ Krishna K. Saha, University of Windsor

University of Windsor, 401 Sunset Ave., Windsor, ON N9B 3P4 Canada

saha5@uwindsor.ca

Key Words: *bias-corrected maximum likelihood, double extended quasi-likelihood, intraclass correlation parameter, over/underdispersion, toxicological data*

We derive expressions for Bias-corrected maximum likelihood (BCML) estimator of the intraclass correlation parameter for data in the form of proportions. A comparison, by simulation, of the BCML estimator, in terms of bias and efficiency, is conducted with two recently proposed estimators, namely, the double extended quasi-likelihood (DEQL) estimator proposed by Lee (2003) and the estimator Q2 proposed by Paul, Saha, and Balasooriya (2003), which is based on optimal quadratic estimating equations of Crowder (1987). The BCML estimator has superior bias and efficiency properties in most instances. Analyses of a set of toxicological data from Paul (1982) and a set of medical data pertaining to chromosomal abnormalities among survivors of the atomic bomb in Hiroshima from Otake and Prentice (1984) show in general much improvement in standard errors of the BCML estimates over other estimates.

Optimal Design for Interobserver Agreement Studies

◆ Mekibib Altaye, Cincinnati Children's Hospital Medical Center

Cincinnati Children's Hospital Medical Center, Center for Epidemiology and Biostatistics, University of Cincinnati, Cincinnati, OH 45229

altq56@CHMCC.ORG

Key Words: *power, interobserver, optimal, cost, agreement*

Power considerations are clearly important in choosing the number of subjects and raters that should be recruited for a study of interobserver agreement. However, most studies are also subject to budget constraints, requiring that the investigator achieve specified power while minimizing the total study cost C. This cost will typically depend on C1, the cost of recruiting a subject (n), C2, the cost of recruiting a rater (m) and C3, the cost of obtaining a measurement on a subject that has already been recruited. We introduce a linear cost function from which the optimal number of raters and subjects can be determined. The optimal values will be derived under the condition that the total cost for sampling and measurement is fixed at C, with the variance of the kappa statistics minimized subject to the restriction on cost given by the proposed model. The results of the minimization process indicate that for a fixed total cost C the effect of varying the allocation of costs among raters, subjects and measurements impacts the required number of subjects much more than varying the number of raters.

Assessing Interobserver Variability in Laser Safety Research

◆ Tzu-Cheg Kao, Uniformed Services University of the Health Sciences; Thomas Johnson, Uniformed Services University of the Health Sciences

Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814-4799

tkao@usuhs.mil

Key Words: *interobserver agreement, Kappa coefficient, risk, laser injury*

In a pre-clinical trial to assess safety in using laser instruments, three researchers independently evaluated the injury status of the skin. Three raters (#1-#3) participated in the evaluation and scoring process. The primary outcome, injury or no, was assessed. Measures were considered: Kappa coefficients, and the proportion of agreement. Either measure was to assess the agreement in injury scores. Results will be compared. Statistical issues related to the assessment of agreement will be discussed.

Rethinking the Adjustment for Chance in Agreement Coefficients

◆ Michael P. Fay, National Institute of Allergy and Infectious Diseases

National Institute of Allergy and Infectious Diseases, 6700A Rockledge Dr., Bethesda, MD

faym@mail.nih.gov

Key Words: *concordance correlation coefficient, kappa, reliability, weighted kappa*

We study the measurement of agreement or reliability between two different instruments. In the standard agreement coefficients (e.g., Kappa, weighted Kappa, concordance correlation coefficient) chance agreement is modeled using expected agreement between two independent random variables (RVs) each distributed according to the marginal distribution of one of the instruments. We propose that the adjustment for chance should also be modeled by expected agreement between two independent random variables, except that both RVs should be identically defined by the mixture distribution of the two marginal distributions. The advantage of the proposed adjustment for chance is that differences between the two marginal distributions will not induce greater apparent agreement. We call the proposed coefficients random marginal agreement coefficients (RMAC). An important RMAC may be estimated by the intraclass kappa, which has previously been proposed to measure agreement for nominal data. For ordinal and continuous data, the RMAC are new (to the best of our knowledge) and may provide more appropriate alternatives to the weighted kappa and concordance correlation coefficient.

227 Bayesian Applications and Case Studies

Section on Bayesian Statistical Science

Tuesday, August 10, 10:30 am–12:20 pm

Who is Connected to the Pittsburgh Jewish Community?

◆ Eloise Kaizar, Carnegie Mellon University; Stephen E. Fienberg, Carnegie Mellon University; Howard Seltman, Carnegie Mellon University

ekaizar@stat.cmu.edu

Key Words: *survey nonresponse, latent class models*

The United Jewish Federation of Pittsburgh commissioned the Pittsburgh Jewish Community Study to enumerate and characterize the Jewish population of the Pittsburgh region. We use the data from the study to answer the question: Which childhood activities are most effective at retaining Jewish children as engaged adult members of the Jewish community? The study collected data via a telephone survey of both known Jewish households and those discovered through random digit dialing. As with most modern telephone surveys, the response rate was low (42% overall), which possibly introduces bias into the usual methods of analysis. By modeling this nonresponse, we conclude there is no significant bias in the answer to the question of childhood activity effectiveness. Using logistic regressions and latent class models, we find evidence that childhood trips to Israel and participation in youth groups are most effective in retaining Jewish children as active adult members of the community. However, we must further refine the models used to confirm this finding.

Hierarchical Models with Migration, Mutation, and Drift

◆ Seongho Song, University of Connecticut; Dipak K. Dey, University of Connecticut; Kent E. Holsinger, University of Connecticut

University of Connecticut, Dept. of Statistics, U-4120, Storrs, CT 06269-4120

seongho@stat.uconn.edu

Key Words: *F-statistics, finite-island model, genetic drift, hierarchical population structure, migration, mutation*

Hierarchical structure arises naturally in genetic models. Individuals belonging to the same population are more similar to one another than are those belonging to different populations. Using properties of moment stationarity we develop exact expressions for the mean and covariance of allele frequencies at a single locus for the two-level hierarchical structure subject to drift, mutation, and migration. For arbitrary mutation and migration matrices, we generalize previous results to multilevel hierarchical model. Consequently, we have closed-form expressions for the mean and covariance of allele frequencies in Wright's finite-island model with constant hierarchical migration and several mutations. It turns out that the correlation among populations and among subpopulations and the correlation between populations and subpopulations vanish for the large size of population and subpopulation. Also we discuss some implications of our results based on Wright's F-statistics as measures of population structure.

Bayesian Analysis in Computer Experiments

◆ Stella W. Karuri, University of Waterloo

University of Waterloo, 680 East 8th Ave., Apt. 212, Vancouver, BC V5T 1T1 Canada

swkaruri@math.uwaterloo.ca

Key Words: *computer experiments, Bayesian analysis*

Mathematical models are commonly used in science and industry to simulate complex physical processes. These models are implemented by complex and expensive (in terms of run time) computer codes, limiting the number of simulation runs. A typical computer experiment involves running the code at a set of predetermined input levels to obtain a set of outputs. To analyze and obtain inference from the resulting output sets, the output is treated as a Gaussian stochastic process, a response surface is then fitted to the code. We outline a Bayesian formulation to this approach, prior uncertainty of the mean output, the variance and correlation between input is specified by means of the Jeffrey's prior. The Jeffrey's prior appeal is that it results in a proper posterior. Using a Gaussian correlation structure, we approximate the Jeffrey's prior on log-transformed correlation parameters with a uniform prior. The approximation method results in a posterior that is easier to approximate and cheaper to compute.

Comparisons using MCMC show that the approximation provides fairly similar results in predictions.

Bayesian Analysis of Dose-response Calibration Curves

◆ Bahman Shafii, University of Idaho; William J. Price, University of Idaho; Ken B. Newman, University of St. Andrews

University of Idaho, Statistical Programs, College of Agricultural and Life Sciences, Moscow, ID 83844-2337

bshafii@uidaho.edu

Key Words: *binomial response, logistic function, Bayesian estimation*

The statistical analysis of dose-response experiments typically models observed responses as a function of an applied dosage series. The estimated "dose-response curve" is used in predicting future responses, however, it is also commonly rewritten in an inverted form where dose is expressed as a function of the response. This modified "calibration curve" is useful in cases where observed responses are available, but their associated dosages are unknown. Traditional statistical techniques for the estimation of unknown doses from the dose-response curve are problematic, involving approximate solutions and methods. Alternatively, this type of inverse calibration problem naturally falls into the framework of Bayesian analysis. That is, one wishes to estimate the probability of an unknown dose value at an observed value of the response given the underlying relationship between the dose and response. This paper examines some potential Bayesian solutions to the calibration problem under various assumptive conditions. The required methodology in each case will be outlined for a dichotomous response variable and a logistic dose-response function.

Gene Expression Patterns during Somatic Embryogenesis in Maize Tissue Culture: A Bayesian Approach

◆ Tanzy M. Love, Iowa State University; Alicia Carriquiry, Iowa State University; Ping Che, Iowa State University; Stephen Howell, Iowa State University; Bronwyn Frame, Iowa State University; Kan Wang, Iowa State University

Iowa State University, 140 University Village Apt. E, Ames, IA 50010

tanzy@iastate.edu

Key Words: *gene expression, maize, somatic embryogenesis*

Regeneration in tissue culture is important for improvement of maize using genetic engineering approaches. However, many important maize lines are difficult to regenerate. The program of gene expression during somatic embryogenesis in tissue culture was examined in regeneration-proficient lines in the effort to find key genes that may limit regeneration in more recalcitrant lines. Somatic embryos were generated from eight embryogenic callus lines developed from immature Hi II embryo explants using protocols developed and utilized at the Plant Transformation Facility at Iowa State.The callus lines were pooled into three groups, and RNA was extracted from samples at six time points during incubation under culture conditions that induce embryo formation, maturation, and germination. RNA was hybridized to 12K gene cDNA microarray chips produced by the Microarray Facility in the Center for Plant Genomics at Iowa State. Statistical design, background cleaning, and normalization issues are explored. Bayesian methods are implemented to detect significant gene expression patterns.

A Bayesian Cross-classified Model for Estimating Teacher Effects on Student Achievement

◆ J. R. Lockwood, RAND Corporation; Daniel F. McCaffrey, RAND Corporation; Louis T. Mariano, RAND Corporation

RAND Corporation, 201 N. Craig St. Suite 202, Pittsburgh, PA 15213

lockwood@rand.org

Key Words: *value-added modeling, teacher effects, student achievement, cross-classified model*

In the education research and policy communities there is increased interest in using "value-added models," relying on longitudinal student-level test score data, to isolate the contributions of teachers to student academic achievement. The fact that students are linked to different teachers as they progress through grades imparts a cross-classified structure to the data that poses both substantive and computational challenges. We introduce a general multivariate, longitudinal model for student outcomes that incorporates these complex grouping structures, subsumes the principal existing modeling approaches, and explicitly allows for the estimation of the long-term effects of past teachers on student outcomes in future years. We present a Bayesian formulation of the model and discuss how the Bayesian framework affords numerous advantages relative to classical mixed-model approaches to value-added modeling. We demonstrate the model and the array of inferences about teacher effects that it facilitates using actual student achievement data.

228 Regression Models in Survival

WNAR, ENAR, Biometrics Section

Tuesday, August 10, 10:30 am–12:20 pm

Statistical Inference for the Additive Risk Model

◆ Yichuan Zhao, Georgia State University

Georgia State University, Dept. of Mathematics and Statistics, Atlanta, GA 30303

yzhao@mathstat.gsu.edu

Key Words: *counting process, normal approximation, partial likelihood, regression, right censoring, martingale*

Cox's proportional hazards model has been the most popular model for the regression analysis of censored survival data. However, the additive risk model provides a useful alternative. We apply empirical likelihood ratio method to additive risk model with right-censoring and derive its limiting distribution. Based on the result we construct a confidence interval for the regression parameter. Useful extensions are discussed. Simulation results are presented to compare the proposed method with that method based on normal approximation.

Censored Linear Regression for Case-cohort Studies

◆ Bin Nan, University of Michigan; Menggang Yu, Indiana University; Jack Kalbfleisch, University of Michigan

University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029

bnan@umich.edu

Key Words: *case-cohort design, censored linear regression, counting processes, martingales, rank statistic*

Two classes of estimates for regression parameters in a linear regression model are proposed in analyzing right censored data from a classical case-cohort design and a stratified case-cohort design, respectively. The estimates are shown to be consistent and asymptotically normal.

Estimation of a Weibull PHM with Time-dependent Stochastic Covariates from Incomplete Data Using a Variant of the ECME Algorithm

◆ Dragan Banjevic, University of Toronto

University of Toronto, 5 King's College Rd., Toronto, ON M5S 3G8 Canada

banjev@mie.utoronto.ca

Key Words: *proportional hazards, Markov process, stochastic covariates, incomplete data, EM algorithm, reliability*

A method for estimation of parameters of a proportional-hazards model (PHM) with time-dependent stochastic covariates from incomplete covariate data is discussed. It is assumed that the failure-time process and the covariate process follow a nonhomogeneous multidimensional Markov process. A variant of the ECME algorithm of Liu and Rubin for simultaneous estimation of both the PHM parameters and Markov process transition rates is considered. Convenient recursive formulas are derived necessary to perform E-step in the calculation, as well as to calculate the observed information matrix, required for direct minimization in Newton-Raphson method and for estimation of the standard errors. The recursive formulas also include solution to an extremely important problem in the calculation, that is, to the scaling of very small probabilities and expectations that may cause underflow in the computation of the likelihood function. The method is applied to real data from industry obtained from equipment condition monitoring and associated failure times.

Estimating the Mixing Proportion in a Semiparametric Mixture Model from Censored Time-to-event Data

◆ Xi Zhang, Columbia University

Columbia University, 362 Riverside Dr., Apt. 8B8, Dept. of Statistics, New York, NY 10025

xz2012@columbia.edu

Key Words: *efficiency, censoring, counting processes, Fredholm integral equations*

In the semiparametric mixture model in treating $H=\theta*F+(1-\theta)*G$, a sample from a mixture of the two unspecified distributions is obtained together with separate samples from each of the two mixing distributions. Interest is in estimating the mixing proportion. Here, the problem is extended to allow for right-censored outcomes. A family of estimating equations is developed, and the optimal member of the family is derived. It is shown that the solution to the optimal estimating equation achieves the semiparametric information bound. An approach to estimating the optimal estimating equation is developed and investigated through simulation experiments.

A Pseudo-partial-likelihood Method for Semiparametric Survival Regression

◆ David M. Zucker, Hebrew University

Hebrew University, Dept. of Statistics, Mt. Scopus, Jerusalem, 91905 Israel

mszucker@mscc.huji.ac.il

Key Words: *Cox regression model, proportional hazards, errors in variables*

An estimator is presented for the regression coefficient vector in the Cox proportional hazards model with covariate error. The estimator is obtained by maximizing a likelihood-type function similar to the Cox partial likelihood. The likelihood function involves the cumulative baseline hazard function, for which a simple estimator is substituted. The method is capable of handling general covariate error structures and can be applied to studies with either an external or internal validation sample. The estimator is shown to be consistent and asymptotically normal, and an estimate of the asymptotic covariance matrix is derived. Some extensions to general transformation survival models are indicated. Two simulation studies are presented: one with a single error-prone binary covariate and one with a single error-prone, normally distributed covariate. These simulation studies show that typically produces estimates with low bias and confidence intervals with accurate coverage rates. Efficiency results relative to fully parametricmaximum likelihood are also presented.

229 Variance Estimation II ▲

Section on Survey Research Methods, Section on Government Statistics

Tuesday, August 10, 10:30 am–12:20 pm

A Generalization of the Coefficient of Variation with Application to Suppression of Imprecise Estimates

◆ Avinash Singh, RTI International; Matthew Westlake, RTI International; Moshe Feder, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

asingh@rti.org

Key Words: *effective sample size, exact binomial, length of confidence interval, length of discrimination interval*

We propose a generalization of the usual coefficient of variation (CV) to address some of the known problems when used in criteria developed to determine the suppression of estimates. Some of the problems associated with the CV in this application include interpretation when the estimate is near zero, and the inconsistency in the interpretation about precision when computed for different one-to-one monotonic transformations. The proposed measure, termed discrimination coefficient of variation (DCV), generalizes the CV using hypothesis-testing ideas involving length of the confidence interval (LCI) and the length of a discrimination interval (LDI). This discrimination interval is used to check whether the sample is large enough or the confidence interval is short enough to discriminate with certain power between the current value and postulated extreme values on either side of the current value. DCV allows for using an exact distribution for computing LCI such as the use of exact binomial when the estimated proportion is very small. The method is illustrated using data from the 2002 NSDUH (National Survey on Drug Use and Health) sponsored by SAMHSA.

Comparison of Variance Estimators under Rao-Sampford Method: A Simulation Study

◆ Fulvia Mecatti, University of Milan-Bicocca; David Haziza, Statistics Canada; Jon N.K. Rao, Carleton University

University of Milan-Bicocca, Dept. of Statistics, Via Bicocca degli Arcimboldi, 8 - Ed U7, Milan, 20126 Italy

fulvia.mecatti@unimib.it

Key Words: *approximate joint inclusion probabilities, efficiency, relative bias, unequal probability sampling without replacement*

Rao-Sampford method of unequal probability sampling without replacement has several desirable properties: inclusion probabilities are exactly proportional to sizes, Sen-Yates-Grundy exactly unbiased variance estimator is non-negative for any sample size, and variance of the estimator is always smaller than the corresponding variance under unequal probability sampling with replacement. Moreover, exact joint inclusion probabilities, and hence the exact variance estimate, can be readily calculated for any sample size, using SAS software. But several approximate variance estimators, based on approximating the joint inclusion probabilities in terms of the first order inclusion probabilities only, have been proposed in the literature. Using simulations, we first compare the relative bias (RB) of the approximate variance estimators, by varying sample size, the population size and the coefficient of variation of the auxiliary variable. We then select those that perform well in terms of RB and compare them with the exact variance estimator in terms of mean squared error.

Analysis of Census 2000 Long Form Variances

◆ Eric Schindler, U.S. Census Bureau

U.S. Census Bureau, Washington, DC 20233

eric.l.schindler@census.gov

Key Words: *raking, successive difference, generalized variance*

The Census 2000 Long Form consisted of a 15% systematic sample of all households. Some 60,000 weighting areas were formed and final weights were calculated using an iterative proportional fitting (raking) procedure. Both the systematic sample and the raking are designed to reduce variance. A successive difference replication (SDR) methodology was developed to estimate the variances in order to properly reflect the variance gains of the systematic sample. Due to resource and timing constraints, the weights were not calculated by raking for each replicate. For each replicate the final production weights were simply multiplied by the appropriate SDR adjustment factors. Thus, the reduction in variances due to raking was not reflected in the variance estimates which, as a result, should overestimate the true variance. Standard errors were estimated for a large number of subtotals in each weighting area, outliers were removed, and generalized design factors were calculated for 60-item groups for four observed sampling rates for each state. This paper examines the consequences of not raking to reweight each replicate by comparing variance estimates with or without the replication of the raking procedure for the SDR method and for a simple jackknife method.

Consumer Expenditure Variance Estimation Incorporating Generalized Variance Functions in Hierarchical Bayes Models

◆ Paul Hinrichs, Peru State College

Peru State College, PO Box 10, Peru, NE 68421

phinrichs@oakmail.peru.edu

Key Words: *consumer expenditure, generalized variance, hierarchical Bayes*

This paper investigates models for sampling error that would improve the stability of mean annual consumer expenditure estimates. Hierarchical Bayes models presented consider mean estimates from five interview and diary sources for diary only estimates when interview estimates are also available. The investigation breaks down the problem into improving the GVF estimate of variance through three-digit classification of UCCs, estimation of sampling variance using hierarchical models and mean expenditure estimation. Model comparison and variance of variance estimates are demonstrated.

Effects of Grouping Continuous Data on First and Second Distribution Moments

◆ Myron Katzoff, National Center for Health Statistics; Jay J. Kim, National Center for Health Statistics; Joe Fred Gonzalez, Jr., Centers for Disease Control and Prevention; Lawrence H. Cox, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

mjk5@cdc.gov

Key Words: *disclosure risk, interval-data, variance, class mark, data summarization, midpoint*

Data such as income are often grouped and released as interval data, considered to be one of the best ways of summarizing data which has disclosure risk implications as well. Class marks (midpoints) of intervals are then used to calculate the mean and variance of the grouped data. In most situations, using midpoints for every observation in the interval smoothes the data, thereby reducing the variance. It can be shown, as in analysis of variance, that, using midpoints, we lose the within-interval variance component if within-interval data have a uniform distribution. However, if distributions within some intervals are peaked or skewed, use of the midpoints of the interval data can result in higher variance estimates than would be obtained with the raw data. Moreover, for those data, the mean of the grouped data based on the use of midpoints is biased. If class (conditional) means are used for calculating overall mean and variance, the mean of the raw data can be recaptured and the variance will be lower. We report some initial results from our study of the impact of accepted practices for approximating moments with summarized data.

Imputation Variance Estimation by Bootstrap Method for the National Ambulatory Medical Care Survey

◆ Yan Li, National Center for Health Statistics; Carrie Lynch, National Center for Health Statistics; Iris Shimizu, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 3206, Hyattsville, MD 20782

YCL5@cdc.gov

Key Words: *complex sample survey, imputation, variance, bootstrap*

The National Ambulatory Medical Care Survey (NAMCS) is a three-stage, stratified, cluster sample survey of visits to office-based physicians in the U.S. The NAMCS measures the utilization of ambulatory medical care services across various types of physicians. Four variables in the current NAMCS public use files are imputed when item nonresponse occurs. The variable “time spent with physician” is imputed using a hot-deck procedure, and the variables sex, race, and age are imputed using a cold-deck procedure. Variance approximations for estimates involving the imputed variables should account for the variation due to imputation. This paper discusses a bootstrap method that was used to assess the magnitude of imputation variance for the 2000 NAMCS and the results.

Statistical Methodology for the Census 2000 Public Use Microdata Samples

◆ Philip Gbur, U.S. Census Bureau; Mary Frances E. Zelenak, U.S. Census Bureau

U.S. Census Bureau, 7975 Central Park Cir., Alexandria, VA 22309

philip.m.gbur@census.gov

Key Words: *sampling, disclosure avoidance, variance estimation*

Public Use Microdata Samples (PUMS) are data files that contain individual housing unit and person records with their associated characteristics. To protect confidentiality, identifying information and detailed geography is excluded. To allow data users to produce any tabulations of interest and to conduct detailed analyses, such as regression analysis and investigation of the relationships among variables, the U.S. Census Bureau created PUMS from decennial census results. Subject to the limitations on sample size and geographic identification, the user may generate tabulations interrelating any desired set of variables. This paper presents the statistical methodologies used in sampling, disclosure avoidance, and variance estimation for the United States, Puerto Rico, Guam, and U.S. Virgin Islands Census 2000 PUMS.

230 Design and Model Selection △ ⊞

IMS

Tuesday, August 10, 10:30 am–12:20 pm

Fitting Main Effect Polynomials in Multifactor Experiments: Unequal Spacing and Unbalanced Data

◆ Charles J. Monlezun, Louisiana State University

Louisiana State University, Dept. of Experimental Statistics, 143 Ag. Adm. Bldg., Baton Rouge, LA 70803

cmonlezu@lsu.edu

Key Words: *cell means, orthogonal decomposition, Gram-Schmidt, weighted averages, sequential sums, regression*

When the factor of interest arises from distinct levels of a quantitative variable, the main effects for that factor may be expressed as a polynomial function of the values of that quantitative variable. The main effect sums of squares for that factor may be partitioned into individual one-degree-of-freedom sums of squares for determining the actual degree of the main effect polynomial function. These one-degree-of-freedom sums of squares are shown to be the sequential sums of squares from the ordinary regression (with no intercept) of the observation vector on specially constructed independent variables. Main effects may be arbitrary weighted averages of cell means. Only the factor of interest need be associated with a bona fide quantitative variable. Construction of the independent variables and implementation of the no intercept regression require no special software, and may be performed using the regression procedure from any standard statistical computing package.

An Optimality Result for a Cross-validation Selection Criterion

◆ Andrew Neath, Southern Illinois University, Edwardsville; Joseph Cavanaugh, University of Iowa; Simon Davies, Centocor, Inc.

aneath@siue.edu

Key Words: *Kullback divergence, Akaike information*

Consider a model selection problem where one looks to decide which input variables are needed in developing the best prediction function. Cross-validation procedures have been seen to work well in practice here. We show how cross validation can be presented using a discrepancy function approach. As an example, the predictive divergence criterion (PDC) is introduced. The determination of the minimum variance unbiased estimate of the target discrepancy leads to an optimal selection criterion under conditions set forth by this cross-validation procedure.

An Algebraic Construction of D-optimal Designs for Weighted Polynomial Regression

◆ Fu-Chuen Chang, National Sun Yat-sen University

National Sun Yat-sen University, 70 Lien-hai Rd., Dept. of Applied Mathematics, Kaohsiung, 804 Taiwan

changfc@math.nsysu.edu.tw

Key Words: *approximate D-optimal design, Chebyshev system, differential equation, matrix, rational function, weighted polynomial regression*

We propose an algebraic construction of (d+1)-point D-optimal designs for d^{th} degree polynomial regression with weight function $\Omega(x)>0$ over an arbitrary compact region [a,b]. Suppose that $\Omega(x)/\Omega(x)$ is a rational function and the information of whether the optimal support contains the boundary points a and b is available. Then the problem of constructing (d+1)-point D-optimal designs can be transformed into a differential equation problem leading us to a certain matrix including a finite number of auxiliary unknown constants, which can be solved from a system of polynomial equations in those constants. Moreover, the (d+1)-point D-optimal interior support points are the zeros of a certain polynomial which the coefficients can be computed from a linear system. In particular, if the functions {1,Ω(x),Ω(x), \ldots, Ω(x)x^{2d}} is a Chebyshev system on [a,b], then the proposed systematic procedure gives the D-optimal designs.

Efficient Cross-over Designs for Comparing Test Treatments with a Control Treatment

◆ Min Yang, University of Nebraska, Lincoln

University of Nebraska, Lincoln, 103 Miller Hall, Lincoln, NE 68583-0712

myang2@unl.edu

Key Words: *cross-over design, repeated measurements, treatments vs. a control, A-optimality, carryover effects*

Comparing t test treatments with one control treatment in cross-over design is considered when the number of periods is less than number of treatments t+1. The currently available result is constructing such designs in a restricted class, i.e., no treatment is allowed to follow itself and the control treatment is uniform in periods. We will present a new result to construct A-optimal designs for some parameters under a much less restricted class, i.e., treatment is allowed to be immediately followed by itself. It also showed that the currently available designs are highly efficient under nearly whole class for those parameters.

Optimal Sequential Planning Techniques

◆ Claudia Schmegner, DePaul University; Michael Baron, University of Texas, Dallas

DePaul University, Mathematics Dept., 2320 N. Kenmore Ave., Chicago, IL 60614

cschmegn@condor.depaul.edu

Key Words: *sequential procedures, cost function, Bayes stopping time, sequentially planned probability ratio test, sufficiency, randomization*

Sequential planning extends and generalizes the classical sequential schemes. Observations are sampled in groups of variable sizes, rather than one by one. A sampling strategy, in terms of batch sizes and the number of batches, is determined sequentially. We discuss optimality of sequential plans under a suitable risk function that, besides the total cost of collecting the observations, includes a new term consisting of the costs for each sampled batch. It is shown that only nonrandomized sequential plans based on a sufficient statistic have to be considered in order to achieve Bayes optimality. Sequential planning on a lattice is considered in detail, and methods of risk evaluation and comparison are derived.

Fractional Error Analysis in Experiments with Multiple Numerical Inputs

◆ George Gardenier, University of Virginia

University of Virginia, 115 St. Andrews Dr. NE, Vienna, VA 22180

ggardenier@virginia.edu

Key Words: *error analysis, multiple inputs, proportional error*

The objective of this presentation is to demonstrate the framework for an interactive quantification of proportionate contributions to the total error of an experiment. Where a number of different related measurements are obtained with various methods, the individual contribution of each measurement to the total error is displayed. The reseacher can then prioritize the necessity for improving the precision of each type of measurement. The software was developed using Microsoft Excel, based on a similar program written in QuattroPro.

231 Statistical Modeling of Earth and Atmospheric Processes ▲

Section on Statistics and the Environment

Tuesday, August 10, 10:30 am–12:20 pm

Least Squares Estimation and ANOVA for Periodic Autoregressive Time Series

◆ Qin Shao, University of Toledo

University of Toledo, OH 43606

qin.shao@utoledo.edu

Key Words: *periodically correlated time, least squares estimation, analysis of variance*

The periodic correlation exists throughout the whole process in a ANOVA-type model where the error terms consist of a periodic autoregressive time series. This paper studies the asymptotic property of least squares estimators and linear testable hypotheses with a modified F test in the analysis of variance akin to periodic autoregressive series. The techniques are applied in making inference on the quarterly streamflow in Asotin, WA.

Space-time Modeling of Groundwater Monitoring and Remediation

◆ Xiaodong Li, University of Illinois, Urbana-Champaign; Barbara A. Bailey, University of Illinois, Urbana-Champaign

University of Illinois, Urbana-Champaign, Dept. of Statistics, 725 S. Wright St., Champaign, IL 61820

xli12@stat.uiuc.edu

Key Words: *space-time modeling, environmental monitoring*

Monitoring ground water and subsurface remedial actions for chemicals such as tritium at sites such as Argonne National Laboratory is an important area of environmental engineering research. The Department of Energy and other federal agencies are making significant investment in the development of field analytical techniques, nonintrusive technologies, and sensor technologies that will impact the way environmental monitoring is conducted. We present a hierarchical space-time model to integrate sources of process knowledge at a site, including geological data, measurements collected during routine groundwater monitoring, and scientific understanding.

Use of Power Prior Distributions to Environmental Data Analysis

◆ Keying Ye, Virginia Polytechnic Institute and State University; Yuyan Duan, Virginia Polytechnic Institute and State University; Eric P. Smith, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics, 410 Hutcheson, Blacksburg, VA 24061

keying@vt.edu

Key Words: *power prior distribution, Bayesian method, environmental statistics, water monitoring*

The elicitation of power prior distributions is based on the availability of historical and other information, and is realized by incorporating a certain power of the likelihood function of the historical data. A modified approach is proposed to construct power prior distributions, and the general conditions for propriety of posterior distributions are examined. We demonstrate the development of the power priors for Bernoulli and normal models, and furthermore discuss the properties of resulted posterior distributions for the parameters of interest as well as for the power parameter. An example of water quality assessment data is presented to illustrate the implementation of power priors under this approach.

Empirical Modeling of Limnetic Temperature Profile and Estimating Thermocline Depth

◆ Victor Chan, Western Washington University

Western Washington University, Dept. of Mathematics, 214 Bond Hall, Bellingham, WA 98225

Victor.Chan@wwu.edu

Key Words: *nonlinear regression, confidence intervals, correlated errors*

A reasonably precise estimate of the thermocline depth or other well-defined boundaries in the temperature profile of a lake is of interest to limnologists. I propose an empirical model based on a statistical distribution to describe the temperature profile. Using the model, the thermocline depth and other well-defined boundaries could easily be estimated. The nonlinear Gauss-Newton method is used to fit the model to the profile data. Examples and the model-fitting procedure are illustrated using the data from Lake Whatcom, WA. I will also present a method to construct an approximate confidence interval for the thermocline depth based on the estimated asymptotic covariance matrix for the OLS estimators.

Detecting Shifts in Hurricane Rates Using a Markov Chain Monte Carlo Approach

◆ Xu-Feng Niu, Florida State University

Florida State University, Dept. of Statistics, Tallahassee, FL 32306/4330

niu@stat.fsu.edu

Time series of annual hurricane counts are examined using a changepoint analysis. The approach simulates posterior distributions of the Poisson rate parameter using Gibbs sampling. A posterior distribution is a distribution of a parameter conditional on the data. The analysis is first performed on the annual series of major North Atlantic hurricane counts from the 20th century. Results show significant shifts in hurricane rates during the middle 1940s, the mid-1960s, and at 1995, consistent with earlier published results. The analysis is then applied to U.S. hurricane activity. Results show no abrupt changes in overall coastal hurricane rates during the 20th century. In contrast, the record of Florida hurricanes indicates downward shifts during the early 1950s and the late 1960s. The shifts result from fewer hurricanes passing through the Bahamas and the western Caribbean Sea. No significant rate shifts are noted along either the Gulf or East coasts. Climate influences on coastal hurricane activity are then examined. Results show a significant reduction in U.S. hurricane activity during strong El Niño events and during the positive phase of the North Atlantic Oscillation.

A Flexible Computer-efficient Class of Spatial Interaction Models and Its Application to Ocean Remote-sensing and Ocean-mapping Data

◆ Ernst Linder, University of New Hampshire

University of New Hampshire, Dept. of Mathematics and Statistics, M 303 Kingsbury Hall, Durham, NH 03824

elinder@cisunix.unh.edu

Key Words: *lattice models, spatial-temporal data, mixed effects models, remote-sensing, ocean-mapping*

We propose a superclass of Gaussian spatial interaction models that includes as special cases CAR and SAR-like models. This class is, unlike traditional CAR models, able to capture spatially smooth phenomena. It is comparable to the Matern class in geostatistics, but inherits the computational advantages of interaction models for lattice data. This model is defined via a spatial structure removing orthogonal transformation. The latter is a one-time preprocessing step and greatly simplifies estimation particularly when it is simulation-based (MCMC). We extend our earlier work, which was within the Bayesian framework, and examine the possibility of frequentist estimation procedure. After transformation this leads to linear mixed effect models with structured variances for the random effects. We discuss estimation and computational issues. We show how this model is applied to large gridded space time in ocean-remote-sensing and ocean-mapping.

232 Robustness and Mixtures

Section on Physical and Engineering Sciences

Tuesday, August 10, 10:30 am–12:20 pm

Simulations for Robust Tests of Many Outliers in Regression

◆ Anthony C. Atkinson, London School of Economics

London School of Economics, Dept. of Statistics, Houghton St., London, WC2A 2AE UK

a.c.atkinson@lse.ac.uk

Key Words: *forward search, simulation envelopes, trimmed estimators*

Tests of outliers in regression need estimates of both the parameters of the linear model and of the error variance. If the outliers are included in the set of observations used for estimation, inconsistent estimates of the parameters will be obtained and the existence and the effect of the outliers will be masked. We therefore consider procedures in which the observations are divided into two groups; the good observations are used to provide estimates of the parameters and the remainder are tested for outlyingness. We are interested in the null distribution of the outlier test. In the simplest case, that of location, we would base our estimates on the m central observations, trimming the remainder. The properties of our estimators would then be those coming from this trimmed sample of n observations. We use this insight, combined with the Forward Search, to provide three excellent approximations to the distribution of the outlier test in regression. The talk will emphasize robust data analysis.

Integration of Detailed and Quick Simulations via a Bayesian Synthesis

◆ Zhiguang Qian, Georgia Institute of Technology; Roshan J. Vengazhiyil, Georgia Institute of Technology; C.F. Jeff Wu, Georgia Institute of Technology

Georgia Institute of Technology, School of Industrial and Systems Engineering, ISYE, Atlanta, GA 30332

zqian@isye.gatech.edu

Key Words: *response surface model, computer experiments, design of experiments, Gaussian process, meta-model, finite elements analysis*

This talk is motivated by collaborative work on robust topology design of cellular material at Georgia Tech. In simulating the material properties finite elements analysis (FEA) can be done based on different physical-mechanistic models. Typically, a more detailed or accurate model will require longer FEA runs while a simplified or rough model will require quicker FEA runs. They are referred to as detailed and quick simulations respectively. Detailed simulations can take up days of CPU time. While they can provide more accurate results, their number can be limited. On the other hand, many quick simulations can be obtained, though the results are less reliable. A new approach is taken here to combine these sources of data to come up with a meta-model that can be used to describe the relationship between the output of FEA runs and input parameters and for prediction. Since the quick simulations form the bulk of the data, they are used to build a semiparametric model based on Gaussian random functions. This fitted model is then “adjusted” by incorporating the information in the detailed simulations. Real data will be used to illustrate this technique.

Online Changepoint Methods for Time Series

◆ Bonnie K. Ray, IBM T. J. Watson Research Center; Hyunyoung Choi, University of Illinois, Urbana-Champaign; Hernando Ombao, University of Illinois

IBM T. J. Watson Research Center, Dept. of Mathematical Sciences, PO Box 218, Yorktown Heights, NY 10598

bonnier@us.ibm.com

Key Words: *spectral analysis, changepoint detection*

We present a new, spectral-based method for detecting changes in correlation structure in a continuous-valued time series. The test statistic can be efficiently computed and evaluated, making it appropriate for application in an online environment. Using simulation, we compare the proposed spectral-based method to existing spectral and parametric-based methods in terms of their performance in an online monitoring framework. We then present an application to monitoring of CPU metrics, for goals such as automated optimization of power consumption.

Detecting Multiple Populations within a Collection of Repairable Systems

◆ Jeffrey Glosup, Sun Microsystems Inc.

Sun Microsystems Inc., 18 Network Cir., Mail Stop UMPK18-216, Menlo Park, CA 94025

jeff.glosup@sun.com

Key Words: *mean cumulative function, repairable system reliability, mixture model*

The recurrence rate of failures in a collection of repairable systems can be studied nonparamentrically with the mean cumulative function. However, this representation of average system behavior is misleading if the rate of occurence of failures is not the same for all systems. In practice, there may be a subset of systems which fail at a much higher rate than the majority of systems. This can be modeled as a mixture of distributions for time to next failure, where each system follows a single distribution. Methods for detecting the presence of such a mixture of distributions are developed and explored. Methods used to categorize individual systems from a collection into multiple groups are also discussed.

Using Censored Data to Test the Location Parameters of Several Exponential Distributions

◆ Jeffrey J. Green, Ball State University; Roger B. Nelson, Ball State University

Ball State University, Economics Dept., Muncie, IN 47306

jgreen@bsu.edu

Key Words: *exponential distribution, location parameter, censored data, hypothesis testing*

A likelihood ratio procedure for testing the equality of the location parameters of two or more two-parameter exponential distributions is presented. This test is based on data censored from above and is constructed under the assumption that the scale parameters of the various distributions are all equal but unknown. The power function for this test is also determined, and a table of critical values for typical situations is given. When testing just two location parameters and the two sample sizes are equal, the test reduces to one recommended by Kumar and Patel. The power function for this latter procedure is also obtained, and it is shown that the likelihood ratio test is just as often as not the more powerful. Besides, Kumar and Patel's technique is shown to be biased.

On the Use of Machine Learning in the Semiconductor Industry: Examples and Case Studies

◆ Theresa Utlaut, Intel Corporation; Kevin Anderson, Intel Corporation

theresa.l.utlaut@intel.com

Key Words: *machine learning, semiconductor*

Semiconductor manufacturers must wring all available improvements in the quality and yield of their products. Some questions an engineer or statistician must ask when confronted with optimizing a 400-operation process are: At which operation to begin? What factors in that operation influence what responses in the process? Have all available opportunities for optimization been exhausted? Huge quantities of data are generated on production processed in a semiconductor facility. However, most of these data are observational in nature. Complex structure, multicollinearities, and sparsity are rampant in these data, and compromise the ability of classical statistical methods to answer the engineer's questions. Machine Learning can be defined as using algorithms that improve their performance by the analysis of data rather than relying on the analyst. Popular methods include Binary Recursive Partitioning, Stochastic Gradient Boosting, and Random Forests. This presentation will offer some theory behind these methods, and discuss their application in semiconductor manufacturing using software named IDEAL. Case studies will demonstrate the utility of these approaches in data analysis.

Comparative Analysis for Artificial Neural Networks and Multiple Linear Regression for Powder Hardfacing Processes

◆ Shu-Yi Tu, University of Michigan, Flint; Jen-Ting Wang, SUNY, Oneonta; Ming-Der Jean, Yung-Ta Institute of Technology & Commerce

University of Michigan, Flint, MI 48502

sytu@umflint.edu

Key Words: *artificial neural networks, multiple linear regression, Taguchi methods, powder hardfacing*

Artificial neural networks (ANNs) are an information processing system with the ability to learn, recall, and generalize from training data. Experiments involving numerous variables are usually analyzed using the analysis of variance (ANOVA) or multiple regression methods. We compare ANNs and the various multiple linear regression analysis in a powder hardfacing process that involves eight variables. The experimental results show that the ANN model effectively performs a better accuracy at predicting the response variable than the multiple linear regression models.

233 Regular Contributed Posters

Business and Economics Statistics Section, Biometrics Section, Section on Risk Analysis, Section on Statistics in Epidemiology, Section on Government Statistics, Section on Quality and Productivity, Section on Risk Analysis, Section on Physical and Engineering Sciences, Social Statistics Section, General Methodology, IMS, ENAR

Tuesday, August 10, 12:00 pm–1:50 pm

Comparisons of the Stein-type and Nonparametric Shrinkage Predictions

◆ Xue Xin, Tulane University; Leann Myers, Tulane University

Tulane University, 71 Centre St., Nutley, NJ 07110

xxin@tulane.edu

Key Words: *shrinkage, prediction, MSEP, PRESS*

Different forms of shrinkage correction have been proposed by researchers when the regression model is used for prediction. The Stein-type and nonparametric shrinkage predictors were compared using resampling and sample-reusing techniques. P (number of predictors), n (sample size), beta, and the correlation structure of the VC matrices were varied. Both the mean and median of the mean squared error for prediction (MSEP) and PRESS were generated from 10,000 simulations (5,000 for PRESS) and the results were compared. Shrinkage predictions gave a uniformly lower MSEP/PRESS than the least squares predictions. The percentage of loss saving went up to 35% under certain conditions. Furthermore, the Stein-type usually gave a lower MSEP/ PRESS than the nonparametric form. The percentage of loss saving decreased as the sample size increased or the number of covariates decreased. The percentages of loss saving were fairly comparable when the covariates were either independent or had low (pairwise r = 0.1) to moderate (pairwise r = 0.4) correlation. The findings were the same whether means or medians of MSEP and PRESS were considered.

Intraindividual Variation in Serum Constituents: A Comparison of Statistical Methods

◆ Cathleen Gillespie, Centers for Disease Control and Prevention; Ralph Donehoo, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, Division of Nutrition and Physical Activity, 4770 Buford Hwy., NE, Mailstop K-26, Atlanta, GA 30341-3724

CGillespie@cdc.gov

Key Words: *Bayesian estimation, repeated measures ANOVA, linear models, coefficient of variation, intraindividual variation, serum*

The coefficient of variation (CV) is the accepted method of quantifying and reporting the laboratory imprecision of serum measurements. Statistical methodology of assessing intraindividual variance in serum constituents varies widely. We examined three methods of assessing intraindividual variance: repeated measures ANOVA, Bayesian estimation, and the observed CV with a linear model to adjust for the regression to the mean effect. We examined these methods in three serum constituents: total cholesterol, retinol, and iron. We compared the variance estimates to those observed among participants in NHANES III with two blood draws, and assessed the effect of the variance estimates on the population distribution and confidence intervals. Bayesian estimation and the linear model allowed for the regression to the mean effect and produced reasonable variance estimates for all three constituents examined. The ANOVA method yielded implausible variance estimates for the constituents. Given the flexibility of Bayesian analysis afforded by the choice of prior distributions, this method could be readily adapted to fit the distribution of a wide variety of serum constituents.

An Algorithm to Construct All 2**n Sequences of Sibling Pairs with Application to Families with Two Monozygotic and One Dizygotic Triplets

◆ Nina Kohn, North Shore LIJ Research Institute; Gene S. Fisch, North Shore LIJ Research Institute; Catherine Derom, Catholic University, Leuven; Martin L. Lesser, North Shore LIJ Research Institute

North Shore LIJ Research Institute, 1129 Northern Blvd., Suite 302, Biostatistics Unit, Manhasset, NY 11030

nina@nshs.edu

Key Words: *intraclass correlation, twins, triplets, Cartesian product, algorithm*

The intraclass correlation (ICC) is commonly used to compute the correlation between k>=2 unordered measures such as IQ. In monozygotic (MZ) triplets, there are three indistinguishable persons and the ICC for k=3 is easily applied. However, in the case of triplets where there are two MZs and one dizygotic twin (DZ), the MZs are distinguishable from the DZ, but not from one another, rendering the usual ICC calculation invalid. For n families of such triplets, we propose a measure of ICC, whereby all 2**n within-family pairwise combinations of one of the two MZs with the single DZ are constructed (a "restricted" Cartesian product) and the standard ICC for k=2 is computed. The mean of those 2**n ICCs is then used to measure correlation between the MZs and the DZ. Construction of datasets consisting of all of these possible combinations is nontrivial. We developed an efficient algorithm using SAS to enumerate all 2n samples. The algorithm is general: it can be used with any number of triplet sets. It can also be used to enumerate all samples constructed by selecting 1 of k siblings or k-1 of k siblings.

More on Tests for Proportion

◆ Wei-Min Huang, Lehigh University

Lehigh University, 14 East Packer Ave., Dept. of Mathematics, Bethlehem, PA 18015

wh02@lehigh.edu

Key Words: *Bernoulli trials, tests for proportion, normal approximation, correction*

Standard tests for proportion based on Bernoulli trials were examined in my previous preliminary study. Additional results are developed in this study. This article examines the exact probability of Type I error and the exact power of these standard tests. We also propose a modified test which has better control on the approximate probability of Type I error. Some additional approximations are also introduced.

The Use of Discrete Data in the PCA: Theory, Simulations, and Applications to the Socioeconomic Indices

◆ Stanislav Kolenikov, University of North Carolina, Chapel Hill; Gustavo Angeles, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Dept. of Statistics and Carolina Population Center, CB#3260, UNC, Chapel Hill, NC 27599-3260

skolenik@unc.edu

Key Words: *welfare index, principal components, polychoric correlation, discrete data, Monte Carlo simulation*

We analyze construction of a welfare index when a more desirable income or expenditure data are not available. The principal component analysis is used to find the weights of variables aggregated into a socioeconomic index. As long as the inputs are often discrete (such as binary indicators of possession of a durable good such as a refrigerator or a car, or ordinal indicators such as the material used in household's dwelling — e.g., straw, wooden, or tin roof), there is a number of ways this discreteness can be taken into account. We advocate the use of polychoric and polyserial correlations as measures of the statistical association between the ordinal indicators (including binary ones), and show that the index based on the polychoric correlations performs better in ranking households as compared to other alternatives (such as using the ordinal variables per se in PCA, or constructing dummy variable indicators for each of the categories). We show the results of a Monte Carlo simulation, as well as applications to the real datasets, of the examples where the differences between methods may be very notable.

Comparing Statistical Software for Linear Mixed Models

◆ Brady T. West, University of Michigan; Brenda W. Gillespie, University of Michigan; Kathleen B. Welch, University of Michigan

University of Michigan, Center for Statistical Consultation and Research, 3550 Rackham Bldg., Ann Arbor, MI 48109

bwest@umich.edu

Key Words: *linear mixed models, statistical software, analysis of correlated data, statistical consulting*

In the past decade, advances in statistical software have resulted in several procedures capable of fitting linear mixed models with normally distributed errors. Currently, such procedures are available in SAS, SPSS, S+/R, STATA and HLM, with each program offering a unique set of available models, options, and defaults. We compare these packages in terms of their ability to fit a variety of models, their default settings, and their options. Some options of interest include types of tests for fixed effects (F-tests, Wald tests, likelihood-ratio tests), estimates of degrees of freedom for F-tests (e.g., traditional ANOVA degrees of freedom vs. adjusted degrees of freedom such as Satterthwaite, Kenward-Roger, or Huynh-Feldt), standard error estimates (model-based vs. robust or sandwich-type), and correlation structures available for repeated measures (e.g., compound symmetry, autoregressive, unstructured, toeplitz). Examples will be presented illustrating sample output for each package, with comparisons of output using the same data. We will make recommendations on appropriate software for various settings.

An Investigation into the Efficiency of Testing Procedures in Heteroscedastic Models

◆ Dennis Oberhelman, University of South Carolina; K. Rao Kadiyala, Purdue University

University of South Carolina, Moore School of Business, Columbia, SC 29208

fsu00018@moore.sc.edu

Key Words: *Behrens-Fisher, simulation, hypothesis test, linear models*

Testing the differences between population means when the population variances differ is referred to as the Behrens-Fisher problem. This paper considers the Behrens-Fisher problem in a multivariate setting. It reviews several recently proposed tests and evaluates their efficiency through Monte Carlo simulation.

Modeling Superpopulation Variance: Its Relationship to Total Survey Error

◆ James R. Knaub, Jr., Energy Information Administration

Energy Information Administration, Dept. of Energy, EI-53, Forrestal Bldg., 1000 Independence Ave., Washington, DC 20585

James.Knaub@eia.doe.gov

Key Words: *regression model, performance measure, scatterplot edits, inherent variance, nonsampling error, sampling error*

In repeated surveys there generally are auxiliary/regressor data available that are related to data collected in a current sample or census survey. These regressor data can be used to edit the current data through scatterplots, and to impute for missing data through regression. Another use for regressor data may be the study of total survey error. To do this, stratify data by regression model application. Then the related scatterplots can be used for editing, and predicted values can be found for data not collected, if any, and also to replace all data that are collected. If ratio regression is used, variance proportionate to the measure of size, then the sum of the predicted values equals the sum of the observed values they replace. The standard error of the total of the predicted values for every member of a finite population, divided by that total, and expressed as a percent, could be labeled as the estimated relative standard error of the superpopulation, or the RSESP. The RSESP would be influenced by (1) the models chosen, (2) inherent variance, and (3) total survey error (sampling and nonsampling error). This paper proposes the RSESP as a survey performance indicator.

Regression to Where?

◆ Amy B. Maddox, Baylor University; John Seaman, Jr., Baylor University

Baylor University, PO Box 97328, Waco, TX 76798

Amy_Maddox@baylor.edu

Key Words: *regression to the mean, regression effect, repeated measures model*

This research concerns the point of regression for regression effect models. There has been extensive debate in the literature about the point of regression. Galton (1886) showed that, under the assumption of normality, the regression effect was to the mean. Over 100 years later, Das and Mulder (1983) claimed, under the assumption of a normal error term, that the regression effect is to the mode. We shall show that, under the assumption of normal errors, the regression effect is to the turning point(s) of the distribution of the measurements, g. If g is symmetric unimodal then that turning point coincides with the mean and regression to the mean occurs. If g is not symmetric but unimodal, then the turning point coincides with the mode and regression to the mode occurs. If g is bimodal, then regression to the three turning points of g occurs. Under non-normal errors regression need not be to the turning point(s) of g.

Simulations: How Many Replications Do We Need?

◆ Jay R. Schaffer, University of Northern Colorado; Myoung-Jin Kim, University of Northern Colorado

University of Northern Colorado, McKee 530, Applied Statistics & Research Methods, Greeley, CO 80639

jay.schaffer@unco.edu

Key Words: *simulations, replications, EWMA control chart*

With the advent of high-speed computers, simulation studies can be easily created. Some of these simulation studies have used up to or even exceeded 10,000 replications. Generally, editors and reviewers will view such studies as having an adequate amount of data to answer the proposed questions. However, no current empirically based recommendations exist. Using an exponentially weighted moving average (EWMA) chart, this study investigates how much data is necessary to obtain an accurate solution to the question and determine the number of replications to use in a simulation. This paper is the first in a series of replication studies being investigated.

Inflated Generalized Poisson Regression Models

◆ Felix Famoye, Central Michigan University; Karan P. Singh, University of North Texas

Central Michigan University, Dept. of Mathematics, Mount Pleasant, MI 48859-0001

felix.famoye@cmich.edu

Key Words: *count data, k-Inflation, estimation, hypothesis-testing*

Zero-inflated Poisson and zero-inflated negative binomial regression models have been proposed for modeling data with too many zeros. A zero-inflated generalized Poisson regression model is proposed as a good alternate to model count data with too many zeros. We propose a k-inflated generalized Poisson regression (k-IGPR) to model count data with too many k-values. Estimation of the model parameters using the method of maximum likelihood is provided. A score test is presented to test whether the number of k-values is too large for the generalized Poisson model to adequately fit the data. The k-IGPR model is illustrated using a numerical dataset. We used the ordinary generalized Poisson regression (GPR), the k-inflated Poisson regression (k-IPR), and the k-IGPR to model the response variable. We found that the k-IGPR model is more appropriate than the k-IPR model for the data.

Unbiased Estimation of Maximum Flow-mediated Brachial Artery Diameter Using Nonparametric Regression Methods

◆ Michael E. Andrew, Centers for Disease Control and Prevention; Shengqiao Li, Centers for Disease Control and Prevention; Joan Dorn, University at Buffalo; Nedra Joseph, University at Buffalo; Violanti John, University at Buffalo; Burchfiel Cecil, Centers for Disease Control and Prevention; Ward Riley, Wake Forest University School of Medicine

Centers for Disease Control and Prevention, 1095 Willowdale Rd., NIOSH/HELD, Mail stop 4020, Morgantown, WV 26505

mandrew@cdc.gov

Key Words: *brachial artery diameter, nonparametric regession*

We consider methods for estimating the maximum value from a sequence of measurements of flow-mediated diameter of the brachial artery. Flow-mediated vasodilation (FMD) is represented using the maximum change from a baseline diameter measurement after release of a blood pressure cuff that has been inflated to reduce flow in the artery. Measurements are taken from ultrasound images measured continuously before and after cuff release. The influence of measurement error on the maximum diameter from raw data leads to overestimation of the maximum change from baseline. When using this approach it is necessary to make a judicious choice of regression methods and smoothing parameters to avoid overestimation or underestimation of FMD. Nonparametric regression models provide a means for overcoming this problem. This study will present results from simulation studies using kernel-based regression methods that provide models resulting in unbiased minimum variance estimates of FMD. Simulations were based on mean diameter series from existing data with additive measurement error. We also present an example using data from the Buffalo Police Health pilot study.

Using Markov Chain Monte Carlo to Model Survival

◆ Andrew Ostarello, California State University, Hayward

California State University, Hayward, 25013 Whitman St. #18S, Hayward, CA 94544

ostarello@yahoo.com

Key Words: *Markov chain Monte Carlo, survival analysis*

New developments in physical and engineering sciences have increased the need for effective simulation before costly development projects are undertaken. There are several techniques that are currently being employed to help researchers more fully understand the behavior and characteristics of fluids, protiens, and particles on a microscopic and molecular level. Some of the major techniques employed to enable these simulations are Markov chain Monte Carlo techniques, Molecular Dynamics Techniques, and most recently gemoetric cluster algorithms. Our research focuses on furthering comparisons between gemetric cluster algorithms and Markov chain Monte Carlo methods. In particular, we focus on the applicablility of these methods to modeling the survival time and reliability of these molecular structures over time. We employed R as our chief simulation tool, and used both discrete-time and continuous-time approaches to compare the two methods.

Comparison of Methods for Analysis of Non-normal Quality-of-life Data

◆ Jennifer L. Beaumont, Evanston Northwestern Healthcare; Lisa Lix, University of Manitoba; Kathleen Yost, Evanston Northwestern Healthcare; Elizabeth A Hahn, Evanston Northwestern Healthcare

Evanston Northwestern Healthcare, Center on Outcomes, Research, and Education, 1001 University Place, Suite 100, Evanston, IL 60201

jbeaumont@enh.org

Key Words: *quality of life, robust estimation, transformation, non-normal distribution*

Assessment of quality of life (QOL) is receiving increasing attention, particularly in areas of chronic illness research. Researchers often use conventional methods based on least-squares estimators (i.e., t-test, ANOVA F-test) to test mean equality across patient groups. However, least-squares estimators are sensitive to the presence of skewed distributions and extreme values, which characterize QOL measures. We compare conventional methods to alternate approaches that are insensitive to non-normal distributions, including transformations and robust estimators (i.e., trimmed means and Winsorized variances). These techniques are applied to data from a comprehensive assessment of the health-related quality of life of blood and marrow transplantation (BMT) survivors.

How Many Replications are Needed in a Simulation Study?

◆ Daniel Mundfrom, University of Northern Colorado; Ampai Ussawarujikulchai, University of Northern Colorado; Dale Shaw, University of Northern Colorado

University of Northern Colorado, Dept. of Applied Statistics and Research Methods, Campus Box 124, Greeley, CO 80639

daniel.mundfrom@unco.edu

Key Words: *simulation, number of replications, multiple comparisons*

Simulation studies are widely used to investigate properties of various statistical procedures that are not easily tracked theoretically. Properties of multiple comparison procedures (MCP), typically experiment-wise error rates and power, have been studied frequently with simulations. In reviewing recent simulation studies of MCPs, no standard seems to exist in regard to the number of replications required to adequately assess the properties being studied. With the continued rise of of computing capabilities, thousands of replications can be quickly performed in such studies. This study compared the results of several studies reconducted with various numbers of replications in an attempt to determine if an optimum number of replications for MCP simulation studies could be found. In general, the number of replications required to achieve the same or very similar results is less than what has often been used.

Minimum L2 Estimators for Poisson Mixtures with Two Components

◆ Shuyi Shen, Southern Methodist University; Ian R. Harris, Southern Methodist University

Southern Methodist University, 3225 Daniels Rd., Rm. 144 Heroy Bldg, Dallas, TX 75275

sshen@smu.edu

Key Words: *divergence, influence function, L2 distance, mixing proportions, robustness*

Finite mixture model is extremely useful in fitting heterogeneous data. Often in practice, the primary interest in fitting a mixture model is to estimate the mixing proportions, while the component densities are already known or can be estimated separately from classified data. Current estimating methods used in the mixture problem include maximum likelihood, method of moments, minimum distance and Bayesian approaches. These methods typically give estimators for the mixing proportions that are not in explicit form and numerical methods are needed to calculate the estimate. A new minimum divergence estimator is developed and applied to estimation of Poisson mixtures with two components. The proposed estimator minimizes the integrated squared difference between densities or distributions (L2 distance). Unlike the other estimators, the proposed estimator for the mixing proportion is in closed form. The method also offers robust estimators while retaining acceptable efficiency compared to the MLE and the Minimum Hellinger Distance Estimator (MHDE). The application of the proposed method to epidemiological data will be presented.

New Tables of the Largest Root of a Matrix in Mulivariate Analysis

◆ William W.S. Chen, Internal Revenue Service

Internal Revenue Service, Statistics of Income, PO Box 2608, Washington, DC 20013

william.w.chen@irs.gov

Key Words: *characteristic roots, extended tables, Fisher-Girshick-Shu-Roy, distribution, percentage points*

The Fisher-Girshick-Shu-Roy distribution (1939), which has interested statisticians more than six decades, is revisited. Instead of using Pillai's K.C.S. method by neglecting higher-order terms of the cumulative distribution function (c.d.f.) of the largest root to approximate the percentage points, we simply keep the whole c.d.f. and apply its natural nondecreasing property to calculate the exact probabilities. We forcus on distribution when there are seven or eight characteristic roots. At the duplicated percentage points we found our computed percentage points consistent to the existing tables. Our tabulations, however, have greatly extended those tables.

Comparison of Imputation Methods in Behavioral Science Research: A Small Longitudinal Study of Sexual Counseling after Prostate Cancer

◆ Dawen Sui, University of Texas M.D. Anderson Cancer Center; Leslie R. Schover, University of Texas M.D. Anderson Cancer Center; Juwon Song, University of Texas M.D. Anderson Cancer Center; Andrea L. Canada, University of Texas M.D. Anderson Cancer Center; Leah E. Neese, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd.-243, Houston, TX 77030

dawensui@mdanderson.org

Key Words: *longitudinal study, missing data, imputation, general linear mixed model, behavioral science study*

The purpose of this paper is to compare imputation methods in handing missing data in a small longitudinal dataset from a pilot, randomized trial of two formats of sexual counseling for couples in which the man had been treated for localized prostate cancer. Fifty-one couples completed the counseling. Both men and their partners filled out self-report questionnaires on entry to the study, at the last session of the intervention, and three-month and six-month follow-up. The responses of interest in this analysis include several individual multiple-choice items. Traditionally, mean imputation is used to estimate the value of scales composed of multiple items, only some of which have missing values. In this study, single and multiple imputation strategies are selected to impute individual missing items, and then imputed values are used to calculate the subscales and response variables. Distributions of the responses are considered to select an appropriate imputation technique. The characteristics of the model fits, parameter estimates, standard deviations, and p values from a general linear mixed model are compared using the complete dataset, traditional mean imputation.

Exact Two-sample Inference with Missing Data

◆ Ken Cheung, Columbia University

Columbia University, Dept. of Biostatistics, 722 West 168th St., Room 630, New York, NY 10032

cheung@biostat.columbia.edu

Key Words: *exact inference, Wilcoxon test*

In comparing follow-up measurements from two independent populations, missing records may arise due to relocation or censoring by terminal events. I will present exact inference for a class of modified U-statistics under missing at random dropouts. The methods are nonparametric in that no distributional assumption is necessary for the outcome variables and the missingness pattern. The methodology was motivated by numerous applications to two-sample comparison in small studies that consist of missing data on a primary endpoint, but have complete records on a baseline variable. Two such datasets would be used as illustrative examples. I will also discuss the computational details for testing and estimation.

On the Parametric Bootstrap of an Asymptotic Variance

◆ Fassil Nebebe, Concordia University

Concordia University, Dept. of Decision Sciences & MIS, 1455 de Maisonneuve Blvd. W, Montreal, PQ H3G 1M8 Canada

fnebebe@vax2.concordia.ca

It is not uncommon that a sample statistic of observations from a parametric distribution has a normal limiting distribution but its variance does not exist for any finite sample size. Under this circumstance, the usual ideal bootstrap variance calculation does not lead to a converging or stable value (as the number of bootstrap sample increases). This paper addresses the issue of estimating numerically the variance of a limiting distribution when the variance does not exist for any finite sample sizes. This "asymptotic" variance is useful for interval estimation and also sample size determination. Contrary to the usual delta method which requires a case by case theoretical treatment, the present approach needs only the resampling of the parametric distribution under study.

Evaluating Sufficient Similarity for Disinfection By-product Mixtures

◆ Zhenxu J. Ma, Battelle; Paul I. Feder, Battelle; Dick J. Bull, MoBull Consulting; K. Schenck, U.S. Environmental Protection Agency; J. E. Simmons, U.S. Environmental Protection Agency; Linda K. Teuschler, U.S. Environmental Protection Agency; Glenn Rice, U.S. Environmental Protection Agency

Battelle, 505 King Ave., Columbus, OH 43201

maj@battelle.org

Key Words: *Disinfection By-Product (DBP) mixtures, Monte Carlo method, bootstrap hypothesis test procedure, sufficient similarity*

Chemical mixtures risk-assessment methods propose sufficient similarity as a key concept for evaluating complex environmental mixtures. If two mixtures are judged to be sufficiently similar, then toxicity data for one mixture can be used as a surrogate when conducting a quantitative risk assessment for the other. An important class of environmental complex mixtures arises as disinfection by-products (DBP). This presentation proposes two alternative statistical methods to assess the similarity of DBP mixtures. In the first method, a multivariate index was developed combining the water chemical composition and the health effects information of each mixture (e.g., carcinogenic, reproductive, neurological effects). A Monte Carlo method was developed to propagate the uncertainty in the health effects specific index governed by assumed distribution, to be used in statistical comparisons among water characteristics such as water sources, treatment plants, and distribution systems. In the second method, a bootstrap hypothesis test procedure under a MANOVA model was developed to test the effects of these factors on the similarity of the DBP mixtures.

Multiple Comparisons Using Medians

◆ Scott J. Richter, University of North Carolina, Greensboro; Melinda H. McCann, Oklahoma State University

University of North Carolina, Greensboro, Mathematical Sciences, Greensboro, NC 27402

sjricht2@uncg.edu

Key Words: *simultaneous inference, pairwise comparisons, median difference, permutation test*

We investigate properties of simultaneous pairwise comparison procedures based on permutation tests and median statistics. Nemyani (1963) and Levy (1979) outlined an exact method, based on Mood's median test, for simultaneous pairwise comparisons of medians, using as the reference distribution the permutation distribution of the maximum of the pairwise Mood statistics. Hochberg and Tamhane (1987) suggest that this method is not valid since it is based on the permutation distribution derived by permuting freely across all treatments. They suggest a valid test must be based on permuting only within pairs of treatments being compared. We propose a similar procedure based on differences between medians instead of Mood statistics. We investigate the power and Type I error rate properties of the tests under both methods of permuting.

Impact of the Sampling Rate on the Estimation of the Parameters of Fractional Brownian Motion

◆ Zhengyuan Zhu, University of North Carolina, Chapel Hill; Murad S. Taqqu, Boston University

University of North Carolina, Chapel Hill, Dept. of Statistics and Operations Research, CB #3260, New West, Chapel Hill, NC 27599

z-zhu@unc.edu

Key Words: *fractional Gaussian noise, self-similarity, long-range dependence, maximum likelihood estimation*

Fractional Brownian motion is a mean zero self-similar Gaussian process with stationary increments. Its covariance depends on two parameters, the self-similar parameter H and the variance C. Suppose that one wants to estimate optimally these parameters by using n equally spaced observations. How should these observations be distributed? We show that the spacing of the observations does not affect the estimation of H (this is due to the self-similarity of the process), but the spacing does affect the estimation of the variance C. For example, if the observations are equally spaced on [0,n] (unit-spacing), the rate of convergence of the maximum likelihood estimator (MLE) of the variance C is √{n}. However, if the observations are equally spaced on [0,1] (1/n-spacing), or on [0, n^2] (n-spacing), the rate is slower, √{n}/\ln n. We also determine the optimal choice of the spacing ∂ when it is constant, independent of the sample size n. While the rate of convergence of the MLE of C is √{n} in this case, irrespective of thevalue of ∂, the value of the optimal spacing depends on H. It is 1 (unit-spacing) if H=1/2.

Model Selection Based on Maximum Likelihood Estimation: Jackknife Approach

◆ Hyunsook Lee, Pennsylvania State University; G. Jogesh Babu, Pennsylvania State University

Pennsylvania State University, 325 Thomas Bldg., University Park, PA 16802

hlee@stat.psu.edu

Key Words: *jackknife, model selection, maximum likelihood estimation, Kullback-Leibler distance, unbiased estimator*

Various model selection methods have been developed to choose a model close to a true model in regression problems, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Minimum description length (MDL), Mallow C_p, cross-validation, bootstrap information criterion (EIC), corrected Akaike Information Certerion (AICc), unified Akaike Information Criterion (AICu), and minimizing R^2. We present an information criterion for model selection by using jackknife method to minimize Kullback-Leibler distance. The jackknife method reduces bias on Kullback-Leiber distance measure between an unknown true model and an estimated model. We discuss the possibility of using the jackknife information criterion (JIC) as a model selection criterion. The asymtotic property of JIC is studied and Monte Carlo simulations are performed to compare JIC to other information criteria.

234 Economic Outlook Speaker Luncheon (Fee Event)

Business and Economics Statistics Section

Tuesday, August 10, 12:30 pm–1:50 pm

Keeping Inflation Under Control

◆ Stephen G. Cecchetti, Brandeis University

Brandeis University, International Business School, Mail Stop 021, PO Box 9110, Waltham, MA 02454

cecchetti@brandeis.edu

Sometime in the last five years monetary policymakers changed gear. They switched from a posture of pushing inflation down, to one of keeping it from going too low. Last year, it got to the point where many prices were falling, and analysts became concerned about the prospects for deflation. Do monetary policymakers need to approach their job differently when inflation is low? What are the prospects for inflation, and what can we expect from the Federal Reserve over the next few months, quarters, and years?

235 Biopharmaceutical Section Roundtable Luncheons (Fee Event)

Biopharmaceutical Section

Tuesday, August 10, 12:30 pm–1:50 pm

Issues in Doing Clinical Trials with Small Populations

◆ Jessica Kim, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852

Kimj@cber.fda.gov

In many clinical trials, the study population can be relatively small. This may bring issues and concerns in analyzing and interpreting the data. It is particularly true when we are dealing with drug development for a rare/serious disease. Topics we will discuss include the importance of defining and measuring endpoints, maintaining a high compliance rate, and developing a pre-specified outcome for sample size determination. We may also discuss the importance of randomization and use of historical controls in studies with limited data availability.

Issues in Choice and Analyses of Endpoints in Antiviral Drug Trials

◆ Victor G. De Gruttola, Harvard School of Public Health

Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02131

victor@sdac.harvard.edu

Key Words: *antiretroviral drug trial, study endpoints, resistance mutations*

Evaluation of efficacy of antiretroviral drugs is generally based on virological endpoints, such as viral load and development of resistance mutations. One issue that arises in defining such endpoints is how time series of viral load measurements can be summarized in a single metric. Common metrics include time to rebound and AUC. For analyses of the latter, nonparametric approaches may be preferable to standard analyses. To reduce viral genetic sequences to single metrics, resistance endpoints may be defined that are based on the drug options available to patients experiencing virological rebound. Such measures are necessary to evaluate the consequences of failing a therapy for patients' long-term prognoses.

Missing Data in Drug Development

◆ Thomas Kelleher, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, Dept. 703, 5 Research Pkwy., Wallingford, CT 06492

thomas.kelleher@bms.com

Key Words: *missing data, drug development, regulatory guidance, HIV clinical trials*

Missing data naturally arise during drug development due to missed study visits and premature treatment discontinuations. Methods for addressing different types of missing data will be discussed. These methods will include those recommended by regulatory authorities as well as alternative methods. Particular examples from HIV clinical trials will be presented.

Graphics: An Underutilized and Misused Tool for Effective Communication

◆ Thomas Bradstreet, Merck & Co., Inc.

Merck & Co., Inc., Mail Stop BL 3-2, 10 Sentry Pkwy., Blue Bell, PA 19422

thomas_bradstreet@merck.com

Key Words: *chartjunk, data analysis, graphical construction, graphical design, perception, presentation graphics*

Communicating effectively through graphics is a prerequisite skill for pharmaceutical professionals whether they are statisticians or nonstatisticians. Many pharmaceutical professionals are responsible for summarizing and communicating the results of experiments through either an oral presentation, a written report, or a refereed publication. To empower pharmaceutical professionals to produce effective graphics, we must train them in implementing the basic definitions, strategies, and practical considerations of graphical perception, design, and construction. Further, there needs to be a harmonization of the ideas behind graphics for presentation and graphics for data analysis. Graphical insight follows a continuum from discovery and understanding in data analysis through to clarity and efficiency in final presentation. Proportionally incorrect and chartjunk-laden default presentations of data analysis and presentation graphics which are generated by many software packages can be improved to be clear and effective graphics.

Stratified Clinical Trials—Some Bread and Butter Issues

◆ Devan V. Mehrotra, Merck & Co., Inc.

Merck & Co., Inc., UN-A102, 785 Jolly Rd., Bldg. C, Blue Bell, PA 19422

devan_mehrotra@merck.com

Key Words: *binary data, interaction, Mantel-Haenszel test, minimum risk weights, random stratum sizes*

We will discuss fundamental statistical issues involving comparative clinical trials with stratification. First, we will differentiate between prognostic and nonprognostic factor stratification. Second, we will stress the importance of clearly stating the null hypothesis being tested. Third, we will review some recently developed alternatives to the popular Mantel-Haenszel procedure, and discuss the related issue of stratum-weighting. Fourth, we will describe how to (and how not to!) test for a treatment by stratum (TxS) interaction, and what to do in the presence of an interaction. Finally, we will comment on the issue of random stratum sizes and discuss the related impact on interval estimates of the “true overall treatment effect.” Illustrative examples and simulation results will be used to reinforce the key points.

236 Section on Bayesian Statistical Science Roundtable Luncheon (Fee Event)

Section on Bayesian Statistical Science

Tuesday, August 10, 12:30 pm–1:50 pm

Bayesian Statistics at FDA, Especially for Clinical Trials of Medical Devices

◆ Gene A. Pennello, U.S. Food and Drug Administration

U.S. Food and Drug Administration, Division of Biostatistics, HFZ-542, Rockville, MD 20850

gxp@cdrh.fda.gov

Key Words: *therapeutic devices and implants, diagnostic test, microarray, hierarchical model, interim analysis, post-market adverse event data-mining*

Bayesian design and analysis is part of an increasing number of pre-market submissions to the Center for Devices and Radiological Health (CDRH) of the Food and Drug Administration (FDA). This initiative, which began approximately five years ago, takes advantage of good prior information on safety and effectiveness that is often available from studies on the same or recent generation devices. To date, about fourteen PreMarket Approval (PMA) applications have been approved by CDRH based on Bayesian analyses, including coronary stents (both drug-eluting and not), spinal fixation devices, a diagnostic breast imager, an artificial urinary sphincter for urinary incontinence, and a female permanent contraception implant. Four were presented to an FDA Advisory Committee Panel of outside experts for a recommendation of approval or disapproval by FDA. Bayesian methodology is also being used at the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) to datamine post-market adverse event reports for associations of events with drugs and vaccines, and CDRH is looking to do the same. Bayesian methodology may also be helpful.

237 Section on Government Statistics Roundtable Luncheons (Fee Event)

Section on Government Statistics

Tuesday, August 10, 12:30 pm–1:50 pm

International Conference on Telephone Survey Methodology

◆ Clyde Tucker, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Rm. 1950, Washington, DC 20212

tucker_c@bls.gov

Key Words: *cellular phones, telephone sample design, nonresponse bias, survey estimation, survey operations*

It has been more than 15 years since the international conference devoted to telephone survey methodology. In the intervening years, there have been rapid changes in telephone systems and telephone survey methodology, including the use of list-assisted designs. The impact of these changes vary internationally. Changes in the telephone system include a revolution in cellular service, new numbering systems allowing greater portability of telephone numbers, and features that protect privacy and access. Telephone subscribers have also changed how they use the telephone. The goals of the conference will be to bring together researchers and survey practitioners from around the world concerned with telephone survey methodology and practice in order to stimulate research papers that contribute to the science of reducing errors attributable to poor telephone survey design, to provide documentation of current practices, and to stimulate new ideas for further research and development. Clyde Tucker is one of the primary organizers of this international conference, which is currently planned for January 2006.

COPAFS and the Federal Statistical System

◆ Edward Spar, COPAFS

COPAFS, 1429 Duke St., Alexandria, VA 22314

COPAFS2aol.com

Key Words: *COPAFS, federal statistics, data-user needs*

The Council of Professional Associations on Federal Statistics (COPAFS) has for 22 years been an advocate for users of federal statistics. COPAFS works closely with statistical agencies, and user communities and Congress to foster the production and use of high quality data. Here's your chance to learn how COPAFS works with federal statistical agencies and stakeholders, and learn about and discuss breaking issues of importance to users of federal statistics.

238 Section on Health Policy Statistics Roundtable Luncheons (Fee Event)

Section on Health Policy Statistics

Tuesday, August 10, 12:30 pm–1:50 pm

Improving the Accuracy of Hospital Report Cards

◆ Peter C. Austin, Institute for Clinical Evaluative Sciences

Institute for Clinical Evaluative Sciences, G1 06, 2075 Bayview Ave., Toronto, ON M4N 3M5 Canada

peter.austin@ices.on.ca

Key Words: *hospital report cards, risk adjustment, provider profiling, statistical modeling, quality of care*

There is an increasing interest in measuring the quality of care that patients receive while in hospital. During this round table luncheon, we will discuss issues concerning risk-adjustment, statistical models for provider profiling, and policy implications of provider profiling.

Pre- and Post-award Role of Statisticians in NIH Grants

◆ Anamaria S. Kazanis, University of Michigan

University of Michigan, G.R. Ford School of Public Policy, 712 Oakland, Room 159, Ann Arbor, MI 48104-3021

akazanis@umich.edu

Key Words: *research, grant, National Institutes of Health (NIH)*

Statisticians play crucial pre- and post-award roles in the evaluation, development, and implementation of research from the National Institutes of Health. Most NIH research is funded through the extramural program, sponsoring scientists working at universities, or private research laboratories. The pre-award roles for statisticians include reviewer, developer, writer, and adviser. Statisticians with medical and behavioral research experience are needed to evaluate proposals, as members of the NIH review study sections, as well as on scientific teams writing and developing the data analysis sections of proposals. Applications for NIH funding must describe the research design, with power analysis to justify the sample size, and the procedures to be used, including data collection, analysis, and interpretation. Once the grant is funded, the statistician should implement the analytical procedures described in the research design, manage data collection, cleaning, and entering. After analysis, the statistician should be coauthor of articles and presentations. From consultant to co-investigator, the role of the statistician will vary according to the level of complexity of the study.

239 Section on Physical and Engineering Sciences Roundtable Luncheon (Fee Event)

Section on Physical and Engineering Sciences

Tuesday, August 10, 12:30 pm–1:50 pm

What's Happening in Computer Experiments?

◆ William Notz, The Ohio State University

The Ohio State University, Dept. of Statistics 1958 Neil Ave., Columbus, OH 43210-1247

win@stat.ohio-state.edu

Key Words: *computer experiments*

This roundtable will provide an overview of computer experiments followed by a discussion of some current issues in the area. We will begin by explaining what a computer experiment is, give some examples of computer experiments, describe some of the statistical issues that one faces, and conclude with a discussion of topics of current interest. This roundtable is intended for people who are interested in the area but know little about it as well as researchers who would like to share some ideas about topics that they believe are of current interest. The goal is to encourage new people to enter this area of research as well to allow active researchers to share some ideas.

240 Section on Quality and Productivity Roundtable Luncheon (Fee Event)

Section on Quality and Productivity

Tuesday, August 10, 12:30 pm–1:50 pm

Communicating Statistics to Nonstatisticians in Industry

◆ Philip R. Scinto, Lubrizol Corporation

Lubrizol Corporation, 29400 Lakeland Blvd., Drop # 152A, Wickliffe, OH 44092-2298

prs@lubrizol.com

Key Words: *statistics, communication, customers*

As statistical consultants in industry, most of us find our work in solving problems and making decisions based upon statistical analysis, logic, and thinking to be both challenging and rewarding. However, we must not overlook the importance of the equally challenging and rewarding necessity of communicating our analysis, logic and decisions to our chemist, physicist and engineer customers, collaborators, and colleagues. Even a high-level understanding of simple statistical concepts such as estimation, bias, variability, power, and probability by our customers would increase productivity by increasing the chances that analyses and decisions are applied correctly with or without the presence of a statistician. We will discuss and share our experiences and our techniques in communicating statistics to nonstatisticians.

241 Section on Statistical Graphics Roundtable Luncheon (Fee Event)

Section on Statistical Graphics

Tuesday, August 10, 12:30 pm–1:50 pm

Tables to Charts and Tables of Charts

◆ Graham Wills, SPSS Inc.

SPSS Inc., 233 South Wacker, 11th Floor, Chicago, IL 60606

gwills@spss.com

Key Words: *tables, charts, multivariate, visualization, trellis, paneling*

A common need for professional statisticians is to take a large, probably multidimensional table and represent it using a graphic of some form. Often this chart's intended viewers are not highly statistically literate and charts appropriate to their level will be required. We will discuss and show examples of automatic techniques for converting tables to charts, including discussion of web generation, automatic variable choice, multidimensional paneling (inclusive of simple paneling, shingling, trellis, and dealing with nested dimensionality), choice of base chart and issues with interactivity. The discussion will make use of the theory and practice of Leland Wilkinson's *The Grammar of Graphics*, but it will not be necessary to have read the book to attend!

242 Section on Statistics and the Environment Roundtable Luncheon (Fee Event)

Section on Statistics and the Environment

Tuesday, August 10, 12:30 pm–1:50 pm

Getting Spatial Statistics into GIS and Other Software

◆Jay M. Ver Hoef, Alaska Dept. of Fish and Game

Alaska Dept. of Fish and Game, 1300 College Rd., Fairbanks, AK 99701

ffjmv@uaf.edu

Key Words: *spatial statistics, GIS, software*

This roundtable will provide a forum for discussion of the some of the most recent developments in spatial statistics, software, and GIS. Discussion will begin with the availability of spatial statistics in various software. I would then like to share some of our experiences with the software. Topics will include the feature sets of various software. For example, can we easily do generalized linear models that include spatial models? How important is it to have spatial statistical analysis tightly linked to a GIS? The roundtable will also provide an opportunity to brainstorm about some of the new directions that are wanted and needed in spatial statistical software.

243 Section on Statistics in Epidemiology Roundtable Luncheon (Fee Event)

Section on Statistics in Epidemiology

Tuesday, August 10, 12:30 pm–1:50 pm

Causal Inferences in Randomized Trials with Noncompliance

◆Xiao-Hua A. Zhou, University of Washington

University of Washington, Dept. of Biostatistics, Box 357232, 1705 NE Pacific St., Seattle, WA 98195

azhou@u.washington.edu

Key Words: *causal inferences, randomized trials, noncompliance*

In a randomized clinical trial on the effectiveness of treatments, one common problem is noncompliance. Noncompliance occurs if subjects do not take their assigned treatments. I will discuss the current methodology for dealing with noncompliance in estimation of the treatment effects.

244 Section on Survey Research Methods Roundtable Luncheons (Fee Event)

Section on Survey Research Methods

Tuesday, August 10, 12:30 pm–1:50 pm

Employment—Which Number to Pick?

◆ George S. Werking, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 4840, Washington, DC 20212

werking.george@bls.gov

Key Words: *employment change, economic cycle*

Monthly employment change is always a highly scrutinized number by policymakers, the financial markets, and forecasters, but these numbers take on a significant added interest during turning points in the economy. Each month, the U.S. Bureau of Labor Statistics publishes two measures of monthly employment change: one measure from the Current Population Survey (CPS), which is a 60,000-unit household survey, and the other from the Current Employment Statistics (CES) survey, which is a 400,000-unit establishment survey. While monthly figures can differ in size and direction, historically the longer-term trends track well. However, the most recent recovery period (November 2001 forward) has shown significantly different trends for the two surveys and has lead to numerous articles being written on the apparent discrepancy. This discussion will focus on the scope and methodology differences between CPS and CES, and what we know about the discrepancy in employment growth.

Enhancing Collaboration between Social Science Researchers and Statisticians

◆ David A. Binder, Statistics Canada

Statistics Canada, Methodology Branch, RH Coats Bldg., Ottawa, ON K1A 0T6 Canada

binddav@statcan.ca

Key Words: *analysis, survey data, survey design, econometrics, health policy research, longitudinal studies*

What are the barriers to involving statisticians in the social sciences, in economics, and in other research disciplines? Are there important “cultural” differences that are a hindrance to cross-disciplinary communications? Do statisticians use language that is meaningless or that has a different meaning to other researchers? Are the norms used by peer reviewers from various disciplines well understood by statisticians when they are consulted? One area where the issue of collaboration with statisticians often arises is in the analysis of survey data. Such data are now being used increasingly by many disciplines, but the subtleties associated with appropriate ways to analyze such data may not be well understood or easy to explain. Most accepted procedures used in the social sciences (often available in commercial software) are based on assuming that the data model incorporates all the relevant information about the mechanisms for selecting respondents. Do social science researchers understand this sufficiently well? What is the role of statisticians here, if any? These are some questions that we will be exploring.

245 Social Statistics Section Roundtable Luncheons (Fee Event)

Social Statistics Section

Tuesday, August 10, 12:30 pm–1:50 pm

Meeting the Response Rate Challenge

◆ Roberta L. Sangster, Bureau of Labor Statistics; ◆ Brenda G. Cox, Battelle

Battelle, 2101 Wilson Blvd., Suite 800, Arlington, VA 22201

coxb@battelle.org

Key Words: *response rates, telephone survey, face-to-face survey, interview survey, data collection, data quality*

Survey researchers are routinely challenged to provide high quality survey data with high response rates, but many indicate that they are expending much more effort now and with less result. This roundtable focuses on sharing information on new tools and old standbys that assist in meeting the response challenges posed by telephone and face-to-face surveys but without sacrificing data quality.

Recent Developments in Poverty Measurement

◆ Daniel Weinberg, U.S. Census Bureau

U.S. Census Bureau, HHES Division, Washington, DC 20233-8500

daniel.h.weinberg@census.gov

Key Words: *poverty*

A 1995 National Academies of Science panel report recommended substantial changes in the way the United States measures poverty. The Census Bureau (and others) have carried out substantial research since then in support of a process that may lead to changes in the the way the Office of Management and Budget defines poverty. This roundtable luncheon will update participants on the current status of that research.

246 Section on Teaching of Statistics in the Health Sciences Roundtable Luncheon (Fee Event)

Section on Teaching Statistics in the Health Sciences

Tuesday, August 10, 12:30 pm–1:50 pm

Assessment of Medical Students and Post-graduate Physicians in Biostatistics Courses

◆ Reena Deutsch, University of California, San Diego

University of California, San Diego, 200 West Arbor Dr., San Diego, CA 92103-8203

rdeutsch@ucsd.edu

Key Words: *assessment, biostatistics training, teaching, evaluation*

There are a plethora of methods available to evaluate academic progress of students: i.e., homework, quizzes, exams, reports, projects, article review, and others. How can these be tailored to medical students, residents, fellows, and other physicians-in-training? What works? What doesn't work? How do we know if a particular method of evaluation "works?" What exactly should be measured, and how do we measure it? How many forms of assessment should be made? How do we balance the various formats and frequency of assessments with our workload and other responsibilities? This should be a lively discussion where we will share our experiences with others who teach in the health sciences. We may not have all the answers, but hopefully we'll gain some insight and acquire some tips about assessment.

247 Section on Statistics in Defense and National Security Roundtable Luncheons (Fee Event)

Section on Statistics in Defense and National Security

Tuesday, August 10, 12:30 pm–1:50 pm

Game Theory and Risk Analysis in Counterterrorism

◆ David Banks, Duke University

Duke University, ISDS Box 90251, Durham, NC 27708

banks@stat.duke.edu

Key Words: *game theory, risk analysis, counterterrorism, portfolio theory*

The federal government is making massive investments in security infrastructure. These investments imply a cost-benefit analysis of risk that is coupled with some kind of game-theoretic examination of the possible actions that may be taken by terrorists. This luncheon will discuss various issues in making both statistical risk analysis and strategic game theory more realistic in the context of the kinds of attacks for which the country is preparing. Some of the topics will include extensions of risk analysis to address adversarial attack, applications of portfolio theory, and the deficiencies of classical game theory.

248 Section on Statistical Education Roundtable Luncheons (Fee Event)

Section on Statistical Education

Tuesday, August 10, 12:30 pm–1:50 pm

Teaching an Undergraduate Course in Statistical Consulting: What, When, Why, and How?

◆ Heather S. Smith, California Polytechnic State University; ◆ John H. Walker, California Polytechnic State University

California Polytechnic State University, Dept. of Statistics, 1 Grand Ave., San Luis Obispo, CA 93407

jwalker@calpoly.edu

Key Words: *teaching, undergraduates, statistical consulting*

There is wide agreement that to be an effective practicing statistician demonstration of good consulting skills is important. How do academic statistics programs help to develop these skills in their students? What are the current best methods for teaching students about these skills? How can we best create an environment that will allow students the latitude to develop and practice these skills? Exploring answers to these questions will be the focus of this roundtable. The hosts have just developed and team-taught a new statistical consulting course at Cal Poly and hope to share their experiences and learn from the experiences of others. If you are currently teaching a statistical consulting course, if you are thinking of developing such a course, or if you are a practicing statistical consultant interested in helping to improve statistics education, please join us.

Using Simulations to Introduce Inference

◆ Murray H. Siegel, Sam Houston State University

Sam Houston State University, Math. & Stats., Box 2206, Huntsville, TX 77341

mth_mhs@shsu.edu

Key Words: *simulation, probability, bias, expectation, chi-square*

Students know the meaning of p value, but do they understand what it means? Simulations are used to help students develop an understanding. The first problem involves the hiring of 30 technicians; 40% of the qualified applicants are female. We expect 12 females to be hired but only eight are. Is there bias? Trials of a binomial simulation (30, .4) are run. The only reason for divergence from the expected is variation. What is the probability, based on the simulations, of hiring eight or fewer females? What does this say about the actual case? The second simulation involves a selection of a favorite cable network. The class votes and the results are recorded. Do the results confirm or refute that the four networks are equally favored? The chi-square statistic is introduced. A simulation is designed with four choices each with a probability of .25. The n is the same as the size of the class. Many simulations are run and for each the results are used to compute a chi-square statistic. What is the probability of obtaining a chi-square statistic greater than or equal to the one obtained for our vote? What does this say about the chance that the networks are equally favored?

How to Best Use a Web Site for Your Course

◆ Marjorie E. Bond, Monmouth College

Monmouth College, 218 East Girard, Monmouth, IL 61462

mebond@monm.edu

Key Words: *education, web site*

How do you use your web site for your course, and how do we get students to use the web site? Is the web site simply a location for documents? For useful links? For solutions? Should students be able to do well in the course without visiting the web site? How can we integrate the web site into our courses without creating extra work for ourselves?

249 ASA Stat Bowl Session II

ASA, ENAR, IMS, SSC, WNAR

Tuesday, August 10, 2:00 pm–3:50 pm

250 Late-breaking Session #1 - Statistical Issues in Forensic Bullet Lead Evidence

ASA, ENAR, WNAR, IMS, SSC

Tuesday, August 10, 2:00 pm–3:50 pm

Forensic Analysis: Statistical Comparison of Bullet Lead Compositions

◆ Karen Kafadar, University of Colorado, Denver; Cliff Spiegelman, Texas A&M University/The Texas Transportation Institute

University of Colorado, Denver, PO Box 173364, Dept. of Mathematics Box 170, Denver, CO 80217-3364

kk@math.cudenver.edu

Since the 1960s, FBI has performed Compositional Analysis of Bullet Lead (CABL), a forensic technique that compares the elemental composition of bullets found at a crime scene to that of bullets found in a suspect's possession. CABL has been used when no gun is recovered, or when bullets are too small or fragmented to compare striations on the casings with those on the gun barrel. The National Academy of Sciences formed a Committee charged with the assessment of CABL's scientific validity. The report, "Forensic Analysis: Weighing Bullet Lead Evidence," included discussion on the effects of the manufacturing process on the validity of the comparisons, the precision and accuracy of the chemical measurement technique, and the statistical methodology used to compare two bullets and test for a "match." This talk will focus on the statistical analysis: the FBI's methods of testing for a "match," the apparent false positive and false negative rates, the FBI's clustering algorithm ("chaining"), and the Committee's recommendations. Finally, additional analyses on data made available for future studies will be discussed.

Expert Testimony on Identity

◆ David H. Kaye, Arizona State University

Arizona State University, College of Law, Tempe, AZ 85287-7906

David.Kaye@asu.edu

Criminalists and other forensic experts often testify about matching items of evidence. DNA, handwriting, tool marks, bite marks, and fingerprints are common examples. A recent report of a committee of the National Academy of Sciences offers important recommendations concerning testimony about identifying bullets with a procedure called "Compositional Analysis of Bullet Lead." This talk will place these recommendations in the broader context of identification evidence generally. It will describe the legal standards that apply to such evidence and offer suggestions on how forensic experts' judgments, statistical, or scientific information should be presented to a lay judge or jury.

Post-NAS Realities of Comparative Bullet Lead Analysis

◆ William Torbin, Consultant

Consultant, 6009 Noranda Dr., Mineral, VA 23117-9647

wtobin@nexet.net

Because the recent report by the National Research Council of the National Academies of Science (NAS) on the forensic practice of comparing bullet lead compositions is circumspect, painstakingly diplomatic, and subtle, the FBI Laboratory Group leader for the practice claims that the study vindicated the FBI position on the practice and continues to testify in court that FBI analysts can link a bullet composition to a specific "molten source." This talk will update attendees on the post-NAS state of affairs regarding the forensic practice of comparative bullet lead analysis (CBLA), the realities of bullet distribution, and why statistical estimation of a "false-positive probability" based on bullet compositions in the FBI bullet lead database may not be of probative significance.

251 *Technometrics* Invited Paper Session △

Technometrics, Section on Quality and Productivity

Tuesday, August 10, 2:00 pm–3:50 pm

Failure Amplification Method: An Information Maximization Approach to Categorical Response Optimization

◆ V. Roshan Joseph, Georgia Institute of Technology; C.F. Jeff Wu, Georgia Institute of Technology

Georgia Institute of Technology, School of Industrial and Systems Engineering, Atlanta, GA 30332-0205

roshan@isye.gatech.edu

Key Words: *accelerated life testing, operating window method, process capability, quality engineering, robust parameter design*

Categorical data arise quite often in industrial experiments because of an expensive or inadequate measurement system for obtaining continuous data. When the failure probability (defect rate) is small, experiments with categorical data provide little information regarding the effect of factors of interest and are generally not useful for product or process optimization. An engineering-statistical framework for categorical response optimization is proposed that overcomes the inherent problems associated with categorical data. The basic idea is to select a factor that has a known effect on the response and use it to amplify the failure probability so as to maximize the information in the experiment. New modeling and optimization methods are developed and illustrated with examples.

252 Methodological Issues Surrounding International Comparisons

International Association of Survey Statisticians, Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section

Tuesday, August 10, 2:00 pm–3:50 pm

Purchasing Power Parities—Statistics to Describe the World

◆ Frederic A. Vogel, The World Bank

The World Bank, 1818 H St., NW, Washington, DC 20433

fvogel@worldbank.org

The International Comparison Program is a coordinated global survey of prices used to compute Purchasing Power Parities which remove the distortions caused when using exchange rates to compare the relative sizes of economies, and to measure poverty levels between countries. The methodology used to determine the basket of items to price, where to price them and the number of times in over 120 countries and the sample and survey framework will be described. The issues faced in determining the sample and survey framework will be presented along with a discussion of methodological issues faced when coordinating this effort across countries with differing capabilities.

Methodological Issues in Conducting Cross-national Attitude Surveys

◆ Roger Jowell, City University

City University, Centre for Comparative Social Surveys, Gloucester Building (G501), School of Social Sciences, Northampton Square, London, EC1V 0HB UK

r.jowell@city.ac.uk

Key Words: *comparative research, attitude-monitoring, survey methods, equivalence*

Comparability in cross-national attitude surveys is difficult to achieve. Cultural incompatibilities and methodological variations affect questionnaire design, sampling, fieldwork, response rates and coding. Problems of 'non-attitudes' and social desirability biases abound. The 22-nation European Social Survey - started in 2001 and funded by the European Commission, the European Science Foundation and the academic funding agencies of all participating nations - tries to mitigate these problems. In its first round, the ESS achieved rigorous probability samples in all countries of a similar "effective" sample size and set high standards in translation, fieldwork and response rates. Accompanying the main dataset is a meta dataset containing details of contact and calling strategies, a file of event and context information, and a comprehensive and transparent technical report, making it a rich resource for methodological research. Its fully documented dataset was made available online in September 2003, only months after the end of fieldwork. The combination of careful design, embedded expertise, ambitious targets, and detailed protocols will be fully described and documented.

Techniques for Calibrating Response Scales Across Countries and Languages

◆ Tom W. Smith, NORC, University of Chicago; Kirk M. Wolter, NORC, University of Chicago

NORC, University of Chicago, 1155 East 60th St., Chicago, IL 60637

smitht@norc.uchicago.edu

Key Words: *cross-national, translation, surveys, responses*

Differences in languages are a major impediment in achieving functional equivalency in cross-national survey research. The difficult issue of translation has been made harder by the relative lack of attention to the testing and standardization of the response scales used in surveys. This paper examines a pilot study in the U.S. and Germany that carries out experiments on how words and phrases used in various response scales are evaluated and used in those countries. Particular attention is given to whether specific response terms (e.g. Somewhat Agree, Strongly Agree) cut an agree-disagree continuum at equivalent points.

253 Genetic Algorithms in Statistical Methodology and Business Applications

Business and Economics Statistics Section, Section on Quality and Productivity, Section on Statistical Graphics, Section on Statistical Computing

Tuesday, August 10, 2:00 pm–3:50 pm

An Overview of Genetic Algorithms and Their Applications

◆ Sangit Chatterjee, Northeastern University

Northeastern University, 219 Hayden Hall, Boston, MA 02130

s.chatterjee@neu.edu

Key Words: *binary digits, stochastic optimization, evolutionary operators, natural selection, statistical modeling*

Genetic Algorithms (GA) are stochastic optimization tools that work on Darwinian models of population biology and are capable of solving for near-optimal solution for multivariable functions without the usual mathematical requirements of strict continuity, differentiability, convexity, and other properties. A GA typically begins by randomly choosing a number of candidate solutions which propagate themselves through a “selection criteria” and are changed by genetic operators. Gas have seen some statistical applications and have been used to solve discrete optimization problems with a considerable degree of success. This talk will give a very broad outline of genetic algorithms and an even broader view of its applications potential. Some difficulties, limitations, and open questions will conclude the talk.

Using Genetic Algorithms to Improve the Construction of Classification, Regression, and Survival Tree Models

◆ J. Brian Gray, University of Alabama; Guangzhe Fan, University of Alabama

University of Alabama, Dept. of IS, Statistics, and Mgt. Science, 300 Alston Hall, Tuscaloosa, AL 35487-0226

bgray@cba.ua.edu

Key Words: *bagging, CART, data-mining, genetic algorithm, random forests*

Tree-structured modeling is a valuable tool for predictive modeling and data mining. Traditional tree-growing methodologies such as CART suffer from “greediness,” i.e., local optimizations at nodes to be split in a decision tree do not always result in global optimization of the tree model. CART solutions are also sensitive to perturbations in the data and can vary greatly across different training sets sampled from the same data. Bagging, random forests, and other ensemble techniques have improved on the predictive performance of CART, but they do not have the interpretability of single-tree models. We describe the TARGET (Tree Analysis with Randomly Generated and Evolved Trees) method of constructing tree models for classification, regression, and survival data. Empirical evidence suggests that TARGET produces smaller trees with better predictive performance than CART. TARGET solutions are also found to be more stable than CART solutions across different training samples from the same data.

An Analysis of Genetic Algorithms for Time Series Pattern Discovery

◆ Vasant Dhar, New York University

New York University, Stern School of Business, Information Systems Dept., 44 W. 4th St., New York, NY 10012-1126

vdhar@stern.nyu.edu

Key Words: *genetic algorithms, automated pattern discovery, machine learning*

Genetic algorithms offer a powerful mechanism for learning from data, based on ideas from Darwinian Natural Selection. I will describe a genetic algorithm called GLOWER and its properties, and some results from using it to learn patterns from financial time series data. In particular, I shall describe the representation it uses for learning patterns, and results relative to those obtained via several other machine learning methods on a specific dataset. I shall also describe real-world experiences from using learned patterns to manage a global financial futures fund completely systematically. The latter are research in progress, including conjectures on when and why learning algorithms tend to “overfit” time series data.

254 Ways to Increase the Number of U.S. Resident Students Who Obtain PhDs in Statistics and Biostatistics ◬ ▦

Section on Teaching Statistics in the Health Sciences, Section on Statistical Education

Tuesday, August 10, 2:00 pm–3:50 pm

An Approach to Increasing Enrollments of American Residents in Statistics and Biostatistics

◆ Edward D. Rothman, University of Michigan

University of Michigan, 3560 Rackham Building, Ann Arbor, MI 48109

erothman@umich.edu

Key Words: *rate limiting step, graduate programs, cost-containment in education*

We identify five critical steps to growing the number of U.S. resident PhDs in statistics and biostatistics. These include the undergraduate pool of potential applicants, the preparation of this pool for graduate schoolwork, the provision of resources to these students, the curriculum in our graduate school, and the mentoring for the dissertation. Anyone of these five could be the rate-limiting step in the process. To maximize throughput within the resources of the departments requires a coordinated plan. The potential barrier to each step is reduced by creation of a stepping stone. A minor in statistics is used to increase the number of majors. The number of majors feeds the honors program. The curriculum for the honors program is coupled to the first-year graduate program. Overlap between programs reduces cost and eases the transition. Looking at the field as a science and not an application of mathematics provides a way to deal with the “too mathematical” barrier. The presentation will also provide survey results.

Making Statistics Interesting for Undergraduate Mathematics Majors

◆ George W. Cobb, Mount Holyoke College

Mount Holyoke College, South Hadley, MA 01075

GCobb@MtHolyoke.edu

Key Words: *curriculum, recruiting, mathematical statistics, statistics education, undergraduate mathematics*

According to legend, Willy Sutton said the reason he robbed banks was because "That's where the money is." For similar reasons, if we are to recruit students to statistics, we should look to undergraduate mathematics departments, because that's where the courses are. Unfortunately, neither the introductory course nor the traditional upper division sequence in probability and mathematical statistics does a good job of getting mathematically talented students engaged with the methods, concepts, and thinking that show what contemporary statistical practice is like. In my talk I shall address a variety of ways to address this problem by making changes to the undergraduate curriculum.

Including the Excluded: Recruitment and Retention of Underrepresented Minority Doctoral Students

◆ Mary W. Gray, American University

American University, Dept. of Mathematics and Statistics, 4400
Massachusetts Ave. NW, Washington, DC 20016-8050

mgray@american.edu

Key Words: *minorities, doctoral students, recruitment, retention, statistics, biostatistics*

A substantial source of potential statisticians has been largely overlooked in efforts to increase the number of U.S. resident students who obtain PhDs in statistics and biostatistics: namely, the African American, Latino, and Native American community. With the tightening of U.S. visa regulations and the perception of an unwelcoming climate, the number of foreign nationals in doctoral programs will certainly not increase to meet the nation's future needs, making it all the more important that all segments of our resident population are included in recruitment and retention efforts. As many potential students from underrepresented minorities graduate from smaller institutions, many will lack adequate preparation for graduate study even though they may have done very well in the courses they have taken. Departments must recognize this by being prepared to provide support for students to make up prerequisites, thus taking longer to finish their degrees. Equally important is to engage faculty and students in providing a supportive atmosphere, including arranging for such mechanisms as study groups, and a safety net, should emergencies arise.

255 IMS Medallion Session 1 ⌘

IMS

Tuesday, August 10, 2:00 pm–3:50 pm

Variational Inference in Exponential families: Alternatives to MCMC

◆ Michael I. Jordan, University of California, Berkeley

University of California, Berkeley, 401 Evans Hall, Berkeley, CA
94720

jordan@stat.berkeley.edu

Key Words: *graphical models, exponential families, MCMC, variational methods, convexity, bioinformatics*

Markov chain Monte Carlo (MCMC) has played an important role in statistics in recent years. MCMC is not a panacea, however, and it is important to explore other methods for probabilistic inference, particularly in the setting of models aimed at large-scale data analysis problems. One general approach, like MCMC having its origins in statistical physics, is provided by variational methodology. Variational methods express computations as the solutions to optimization problems, and derive approximations by "relaxations" of this optimization problem. These methods are particularly powerful in the context of exponential family distributions, where tools from convex analysis and convex optimization come into play. They also are particularly natural in the setting of graphical models, where the graphical structure of the model can aid in developing relaxations. I tell a story with three interrelated themes: exponential families, graphical models, and variational inference. I will illustrate these ideas with examples taken from bioinformatics and information retrieval. (Joint work with Martin Wainwright.)

256 Rating Models and Balance in Competition ▲

Social Statistics Section, Section on Statistics in Sports

Tuesday, August 10, 2:00 pm–3:50 pm

Measuring the Impact of Free Agency on Professional Team Sports in North America

◆ Michael J. Mondello, Florida State University; Joel Maxcy, Ithaca College

Florida State University, 114 Tully Gym, Dept. of Sport Management, Tallahassee, FL 32306

mondello@coe.fsu.edu

Key Words: *competitive balance, professional sports, free agency*

This paper considers the effects on competitive balance (CB) given institutional changes permitting professional athletes to become free agents. Economists advocate that free agency will not alter distribution of talent because players remain directed to their highest-valued user. Unrestricted free agent rights have been present in Major League Baseball (MLB) and the National Basketball Association (NBA) since the late 1970s, but were unavailable in the National Football League (NFL) and National Hockey League (NHL) until the 1990s. Empirical tests in MLB offer limited evidence that free agency has changed CB. This study develops time-series regression models for other leagues. Alternative measures of CB accounting for seasonal dispersion in team win-percents and outcomes are derived for the 1951-2002 seasons. Results from the NHL provide clear evidence of improved CB since free agency. Evidence of NFL improvement is just short of statistical significance. There is little effect on the NBA, although imposition of a payroll cap in 1984 has adversely affected CB. Changes in CB given alternate free-agent rules varies across leagues, but in no case is CB critically affected.

Rating and Competition in Soccer

◆ Ruud Koning, University of Groningen

University of Groningen, Dept. Economics, PO Box 800, Groningen, 9700AV The Netherlands

r.h.koning@eco.rug.nl

Key Words: *rating models, competitive balance, soccer*

Competitive balance in soccer is a topic that has been getting a lot of attention the last couple of years. Smaller teams claim that the proceeds of the sale of national television rights are not distributed evenly. Teams from smaller countries claim that they are not receiving a fair part of the proceeds of international competitions. The burst of the telecom boom has left its marks on soccer as well. I measure the development of competitive balance in soccer over time on three levels. First, on the national level, then on the level of international competitions for teams, and finally on the level of national teams. I make a distinction between team qualities and home advantage. I try to assess whether important changes in the institutional environment like the Bosman ruling have resulted in any marked changes of the balance in competition.

Competitive Balance in Golf

◆ Scott M. Berry, Berry Consultants

Berry Consultants, 5124 Bellerive Bend, College Station, TX 77845

scott@berryconsultants.com

Key Words: *Bayesian analysis, hierarchical models, smoothing splines*

In all sports in which an objective measure is available athletes in each successive era are better than the previous era. In sports where there are no objective measures it is more difficult to judge the level of competition through time. I analyze the results of golf tournaments over the last 75 years addressing the level of ability of the golfers. I measure the strength of the fields and the level of individual players through time. I employ Bayesian hierarchical models with smoothing splines for the effects of aging and the quality of the "field."

257 Data Monitoring Committees - Best Practices ▲ ⌘

Biopharmaceutical Section

Tuesday, August 10, 2:00 pm–3:50 pm

Data Monitoring Committees—an NIH Approach

◆ Lawrence Friedman, National Heart, Lung, and Blood Institute

National Heart, Lung, and Blood Institute, Bldg. 31, Room 5A47, 31 Center Dr., MSC 2490, Bethesda, MD 20892-2490

friedmal@nhlbi.nih.gov

NIH has used data monitoring committees (DMCs) since the mid-1960s. At first they were only used for selected clinical trials and varied considerably in how they were constituted and functioned. Over time, their use broadened, and today they are used for all late-phase trials. Many early phase trials, such as those with particularly invasive or dangerous interventions or vulnerable populations, also have DMCs. Even those NIH trials without DMCs must have data monitoring plans. A DMC's primary responsibility is to make sure that trial participants are not unnecessarily harmed. Secondarily, it ensures the trial's scientific integrity. DMCs usually monitor data from single trials. Exceptions are networks, where they monitor all studies conducted by the network, and trials done at NIH, where DMCs will monitor many trials in particular scientific areas. Members of DMCs are independent of investigators and sponsoring NIH institutes. They should have no major financial or other conflicts of interest. Members have expertise in the field under investigation, biostatistics and, often, bioethics or patient advocacy. Operational procedures of NIH DMCs will be described.

Experiences with Interim Analyses and DSMBs—a CRO Perspective

◆ John C. Conlon, Covance, Inc.

Covance, Inc., Clinical Development Services, 210 Carnegie Center, Princeton, NJ 08540-6233

john.conlon@covance.com

Key Words: *O'Brien-Fleming, Lan-DeMets, Lan-Trost, power reassessment, Type I error, interim analysis*

Data management and biometrics support (including interim analyses) for DSMBs forms an integral part of life within our CRO. Challenges can be administrative as well as technical, and development of statistical models can be paramount to both. This paper illustrates some of the experiences and challenges encountered whilst planning and analyzing two large phase III oncology studies. Administrative challenges discussed include the organization of DSMB meetings at times to match required numbers of endpoints, development of DSMB communication plans and contingency planning for all outcome eventualities. Technical statistical challenges from the two studies include analysis using O'Brien Fleming implementation of Lan & DeMets procedures and the implementation of a conditional power reassessment according to Lan & Trost (1997). The latter covers not only the establishment of futility analysis criteria and the determination of the actual Type I error rate, but also the discussions with the FDA in getting the procedure approved.

Why Specify Best Practices? The Case for Anarchy

◆ Janet Wittes, Statistics Collaborative, Inc.

Statistics Collaborative, Inc., 1710 Rhode Island Ave., NW, Washington, DC 20036

janet@statcollab.com

Key Words: *Data Monitoring Committees, clinical trials*

Bad practice in Data Monitoring Committees (DMCs) is easy to spot. Unclear tables, inaccurate data, biased reporting, failure to provide adequate timely information about serious adverse events, or obvious conflicts of interest quickly provide clues to the members of an experienced DMC that the information they are receiving may lead them to nonoptimal decisions. Thus, a first good step in a search for best practices is a statistical version of our medical colleagues' rule: first, do nothing wrong. Good practice is also easy to recognize ("I know it when I see it", as Justice Stewart said in another context). The search for "best practices," however, may be elusive—what is best depends on the disease under study, the state of knowledge of the biology, and the culture of the medical specialty. This paper explores a series of basic principles that should help guide the newcomer to DMCs. It takes the view that many quite different practices are acceptable statistically and ethically as long as all parties involved understand the purpose of the committee and give each other the latitude to perform their responsibilities efficiently.

258 Bayesian Nonparametric and Semiparametric Biostatistical Methods ▲

Section on Statistics in Epidemiology, Section on Bayesian Statistical Science, WNAR

Tuesday, August 10, 2:00 pm–3:50 pm

Nonparametric Bayesian Data Analysis: Inference for Differential Gene Expression

◆ Peter Mueller, University of Texas MD Anderson Cancer Center

University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Dpt. Biostatistics, Houston, TX 77030

pm@odin.mdacc.tmc.edu

Key Words: *density estimation, microarray, mixture models*

We review methods and models used in nonparametric Bayesian inference, using as a motivating example the analysis of microarray data for differential gene expression. The discussion will highlight advantages and limitations of inference based on full (posterior) probability models for the relevant unknown distributions. We will compare resulting inference with ad-hoc exploratory data analysis, with a similar parametric empirical Bayes approach proposed in Efron et al. (2001), and with fully parametric finite mixture models. We will discuss problems of the empirical Bayes inference that are mitigated by a nonparametric Bayesian approach. As a specific example of nonparametric Bayesian modeling we will focus on an extension of Dirichlet process mixtures of normal distributions. As in most nonparametric Bayesian inference, efficient computation is key to a successful implementation. We will review computational problems and strategies that are relevant for a wide class of nonparametric Bayesian models.

Bayesian Semiparametric Models for Survival Data with Time-dependent Covariates

◆ Wesley O. Johnson, University of California, Davis; Timothy E. Hanson, University of New Mexico

University of California, Davis, 369 Kerr Hall, Dept. of Statistics, Davis, CA 95616

wojohnson@ucdavis.edu

Key Words: *time-dependent covariates, mixture of Polya trees prior, Markov chain Monte Carlo methods*

We discuss three models for survival data with time-dependent covariates(TDC); one by Cox and Oakes (1984), the second by Cox (1972), and a new model for TDCs that parallels Cox's original construction, only for the Accelerated Failure Time Model (AFT). We develop a Markov chain Monte Carlo approach based on modeling the baseline survival distribution as a mixture of Polya Tree processes for each model. A substantive illustration is presented using Cerebral Edema data, with emphasis on the practical implementation and use of these models.

Bayesian Nonparametric Modeling for Spatial Data

◆ Athanasios Kottas, University of California, Santa Cruz; Alan E. Gelfand, Duke University; Steven MacEachern, The Ohio State University

University of California, Santa Cruz, Baskin School of Engineering, 1156 High St., Santa Cruz, CA 95064

thanos@ams.ucsc.edu

Key Words: *dependent Dirichlet processes, Dirichlet process mixture models, Gaussian processes, nonstationarity, spatial generalized linear models*

Customary modeling for continuous point-referenced spatial data assumes a Gaussian process, which is often taken to be stationary. Here, we propose a random spatial process that is neither Gaussian nor stationary. We first develop a dependent Dirichlet process (DP) model for spatial data. Next, we introduce mixing through dependent DPs and examine properties of the resulting models. In the Bayesian framework, posterior inference is implemented using Markov chain Monte Carlo (MCMC) methods. Spatial prediction raises interesting questions but can be handled using the MCMC output. The hierarchical nature of our modeling allows important extensions to nonparametric spatial generalized linear models (GLMs). The dependent DP framework yields interesting correlation structures and provides more flexible distributional specifications than traditional spatial glms, e.g., binomial response and Poisson count models. Applications include modeling and prediction for disease incidence employing the dependent DP model for the spatial variation in the probability of disease.

259 Recent Developments in Statistical Methods for Drug Synergy ▲ ⌘

Biopharmaceutical Section, Section on Physical and Engineering Sciences

Tuesday, August 10, 2:00 pm–3:50 pm

A History of Response Surface Models for Assessing the Nature and Intensity of Interaction among Multiple Agents

◆ William R. Greco, Roswell Park Cancer Institute; Donald B. White, University of Toledo; Leonid A. Khinkis, Canisius College

Roswell Park Cancer Institute, Elm & Carlton Sts., Dept. of Biostatistics, Buffalo, NY 14263

William.Greco@RoswellPark.org

Key Words: *synergy, additivity, antagonism, response surface, pharmacometrics*

Pairwise/multiple agent interaction assessment is important for the disciplines of pharmacology, toxicology, epidemiology, food science, and others. Heated disagreements are common in this field. Although it is our view that the future of the field will be dominated by response surface approaches, there is no current consensus on the best RSM. Characteristics for a good RSM for multiple agent interaction include: (a) ability to accommodate arbitrary patterns in dose response slope, maximum effect, and minimum effect parameters; (b) ability to accommodate irregularly shaped isobols at different effect levels, including simultaneous regions of synergism, additivity, and antagonism on the same surface; (c) ability to accommodate antagonism, potentiation, inhibition, and/or coalism between specific agent pairs in a complex mixture; (d) ability to accommodate three or more agents; (e) a hierarchical structure; (f) good statistical characteristics for parameter estimation, model-fitting, and interaction quantitation; and (g) good accompanying graphical procedures. This review will set the stage for reports on current work in the field. Supported by NIH RR10742 and CA16056.

Modeling Effects of Agent Combinations Using Nonlinear Mixture Experiment Methods

◆ Donald B. White, University of Toledo; William R. Greco, Roswell Park Cancer Institute; Leonid A. Khinkis, Canisius College

University of Toledo, Dept. of Mathematics, University Hall 2040, Toledo, OH 43606-3390

donald.white@utoledo.edu

Key Words: *mixture amount model, process variable, nonlinear regression, synergism, antagonism, response surface*

In pharmacology and toxicology, modeling the effects of chemical agents is important for developing optimal experimental designs, making effect predictions, optimizing treatments, and increasing scientific understanding. Extending such models to multiple agents is important for research in cancer therapy, antibiotic treatment, herbal remedies, toxic waste hazards, and many other applications. We introduce a new hierarchical approach to modeling the effects of multiple agents using a combination of nonlinear response surface modeling with mixture experiment methods. This approach allows us to model more than two agents, modulating agents having no effects singly but which enhance the effects of other agents, and complex patterns of synergism and antagonism. We are also able to produce effective descriptive graphics, well-grounded statistical inference, and methods for creating optimal experimental designs. These results are illustrated with three studies: a large in-vitro study of three cancer drugs, another large in-vitro study of two cancer drugs and a modulator, and a small in-vivo study of agents affecting cholesterol levels in mice. Supported by NIH RR10742 and CA16056.

Multiple Drug Interactions—A Flexible Approach

◆ Matthew L. Fidler, University of Utah; Steven E. Kern, University of Utah

University of Utah, 421 Wakara Way, Suite 318, Salt Lake City, UT 84108

fidler@math.utah.edu

Key Words: *response surface, synergy, antagonism, additivity, interaction, modeling*

Drug therapy strategies maximize the desired effect(s) and minimize the negative side effect(s) often employing combination therapies. To classify, predict, and summarize these interactions, modeling is used. Modeling classification is based on Loewe standards, which adherence to caused inflexibility in changes in maximum effect and Hill slope. Minto suggested a model that deviated from this allowing flexibility, butleaving behind the single-constants to indicate interaction type. We propose a model similar to Minto's that allows single-parameter interaction statements, compares real data fits to Minto's model, and compares the error between Loewe reduction and Minto's noninteracting reduction. The following conclusions are made:(1) the current model fits as well as Minto (compared by AIC), (2) the model allows approximate assessment of additivity as proposed by Loewe, and (3) the model allows flexibility in surfaces, and (4) the model allows parameter-based statements, letting statistical tests such as interaction type and asymmetry to be performed.

Statistical Methods for Analysis of Mixtures of Many Chemicals Using a Ray Design

◆ Chris Gennings, Virginia Commonwealth University; Hans Carter, Virginia Commonwealth University

Virginia Commonwealth University, Dept. of Biostatistics, 1101 E. Marshall St., Richmond, VA 23290-0032

gennings@hsc.vcu.edu

Key Words: *synergy, dose-response, interaction*

The motivating strategy of our research program is to develop statistical techniques that are useful in detecting departure from additivity in mixtures with many chemicals while maintaining the fundamental definition of additivity as found in the toxicology literature. By focusing the inferential region to relevant mixing ratios of the chemicals through use of a ray design, it becomes experimentally practical for studies to be conducted of mixtures with many components. Single chemical dose-response data are used to estimate an additivity model for a fixed-ratio mixture. An "unconstrained" model is fit to observed mixture data along the ray. A test of the coincidence of these two dose-response curves is a test of additivity. A variation of this approach is demonstrated when the parametric form of the underlying response surface is assumed. Methods are described for testing hypotheses regarding the impact of subsets of chemicals in the mixture. A statistical model allowing for an interaction threshold parameter is presented. Finally, we describe a method of analysis where the additivity model is built implicitly. Each method is illustrated with various toxicology studies.

Robust Design Strategies for Bioassay and Drug Synergy

◆ Timothy E. O'Brien, Loyola University Chicago

Loyola University Chicago, Dept. of Math & Statistics, 6525 N. Sheridan Rd., Chicago, IL 60626

tobrien@math.luc.edu

Key Words: *optimal design, generalized nonlinear modeling, drug synergy, relative potency, interaction, oncology*

The validity of biomedical and scientific research depends upon the accuracy of the statistical models and the efficiency of the experimental designs used by these researchers. Optimal design theory provides researchers with the means to select the "optimal" settings for their experiment in the sense of yielding a design or plan that will require the fewest repetitions to yield accurate results. However, since statistical models are often not known with complete certainty, these designs may be inappropriate and often have no ability to signal significant lack of fit. Thus, so-called robust designs are desired to estimate model parameters and accurately make predictions, on the one hand, and to highlight model misspecification, on the other. The focus of this paper is to outline recent developments in modeling and design strategies for the detection of relative potency and synergy for situations where the assumed (nonlinear) model function, the initial parameter estimates, or the error structure is not completely known.

260 Innovations in Sample Weighting

Section on Government Statistics, Section on Survey Research Methods

Tuesday, August 10, 2:00 pm–3:50 pm

Evaluating Alternative Calibration Schemes for an Economic Survey with Large Nonresponse

◆ Chadd Crouse, National Agricultural Statistics Service; Phillip S. Kott, National Agricultural Statistics Service

National Agricultural Statistics Service, 3251 Old Lee Hwy., Room 305, Fairfax, VA 22030-1504

ccrouse@nass.usda.gov

Key Words: *nonresponse, quasi-random model, coverage adjustment, delete-a-group jackknife*

The U.S. Department of Agriculture conducts an annual economic survey of farm operations called the Cost and Returns Report (CRR). This survey is fairly long and is based on a multiphase sample with several opportunities for nonresponse. As a result, CRR response rates are low by government standards. The USDA has historically "trued up" the estimated number of farms within a region and economic-size class using more reliable aggregates determined from other sources. Recently, the two agencies involved in the survey have begun employing a more sophisticated linear calibration scheme to force a number of the CRR estimates to match outside totals, not just a few mutually exclusive farm counts. Delete-a-group jackknife methodology is being used to evaluate alternative calibration schemes.

Composite Response Rates for Surveys with Nonresponse Follow-up

◆ Jill A. Dever, RTI International; Avinash Singh, RTI International; Vincent G. Iannacchione, RTI International

RTI International, 1615 M St. NW, Suite 740, Washington, DC 20036

jdever@rti.org

Key Words: *composite response rate, dual-frame estimation, weight calibration*

Unweighted response rates (RR) can sometimes lead to biased conclusions about the response propensity of a surveyed population. Using the AAPOR RR3 definition, we calculated a 54.4% RR to a stratified sample of 10,301 Gulf War veterans (GW). This RR was almost 10 percentage points higher than the corresponding design-weighted response rate (WRR) of 44.9% due to oversampling females and reservists. In an effort to reduce potential NR bias, we conducted a telephone follow-up of 1,000 mail nonrespondents, achieving a 55.5 RR. For a survey with a NR follow-up, the WRR offers a way to measure the combined response to the initial survey and the follow-up. While the usual double-sampling estimator provides an unbiased estimate of the combined RR, it may be unstable especially if the follow-up subsample is small. We show that the ideas underlying dual-frame estimation together with weight calibration produce a composite RR for the GW survey that is more efficient than a double-sampling estimator.

Mean Squared Error Estimation for the Coverage and Nonresponse-adjusted U.S. Census of Agriculture

◆ Matthew J. Fetter, U.S. Dept. of Agriculture; Phillip S. Kott, National Agricultural Statistics Service

U.S. Dept. of Agriculture, 14th St. & Independence Ave., Room 4818, Washington, DC 2025

Key Words: *calibration, quasi-randomization model, integerization, smoothing*

The National Agricultural Statistics Service weights the records on the Census of Agriculture to adjust for nonresponse and then conducts a second calibration step to account for the undercoverage of its census mailing list. After that, the nonresponse- and coverage-adjusted weights are integerized. Since they are in part based on sparse area samples, the indications used to determine the calibration targets are smoothed using components-of-variance modeling. We will outline how NASS is computing the mean squared error of its estimated census totals acknowledging these various sources of measurable error.

A Model-based Approach to Weight-trimming

◆ Michael R. Elliott, University of Pennsylvania School of Medicine

University of Pennsylvania School of Medicine, Center for Clinical Epidemiology and Biostatistics, 612 Blockley Hall, 423 Guradian Dr., Philadelphia, PA 19104

melliott@cceb.upenn.edu

Key Words: *sample survey inference, sampling weights, regression estimators, random effects*

In unequal-probability-of-selection sample, correlations between the probability of selection and the sampled data can induce bias. Weights equal to the inverse of the probability of selection are often used to counteract this bias. Highly disproportional sample designs have large weights, which can introduce unnecessary variability in statistics such as the population mean estimate. Weight-trimming reduces large weights to a fixed cutpoint value and adjusts weights below this value to maintain the untrimmed weight sum. This reduces variability at the cost of introducing some bias. Standard approaches are not "data-driven": they do not use the data to make the appropriate bias-variance trade-off, or else do so in a highly inefficient fashion. This presentation develops Bayesian methods for weight-trimming to supplement standard, ad hoc design-based methods in disproportional probability-of-inclusion designs where variances due to sample weights exceeds bias correction. These methods are used to estimate linear and generalized linear regression model population parameters in the context of stratified and poststratified known-probability sample designs.

261 GAP Analysis for Geospatial Biodiversity Assessment: Improving Statistical Methodology ▲

Section on Statistics and the Environment

Tuesday, August 10, 2:00 pm–3:50 pm

Techniques for Classifying and Understanding Vegetation Alliances

◆ Michael D. Jennings, U.S. Geological Survey/University of Idaho

U.S. Geological Survey/University of Idaho, 530 S. Asbury St., Suite 1, Moscow, ID 83843

jennings@uidaho.edu

Key Words: *vegetation alliance classification, Mantel, MRPP, NMDS*

Little is known about ecological characteristics of alliances described in the U.S. National Vegetation Classification, either generally or individually, yet they are commonly the fundamental unit of vegetation used in large-area spatial datasets of habitat and biodiversity. Few alliance descriptions are based directly on original field plot data, and such data are rarely available for reexamination. This presentation focuses on applications of the Mantel Test, Non Metric Multidimensional Scaling, and Multi-Response Permutation Procedures in (a) evaluating and testing standardized field plot data from heterogeneous sources, and (b) measuring the relationship between the dominant plant species of alliances (which are used to identify and delineate the occurrence of an alliance) and the understory species which make up the composition of an alliance. The work presented here is based on an initial dataset containing almost 40,000 field plot records of 11 different sources. Additional datasets of climate, biomass productivity, and morphological traits of plant species were integrated with the field plot data. The techniques presented are central to achieving a numerical description.

Nationwide Mapping of Forest Type and Biomass: Comparing Nonparametric Modeling Tools

◆ Gretchen Moisen, U.S. Forest Service

U.S. Forest Service, 507 25th St., Ogden, UT 84401

gmoisen@fs.fed.us

Key Words: *forest inventory, MARS, GAM, CART, RandomForest, satellite imagery*

The USDA Forest Service Forest Inventory and Analysis program (FIA) collects forest inventory data on a systematic grid across all lands in the United States. Moderate resolution maps of forest biomass and forest type are currently being developed nationwide by modeling response data collected on FIA sample plots as nonparametric functions of numerous ancillary predictor layers. These layers include: 16-day Moderate Resolution Imaging Spectrometer (MODIS) composites and associated vegetation indices and MODIS percent tree cover; vegetative diversity and type synthesized from the National Land Cover Dataset; topographic variables derived from Digital Elevation Models; monthly and annual climate parameters; and other ancillary variables. The nonparametric modeling tools of choice for these production mapping efforts include See5 for the categorical response (forest type) and Cubist for the continuous response (biomass). We compare the performance of these two tools against other nonparametric alternatives using data collected in the interior Western United States.

An Investigation into the Use of Maximum Posterior Probability Estimates for Assessing the Accuracy of Large Maps

◆ Brian M. Steele, University of Montana

University of Montana, Dept. of Mathematical Sciences, Missoula, MT 59802

steele@mso.umt.edu

Key Words: *land cover map, classification, cross-validation, prediction rule, accuracy assessment*

Assessing the accuracy of land cover maps using conventional methods is often prohibitively expensive because of the difficulty of adequately sampling the range of geographic and environmental variation. Often, probability sampling is not feasible, and conventional accuracy estimators based on the proportion of correctly classified observations among a sample of map units with known land cover type are unreliable. An alternative approach using maximum posterior probability (MPP) estimators was investigated in this research. This talk introduces MPP-based estimators and reports on a comparative study of conventional and MPP estimators for a land cover mapping problem involving nine Landsat TM scenes. We simulated conventional and MPP-based accuracy estimators derived from post classification samples and training sample cross-validation. The results show that substantial reductions in the mean square error can be obtained from MPP-based estimators compared to conventional estimators. Additionally, MPP-based estimators produced reasonably good accuracy estimates even when biased designs were used to draw the training samples.

Classification Approaches for Modeling Animal Species Distributions

◆ Thomas C. Edwards, Utah State University

Utah State University, Utah Cooperative Fish & Wildlife Research Unit, USGS Biological Resources Division, Logan, UT 84322-5290

tce@nr.usu.edu

Key Words: *classification models, species presence absence, gap analysis, conservation planning*

Classification techniques are powerful tools for modeling linkages of animal species with habitat type, especially when coupled with spatially explicit environmental data structures. Unfortunately, a common problem confronting such models is adequate presence absence data for model building. Too often absence data, irrespective of whether collected from a design or purposive-based approach, are lacking, essentially eliminating the application of classification tools for modeling purposes. Several different approaches exist for generating so called "pseudo-absence" data, each with its own assumptions and resultant impacts on the final classification models. I evaluate the effect of two different approaches for generating "pseudo-absence" data on resultant animal distribution models, comparing: (1) random absences generated from within the known range of collected presence data, and (2) a weighted approach based on density isoclines of presence data. Example species for analysis and evaluation come from the Utah Gap Analysis data structure, as well as other environmental data structures used in ongoing conservation planning.

262 Statistical Graphics and Computing Sections Student Competition Papers

Section on Statistical Computing, Section on Statistical Graphics

Tuesday, August 10, 2:00 pm–3:50 pm

Analysis of Nonlinear Variation in Thermal Performance Curves

◆ Rima Izem, University of North Carolina, Chapel Hill; J. Stephen Marron, University of North Carolina; Joel G. Kingoslver, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Statistics and Operations Research Dept., New West 107, CB #3260, Chapel Hill, NC 27599-3290

rizem@email.unc.edu

Key Words: *functional data analysis, evolutionary biology, analysis of variation, nonlinear variation, Frechet mean and variance, common shape curves*

Thermal performance curves are an example of common shape curves from evolutionary biology. Three modes of variations in these curves are of interest to biologists because each mode induces a different evolutionary response of the population to selection. Since two of those modes are nonlinear, traditional methods of analysis of variance in curves such as principal component analysis fail to decompose the variation into biological modes or to quantify each mode. We present the result of a new method: Template Mode of Variation in the decomposition of the variation in the thermal performance curves into modes of variation of biological interest.

Implementation of Backward Induction for Sequentially Adaptive Clinical Trials

◆ Jay K. Wathen, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77030-4009

jkwathen@mdanderson.org

Key Words: *backward induction, implementation, clinical trial*

In many clinical trials, patients are enrolled and data are collected sequentially, with interim decisions based on the accruing data. This naturally leads to application of Bayesian sequential procedures for trial monitoring. I will discuss the implementation and computational tasks involved in the use of backward induction for making decisions during a clinical trial. I will address general computational needs, and illustrate the ideas with specific results for a two-arm trial with a binary outcome and a sample size of 200 patients. In addition, simulation results will be provided.

The Billion Byte Brain: Combining Physiological Data and Gigabytes of Images to Improve Maps of Brain Activity

◆ Kary L. Myers, Carnegie Mellon University

Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213

kary@stat.cmu.edu

Key Words: *brain imaging, optical imaging, image analysis, large datasets, physiological noise, video data*

Scientists use optical imaging to make maps of stimulus-induced brain activity. However, physiological fluctuations unrelated to the stimulus can mask the changes due to activation. I will present a method for overcoming the physiological noise to reveal the signal of interest. Two features distinguish this work from that reported by others. First, the size of the data I consider is much larger; while a typical article reports experiments with about 12 megabytes of raw data, the experiment I analyze produced over six gigabytes of raw data (almost 25 gigabytes of floating point numbers). As I write this, my working directory contains 750 gigabytes of intermediate results. Second, I present a scientifically appealing way of reducing the noise masking the signal: instead of using decompositions like PCA to separate signal from noise, I augment the images with simultaneously recorded physiological measurements. My current procedure, multiple linear regression, is simple statistically but challenging computationally: I perform almost 20 million individual regressions with 20,000 observations each. By cleaning the data of physiological noise, I hope to make better maps of brain activity.

Reduction Algorithm for the MLE for the Distribution Function of Bivariate Interval-censored Data

◆ Marloes H. Maathuis, University of Washington

University of Washington, Dept. of Statistics, Box 354322, Seattle, WA 98195-4322

marloes@stat.washington.edu

Key Words: *computational geometry, parameter reduction, maximal clique, maximal intersection, regions of possible mass support*

We study computational aspects of the nonparametric maximum likelihood estimator (MLE) for the distribution function of bivariate interval censored data. The computation of the MLE consists of two steps: a parameter reduction step and an optimization step. We focus on the reduction step. We introduce two new reduction algorithms: the Tree algorithm and the HeightMap algorithm. The Tree algorithm is only mentioned briefly. The HeightMap algorithm is discussed in detail and also given in pseudo-code. It is a very fast and simple algorithm of time complexity $O(n^2)$. This is an order faster than the best known algorithm thus far, the $O(n^3)$ algorithm of Bogaerts and Lesaffre (2003). We compare our algorithms with the algorithms of Gentleman and Vandal (2001), Song (2001) and Bogaerts and Lesaffre (2003) using simulated data. We show that our algorithms, and especially the HeightMap algorithm, are significantly faster.

263 Bayesian Application to Health Economics ▲

Section on Health Policy Statistics, Section on Bayesian Statistical Science

Tuesday, August 10, 2:00 pm–3:50 pm

Use of Cost-effectiveness Elasticity to Address the Problem of Small Denominators in Cost-effectiveness Analysis

◆ Ya-Chen T. Shih, University of Texas M.D. Anderson Cancer Center; Scott B. Cantor, University of Texas M.D. Anderson Cancer Center; Peter Muller, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, Dept of Biostatistics & Applied Mathematics, 1515 Holcombe Blvd. Box 196, Houston, TX 77030

yashih@mdanderson.org

Key Words: *cost-effectiveness analysis, Bayesian cost-effectiveness analysis, cost-effectiveness elasticity, incremental cost-effectiveness ratio, small denominators*

The conventional CEA summarizes results in an incremental cost-effectiveness ratio (ICER) and compares it with a specific threshold to determine whether a new intervention is cost-effective. The ICER measures the absolute difference in costs against the absolute difference in effectiveness; it does not differentiate small denominators caused by poor clinical benefit (bad intervention) from those due to treating conditions with a limited lifespan to show clinical improvement (bad disease). Therefore, decisions based on ICER may be biased against the latter type of interventions. We propose a relative CE measure called the CE-elasticity to address this issue. We used the Bayesian approach to analyze a simulated data with the same incremental effectiveness for diseases with very short, short, and moderate life expectancy; the estimated posterior mean CE-elasticity for these diseases was 0.47, 2.68, and 5.45, respectively. The probability that the new intervention had an elasticity larger than 2 is 0.63, 0.50, and 0.48, respectively. Results showed that the new CE measure differentiate bad interventions from bad diseases in CEA and thus may minimize biased decision-making.

Bayesian Hierarchical Analysis of Health Services Outcomes Data

◆ Benjamin N. Bekele, University of Texas M.D. Anderson Cancer Center; Linda Elting, University of Texas M.D. Anderson Cancer Center; Catherine Cooksley, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd.-447, Houston, TX 77030

bbekele@mdanderson.org

In many health services research studies, observations are correlated, resulting in dependence among outcomes (e.g., subjects nested within hospitals or measurements nested within patients). We will discuss Bayesian approaches that can handle three levels in a hierarchy (i.e., patients nested within surgeons nested within hospitals). Within this Bayesian modeling context, we will discuss methods that can be used to accommodate “imperfect nesting” (i.e., some surgeons practice at more than one hospital). We will give two examples where we fit Bayesian hierarchical regression models (binary outcomes and survival outcomes) from data obtained through public-use databases. We show how to fit such models using WinBugs 1.4 and how to use the new capabilities in WinBugs (e.g., calling WinBugs from S-plus) to more automate the analysis process. (In the past, running univariate models could be very time-consuming.)

Cost-utility Analysis of HIV Prevention Interventions: Estimates and Uncertainties

◆ Matthew J. Hayat, National Institutes of Health; Prakash Laud, Medical College of Wisconsin; Raymond G. Hoffmann, Medical College of Wisconsin; Ana Johnson-Masotti, McMaster University

National Institutes of Health, NIDCD, 6120 Executive Blvd., MSC 7180, Bethesda, MD 20892

hayatm@nidcd.nih.gov

Key Words: *HIV, cost, quality of life, Bayesian, effectiveness, utility*

Cost-utility analyses provide useful guidance to policymakers in comparing different treatments for a disease or in allocating societal resources to alleviating the ill effects of different diseases. We first describe a randomized study conducted to assess the effectiveness of three different strategies designed to increase protective behavior in mentally ill adults. Next we give an overview of statistical measures used in cost-utility analyses, including a discussion of some related pitfalls. A further complication arises when costs or utilities cannot be measured directly, as in the case of HIV prevention programs. Observable behavioral changes are translated via models into QALYs (quality adjusted life years) saved by a prevention program. These models contain parameters about which only partial information is available. We consider inference models that combine this information with an accounting of the sampling variability in the behavioral changes. Resulting methods are illustrated on data from the above mentioned HIV prevention study.

Bayesian Inverse Decision Theory

◆ Dennis D. Cox, Rice University; Kalatu Davies, Rice University; Scott B. Cantor, University of Texas M.D. Anderson Cancer Center; Richard Swartz, University of Texas M.D. Anderson Cancer Center

Rice University, Dept. of Statistics MS-138, 6100 Main St., Houston, TX 77005

dcox@rice.edu

Key Words: *Bayesian sequential methods, Bayesian inference, medical decision-making, cost-benefit analysis, inverse decision theory*

Inverse decision theory concerns inferences about losses given a presumed optimal decision rule. In most settings, this provides only linear constraints on the losses. We consider settings where we have prior knowledge about the losses which is captured by a prior distribution on the losses. We show how this can be combined with the constraints from an inverse decision theoretic analysis to obtain a posterior on losses. Generalizations include (a) sequential procedures, where one also considers costs associated with data collection, and (b) uncertainty about the probabilistic model for the decision problem, which is also treated from a Bayesian perspective to obtain a unified approach to the overall problem. An application is given to the current standard of care for detecting precancerous lesions of the uterine cervix.

Use of Bayesian Methods in Budget Impact Analysis

◆ Shu Han, Rice University; Ya-Chen T. Shih, University of Texas M. D. Anderson Cancer Center

Rice University, 6100 Main St., MS-138, Houston, TX 77005

shuhan@rice.edu

Key Words: *inhomogeneous Markov chain, budget impact analysis, Bayesian, generic entry, economic evaluations*

An inhomogeneous Markov chain model was proposed to include variations in patients (pts) mix and drug prices in BIA using Bayesian methods. Markov states were categorized by whether a pt was treated with a generic drug, an existing brand-name drug, or a new drug. Modifications were made at each cycle for newly diagnosed incident cases and exiting cases due to cure or death. Also considered is the difference in treatment preference b/t the current and newly diagnosed pts. A case study from simulated data was used to compare the budget impact of including vs. excluding a new drug in a health plan, taking a payer's perspective and a five-year time frame. Results were presented in a probabilistic plot similar to cost-effectiveness acceptability curve. Adding the new drug to the plan was found to increase the budget increase in the short run but not in the long run. The probability that including the new drug would increase in the budget by 10% is 9%, 26% in a one- and two-year time frame, and it becomes cost neutral in the five-year time frame. Our model provides a framework to examine time-varying parameters in BIA and generates estimates that better reflect the real health care market.

264 Design and Analysis Issues for Cancer Studies ▲ ⌘

Biometrics Section, Biopharmaceutical Section

Tuesday, August 10, 2:00 pm–3:50 pm

An Improved Design for Clinical Trials Evaluating Multiple Agents

◆ Christine E. McLaren, University of California, Irvine; Vernon M. Chinchilli, Pennsylvania State University; Jennifer J. Roan, Minitab Inc.; Wen-Pin Chen, Chao Family Comprehensive Cancer Center; Frank L. Meyskens, Chao Family Comprehensive Cancer Center

University of California, Irvine, Epidemiology Division, 224 Irvine Hall, Irvine, CA 92697-7550

cmclaren@uci.edu

Key Words: *2x2 factorial design, Cochran-Armitage trend test, simple loop alternative, contrast statistic, Jonckheere-Terpstra test, chemoprevention*

For clinical trials that intend to answer multiple questions, new designs are needed to achieve results with fewer patients and less expense. Adenomatous polyps are precursors to colorectal cancer, the fourth most common incident cancer in the United States. We propose a 2x2 balanced factorial design to test whether a combination of agents, difluoromethylornithine and sulindac, is better than the individual treatments or placebo with regard to reducing the proportion of recurrent adenomas. To compare proportions of patients developing at least one adenoma between treatment groups, we derive a pair of two-sided tests for the ordered alternative of increasing or decreasing proportions across treatment groups, called the simple-loop alternative, and a method for power and sample size calculations. We compare this test with alternative tests under comparable assumptions. The proposed design to study an intermediate marker for cancer incidence, achieves a reduction in sample size of over 30% and should lead to a significant reduction in the cost of the clinical trial. We suggest the use of this experimental design to achieve economy in terms of numbers of patients and monetary expense.

The Use of Mediation Models to Identify and Validate Biomarkers

◆ John Kittelson, University of Colorado Health Sciences Center; Patrick Blatchford, University of Colorado Health Sciences Center

University of Colorado Health Sciences Center, Preventive Medicine and Biometrics, 4200 E. 9th Ave., B-119, Denver, CO 80262

john.kittelson@uchsc.edu

Key Words: *biomarker, mediation model, lung cancer*

Survival rates in lung cancer patients are low because patients have disseminated disease at the time of diagnosis. Biomarkers are needed for early detection, and a major effort in lung cancer research is to identify and validate biomarkers. A very large number of candidate biomarkers are available from biochemical, cellular, and genetic measurements (e.g., DNA microarrays) on bronchial biopsy tissue and sputum samples. The standard approach to identifying biomarkers is to evaluate their expression in subjects with and without lung cancer using measures such as sensitivity and specificity. When evaluating many biomarkers, there is an elevated false positive rate. We discuss the use of mediation models to identify biomarkers that mediate the relationship between smoking and lung cancer. We show that this approach reduces the risk of falsely positive biomarkers. We discuss the implications of this approach to study design, and describe how smokers, nonsmokers, cancer cases, and controls should be selected to support this analysis. The methods are illustrated using biomarkers from bronchoscopy tissue specimens collected as part of the Colorado Lung SPORE cohort.

Estimation of Binary Response with Order Restrictions

◆ Jeremy M.G. Taylor, University of Michigan; Lu Wang, University of Michigan

University of Michigan, Dept. of Biostatistics, 1420 Washington Heights, Ann Arbor, MI 48109

jmgt@umich.edu

Key Words: *biomarkers, isotonic regression, order restrictions, multiple imputation*

We consider the situation of two or more ordered categorical variables and a binary outcome variable, where one or more of the categorical variables may have missing values. The goal is to estimate the probability of response of the outcome variable for each cell of the categorical variables while incorporating the ordering. The probability of response is assumed to change monotonically as each of the categorical variables changes. Let q_ij be the probability that the categorical variables fall in cell ij, and p_ij be the probability of response. A model is developed in which the number of observations in each cell is multinomial(q) and the response is binomial with parameters p_ij for each ij. Three estimation approaches are compared, Gibbs sampling with order restrictions on p_ij, multiple imputation and isotonic regression. This problem is motivated by studies in which multiple biomarkers are measured on stored specimens and the outcome measure is response to treatment. Missing biomarker data is common in such settings. It is also biologically reasonable to assume that the probability of response changes monotonically as the biomarker values change.

Statistical Considerations in the Development and Validation of Prognostic Profiles Derived from Gene Expression Microarray Data

◆ Lisa McShane, National Cancer Institute; Lara Lusa, Istituto Nazionale Tumori of Milan; Michael Radmacher, Kenyon College; Joanna H. Shih, National Cancer Institute; George Wright, National Cancer Institute; Richard Simon, National Institutes of Health

National Cancer Institute, 6130 Executive Blvd., Room 8126, Biometric Research Branch, MSC 7434, Bethesda, MD 20892-7434

lm5h@nih.gov

Key Words: *model validation, microarray, cross-validation, prognostic models, expression profiles*

DNA microarray analysis has been used to develop classifiers based on expression profiles that predict disease outcome for cancer patients. Often some preliminary validation of the predictive power of a profile-based classifier is carried out using the same dataset that was used to derive the classifier. Techniques such as cross-validation can be used in this setting to assess predictive power, and if applied correctly, will result in a nearly unbiased estimate of predictive accuracy of a classifier. However, some investigators conducting these gene expression profiling studies have applied cross-validation techniques incorrectly and have obtained highly biased estimates of predictive accuracy. Also, some cross-validated estimates of predictive accuracy can have very high variance. Alternatively, some investigators have tried to validate predictors by testing the statistical significance of the association between cross-validated predicted classes and disease outcomes. We will discuss the limitations of these various approaches and make recommendations for proper validation strategies.

Clinical Trial Designs for Dose-seeking, Non-MTD Trials with Biomarker Endpoints

◆ Sumithra J. Mandrekar, Mayo Clinic; Wei Zhang, University of Iowa; Susan M. Geyer, Mayo Clinic; Vera Suman, Mayo Clinic; Karla Ballman, Mayo Clinic; Antje Hoering, Mayo Clinic; Daniel Sargent, Mayo Clinic

Mayo Clinic, 200 First St. SW, Kahler 1A, Rochester, MN 55905

mands@mayo.edu

Key Words: *biomarkers, clinical trials, designs, simulations*

Historically, designs for dose-seeking trials have been geared towards finding the maximum tolerated dose (MTD), with safety as the primary outcome. With target-based anticancer agents whose dose response curves are unknown and whose dose/toxicity relationship are expected to be minimal, alternative designs to identify the biologically optimal dose have become necessary. The type of dose response relationship is also usually unknown for these agents. Using simulation, we compare various designs (e.g., cohorts of k individuals, randomization among d predetermined dose levels, continuous reassessment methods, accelerated titration) for determining the optimal dose of a targeted therapy under different types of dose-response relationships (e.g., linear, quadratic, increasing with a plateau). We assume the therapy is nontoxic for the dose range of interest and that the endpoint can be assessed in a relatively easy, reliable, and timely manner. The designs are compared with respect to expected sample size, the number of patients treated at a suboptimal dose, distance the selected dose is from the optimal dose, expected trial duration, plus several other pragmatic considerations.

265 Causality and Bias in Biomedical Studies

ENAR, Section on Statistics in Epidemiology, Biometrics Section, Section on Health Policy Statistics, Section on Statistics and the Environment

Tuesday, August 10, 2:00 pm–3:50 pm

A Formulation of Causal Inference for Population Interventions Based on an Experimental Paradigm

◆ Steven Mark, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd., EPD Rm. 8036, Rockville, MD 20852-4910

sm7v@nih.gov

Key Words: *causal inference, attributable risk, populaton interventions, noncompliance, absolute risk, confounding*

Population interventions are considered when there is evidence that altering the prevalence of an exposure will change the rates of disease; there are feasible mechanisms for altering population levels of exposure; the predicted magnitude in the change in disease incidence is sufficient to justify a particular intervention. I discuss the assumptions required for investigators to accurately predict disease rates in a post-intervention population. The assumptions and inference are formulated in terms of correspondences between the probability distributions of random variables in the pre-intervention and post-intervention populations. The inferential structure is organized to correspond to the usual temporal sequence of population intervention experiments.

Approximate Causality for Longitudinal Binary Outcomes and Nonadherence

◆ Thomas R. Tenhave, University of Pennsylvania; Dylan Small, University of Pennsylvania; Kevin Lynch, University of Pennsylvania; David Oslin, University of Pennsylvania; Jing Cheng, University of Pennsylvania

University of Pennsylvania, CCEB, Blockley Hall Rm 607, 423 Guardian Dr., Philadelphia, PA 19104-6021

ttenhave@cceb.upenn.edu

Key Words: *random effects, logistic, exclusion restriction, randomization, encouragement studies, mental health*

We present a random effects logistic approach for longitudinal binary outcomes that adjusts for longitudinally measured treatment nonadherence in the context of two randomized behavioral intervention trials. Such an approach is an extension of the Nagelkerke et al. (2000) approximation for the cross-sectional binary outcome case, and is a response to requests for as-treated analyses to supplement intent-to-treat longitudinal analyses based on analogous random effects logistic models. By as-treated, we mean contrasting a condition with treatment versus a condition without treatment. For both trials, naively comparing treated and untreated conditions without consideration for the potential biases associated with treatment nonadherence led to unexpected results: the naïve as-treated contrasts were of smaller magnitude than the intent-to-treat contrasts, which is the converse of the typical case when there is a significant intent-to-treat effect. For both trials, we present an as-treated analysis that adjusts for treatment nonadherence such that the as-treated contrast is greater than the intent-to-treat contrast, while retaining the intent-to-treat inference based on p values.

Evaluating and Correcting Guess Effect in not Perfect Double-blinded Clinical Trials

◆ Jianfeng Cheng, Columbia University; Eva Petkova, Columbia University

Columbia University, 1051 Riverside Dr., PO Box 48, New York, NY 10032

jc794@columbia.edu

Key Words: *double-blinded clinical trial, assessing guess effect, correcting guess effect*

Failure to maintain blindness is a common phenomenon in double-blinded clinical trials testing the efficacy of antidepressant medications, which may affect the magnitude of the drug effects and makes the efficiency of psychotropic drugs to unknown degrees untrustworthy. It is most often thought that the bias of the researchers' rating will benefit the active treatment in a placebo controlled experiment or the competing agent in trials where the control is some traditional treatment. We propose a method for assessing the presence of possible rater'sbias due to treatment guessing in antidepressant clinical trials and offer a proposed estimator of treatment efficacy by correcting such guess effect. Simulation studies showed the properties of the proposed estimators: unbiasedness and smaller mean square error than the naive ones. Through the examples, we presented how to apply the methods to the clinical trials, even for those without guess data collected. The proposed approach allows for reevaluation of previously conducted clinical trials and provides a method for measuring and correcting the potential bias.

Analysis of Behavioral Weight Loss Data

◆ Wei Lang, Wake Forest University School of Medicine

wlang@wfubmc.edu

Key Words: *adherence, clinical trial, parametric model, repeated measures*

In clinical trials evaluating the effect of treatment on outcomes measured repeatedly over time, participant's adherence to the treatment is a key issue. Analysis results based on data obtained on participants who adhere to the treatment need to be adjusted in order to draw valid conclusion on the treatment effect. Adherence can be analyzed as either a 0-1 outcome at each time point or as a time-to-event analysis with the event being time to nonadherence. We construct a copula-based parametric model (CPM) for the joint analysis of correlated discrete and continuous outcomes. A separate CPM is also proposed for the joint modeling of time-to-event and longitudinal outcomes. Data from a behavioral weight loss study and postmenopausal estrogen/progestin intervention clinical trial are analyzed using the proposed method.

266 Bayesian Methods for Survival Analysis and Software Reliability

Section on Bayesian Statistical Science

Tuesday, August 10, 2:00 pm–3:50 pm

A Simple Approach to Fitting Bayesian Survival Models

◆ Paul Gustafson, University of British Columbia; Dana Aeschliman, University of British Columbia; Adrian R. Levy, University of British Columbia

University of British Columbia, 333-6356 Agricultural Rd., Vancouver, BC V6T 1Z2 Canada

gustaf@stat.ubc.ca

Key Words: *Bayesian survival analysis, copula model, Markov chain Monte Carlo, semiparametric hazard, time-dependent covariate effects*

There has been much recent work on Bayesian approaches to survival analysis, incorporating features such as flexible baseline hazards, time-dependent covariate effects, and random effects. Some of the proposed methods are quite complicated to implement, and we argue that as good or better results can be obtained via simpler methods. In particular, the normal approximation to the log-gamma distribution yields easy and efficient computational methods in the face of simple multivariate normal priors for baseline log-hazards and time-dependent covariate effects. While the basic method applies to piecewise-constant hazards and covariate effects, it is easy to apply importance sampling to consider smoother functions.

Dynamic Software Reliability Models for Typed Defects

◆ Nalini Ravishanker, University of Connecticut; Bonnie K. Ray, IBM T.J. Watson Research Center; Zhaohui Liu, University of Connecticut

University of Connecticut, Dept. of Statistics, U-4120, 215 Glenbrook Rd., Storrs, CT 06269

nalini@stat.uconn.edu

Key Words: *Bayesian inference, growth curve, power prior*

We describe a model for software reliability characterization of typed defects using a growth curve formulation that allows model parameters to vary as a function of covariate information. For instance, it is possible that an increase in the discovery rate for Type II (assignment/initialization) defects might result as more Type I (checking) defects are found, with possible feedback relationships as well. In general, the dependence of discovery of defects of Type I on an arbitrary number of other defect types can be handled in a dynamic growth curve framework by modeling the reliability growth of Type I defects using covariates to represent the number of defects of other types found up to a certain time. We describe a Bayesian framework for inference and model assessment, using Markov chain Monte Carlo techniques, which allows for incorporation of historical information and expert opinion in the form of prior distributions on the parameters. The use of power priors to combine data-based and qualitative prior information is also introduced in the growth curve modeling context. The methods are illustrated using simulated and real defect data.

Bayesian Analysis of a New Joint Model with Cure Fraction

◆ Debajyoti Sinha, Medical University of South Carolina; Joseph G. Ibrahim, University of North Carolina, Chapel Hill; Ming-Hui Chen, University of Connecticut

sinhad@musc.edu

Key Words: *biological markers, longitudinal data, cure rate survival model*

Vaccine therapies have received a great deal of attentional recently as potential therapies in cancer clinical trials. However, little is currently known about the biological activity of vaccines and that of finding good markers for assessing the efficacy of these vaccines and predicting clinical outcome. The antibody immune measures IgG and IgM have been proposed as potential markers in melanoma clinical trials because of their observed correlation with clinical outcome in pilot studies. To better understand the role of IgG and IgM for a particular vaccine, we must better understand the relationship of clinical outcome to an individuals antibody (IgG or IgM titires) history over time. Gibbs sampling methodology is given for carrying out Bayesian inference. We illustrate our methodology with a recently completed melanoma clinical trials using vaccine.

Bayesian Software Release Strategies

◆ Refik Soyer, George Washington University

George Washington University, Dept. of Management Science, Monroe Hall 403, Washington, DC 22315

soyer@gwu.edu

Key Words: *software reliability, decision analysis, optimal testing*

We address the problem of how to terminate the testing and release the software during the development phase. In so doing, we present a Bayesian decision theoretic approach and develop Monte-Carlo-based methodologies that replace the preposterior analysis by surface fitting methods. We illustrate our approach by considering single and two-period problems.

Stopping Rule for Software Testing Based on a Binary Segmentation Procedure

◆ Tae Yang, Myongji University; Lynn Kuo, University of Connecticut

Myongji University, Dept. of Mathematics, Yongin, Kyonggi, 449-728 Korea

tyang@mju.ac.kr

Key Words: *binary segmentation procedure, deviance criterion, reliability growth pattern, stopping rule*

We develop a method that can be used to evaluate the reliability of software and to determine when to stop testing. In software testing, the time-domain software reliability growth models are assumed to make dynamic changes at each failure and the successive failure rates may be relatively more homogeneous for a series of related tests. The homogeneous clustering of these rates can be carried out using a binary segmentation procedure. The Bayes estimate of a clustered failure rate provides a direct reliability assessment for these clusters. We introduce the notion of reliability growth based on these clusters, where software reliability growth can be explained by the gradual reduction of the clustered growth rates from earlier to later period with perturbations. Stopping rules on software testing are developed based on the clustered reliability growth model.

267 Implementing Federal Legislation and Regulations

Section on Government Statistics, Social Statistics Section

Tuesday, August 10, 2:00 pm–3:50 pm

An Integrated Approach to Implementing Data Stewardship Policies through PIAs

◆ Wendy L. Alvey, U.S. Census Bureau

U.S. Census Bureau, Policy Office, Room 2430 FB 3 MS 3700, Washington, DC 20233

wendy.l.alvey@census.gov

Key Words: *data stewardship, privacy principles, E-Government Act, Privacy Impact Assessments, policy development*

For the past three or four years, the U.S. Census Bureau has been working on an integrated approach to developing corporate data stewardship policies. At the past few JSMs, we have reported on the vision we had in mind, the framework for that effort, and the underlying Privacy Principles that support the data stewardship approach. We have also briefly discussed some of the policies that the Data Stewardship Executive Policy Committee (DSEP) has developed to address gaps in current practice identified by those Privacy Principles and described some preliminary work we have done on developing a tool to capture the results as part of the E-Government Act's required Privacy Impact Assessments (PIAs). The DSEP's focus over the past couple of years has been on establishing these new internal policies to strengthen the Census Bureau's role as a good data steward. Of about 20 policies under development, just over 20% have been adopted. The problem is that that is when the hard work begins—we have learned that implementation involves a whole lot more than just announcing the new policy. This paper will talk about the Census Bureau's approach to implementation of DSEP policies, designed to ensure that the policies are applied efficiently and consistently Bureauwide, with a minimum of additional burden to the individual program areas. It will focus on the use of the PIA process to integrate the various policies with existing processes.

From PIA to Policy: Administrative Records and Record Linkage

◆ Lisa M. Blumerman, U.S. Census Bureau

U.S. Census Bureau, Policy Office, Room 2430 FB 3 MS 3700, Washington, DC 20233

Lisa.M.Blumerman@census.gov

Key Words: *administrative records, record linkage, privacy and confidentiality, data stewardship*

Vital to the U.S. Census Bureau's livelihood is the proper management, support, and use of administrative records data. These files are maintained by the source agencies for the purpose of administering programs and providing services. The Census law encourages us to use administrative records to improve and enhance our programs. Their use, however, carries responsibilities. The Census Bureau has taken great strides to develop a Data Stewardship Program and its respective policies. Our concern for privacy and confidentiality have led to increased safeguarding and data protection measures. One example is the Census Bureau's implementation of Privacy Impact Assessments (PIAs), required as part of the E-Government Act. PIAs serve as documentation of our coordinated approach to the integration of data stewardship policies. This paper will focus on one specific section of the PIA—the use of administrative records and record linkage techniques—and discuss how this legislation translated into policy. Additionally, I describe the legal and policy requirements that relate to the acquisition, use, and protection of administrative records data and subsequent record linkages.

A Policy Application: Using Administrative Records to Supplement Census Bureau Programs

◆ Sally M. Obenski, U.S. Census Bureau; Ron Prevost, U.S. Census Bureau

U.S. Census Bureau, Office of Planning, Research and Evaluation Div., Room 1107 FB2 MS 9200, Washington, DC 20233

Sally.M.Obenski@census.gov

Key Words: *administrative records, demographic programs, linkage*

Administrative records coupled with state-of-the-art record linkage methodologies are proving to be a powerful tool for supplementing the Census Bureau's demographic, decennial, and geographic programs. For example, rather than supplanting a traditional decennial census as envisioned earlier in the administrative records program's history, research and analysis results point instead to a supporting role. Likewise, instead of replacing traditional household surveys with administrative data that provide similar information, it appears that administrative data and their analyses are most effective in strengthening and augmenting direct data collection programs. Administrative records have long been considered a potential means for reducing cost and respondent burden while improving data quality. Key events have shaped the current program solving substantive problems heretofore having vexed advances in administrative records uses. Due to these breakthroughs, administrative records are already improving existing programs and emerging as powerful tools for an array of survey, census, and program uses.

Performance Measurement: A Technical Perspective

◆ Clyde Tucker, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Rm. 1950, Washington, DC 20212

tucker_c@bls.gov

Key Words: *program evaluation, GPRA, market value, quality indicators, measurement error*

Performance-based evaluation is, in principle, a great step toward government accountability. As usual, however, the devil is in the details. This paper focuses on issues surrounding the implementation of the Government Performance Review Act (GPRA) in federal statistical agencies. The paper does not address political issues concerning GPRA, but, instead, deals with technical concerns. The topics covered include, but are not limited to, the measurement of future costs and benefits of statistical programs, determining the market value of government products, measuement issues involved in developing indicators of performance, and the resources required to carry out the GPRA mandates.

NCES Implementation of GPRA via Customer Satisfaction Surveys

◆ Sameena M. Salvucci, Synectics for Management Decisions, Inc.; Stephen R. Wenck, Synectics for Management Decisions, Inc.; Susan Hamsher, Synectics for Management Decisions, Inc.; Arnold Goldstein, National Center for Education Statistics; Barbara Marenus, National Center for Education Statistics; Daniel Slattery, National Center for Education Statistics

Synectics for Management Decisions, Inc., 1901 N. Moore St., Suite 900, Arlington, VA 22209

sams@smdi.com

Key Words: *service standards, strategic planning, performance measures, performance indicators, performance planning*

The National Center for Education Statistics (NCES), as well as many other areas of the government, is required under the Government Performance and Results Act (GPRA) enacted on August 3, 1993, to "provide for the establishment of strategic planning and performance measurement." In support of its mission, NCES has established GPRA goals and objectives in three functional performance areas: collection, analysis, and dissemination of information on the condition of education in the U.S. and comparative international statistics. Between 1996 and 2003 NCES conducted five customer satisfaction surveys. This paper will describe how results from these surveys were used as performance indicators of the Center under GPRA. The paper will also describe how NCES has used these performance indicators to identify areas of performance planning. Finally, the paper will provide "lessons learned" about the design of the survey in relationship to its usefulness for providing relevant information for performance planning.

268 Kernel Smoothing ⌘

IMS

Tuesday, August 10, 2:00 pm–3:50 pm

A Note on the Interpolated Kernel Density Estimates

◆ Chien-Tai Lin, Tamkang University; J.S. Wu, Tamkang University; C.H. Yen, Tamkang University

Tamkang University, Dept. of Mathematics, Tamsui, 251 Taiwan

chien@math.tku.edu.tw

Key Words: *frequency polygons, integrated mean squared error, nonparametric density estimation, smoothing*

We investigate the sufficient condition for the interpolated kernel density estimate (IKDE) to be a probability density. Under the condition, the appropriate class of the kernel functions and the ratio of grid distance versus smoothing parameter can be easily derived. The asymptotic integrated mean squared error properties of IKDE in this class are also studied. We finallyshow that the optimal IKDE can be obtained if the kernel functions are the finite-degree polynomials of absolute-valued variables with a specified ratio.

Local Isotonic Regression

◆ Derick R. Peterson, University of Rochester

University of Rochester, Dept. of Biostatistics & Computational Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

peterson@bst.rochester.edu

Key Words: *local monotonic regression, nonparametric regression, local polynomial regression*

A new Local Isotonic (LOIS) regression estimator is proposed for the standard nonparametric regression problem. The idea is to approximate the regression function at each point by a local isotonic regression rather than, say, a local polynomial one. Since this local model is clearly more flexible than that of local linear regression, itself a locally monotone model, the LOIS regression estimator has smaller bias at most points. Moreover, in stark contrast to local polynomial regression, LOIS regression is very insensitive to the choice of bandwidth unless the regression function oscillates frequently. And if the data are monotone then, for any bandwidth, LOIS reproduces the data, is thus conditionally unbiased at each observation, and is equivalent to the nonparametric maximum likelihood estimator for monotone regression functions. Since the price to be paid for the added flexibility of the local model is increased variability, and asymptotically this price is a steep one, the performance of LOIS for practical sample sizes is investigated via a simulation study.

Nonparametric Coefficient of Determination

◆ Li-Shan Huang, University of Rochester

University of Rochester, Dept. of Biostatistics, URMC, 601 Elmwood Ave., Box 630, Rochester, NY 14642

Lhuang@bst.rochester.edu

Key Words: *local linear regression, nonparametric regression, coefficient of determination, R-squared, degree of freedom*

A new nonparametric coefficient of determination (R-squared) is proposed for local linear regression. It aims to measure the proportion of variation of a response explained by a nonparametric curve fitted by local linear regression for a given bandwidth. An adjusted nonparametric R-squared is also presented by adjusting the sums of squares with the corresponding degree of freedom. Asymptotic results are given as a theoretical basis for the new nonparametric R-squared. We examine practical performance of the new nonparametric R-squared via simulations, and compare with a correlation-based nonparametric R-squared studied by Doksum and Samarov (1995).

Data Sharpening under Unimodal Constraint in Kernel Density Estimation

◆ Kee-Hoon Kang, Hankuk University of Foreign Studies; Peter Hall, Australian National University

Hankuk University of Foreign Studies, Mohyun, Yongin, 449-791 Korea

khkang@hufs.ac.kr

Key Words: *unimodal density, data sharpening, density estimation*

We discuss a robust data-sharpening method for rendering a standard kernel estimator, with a given bandwidth, unimodal. It has theoretical and numerical properties of the type that one would like such a technique to enjoy. In particular, we show theoretically that, with probability converging to~1 as sample size diverges, our technique alters the kernel estimator only in places where the latter has spurious bumps, and is identical to the kernel estimator in places where that estimator is monotone in the correct direction. Moreover, it automatically splices together, in a smooth and seamless way, those parts of the estimator that it leaves unchanged and those that it adjusts. Provided the true density is unimodal our estimator generally reduces mean integrated squared error of the standard kernel estimator.

Local Likelihood Intensity Estimation

◆ Jennifer L. Prokop, University of Western Ontario

University of Western Ontario, Dept. of Statistics & Actuarial Science, Western Science Centre, London, ON N6A 5B7 Canada

jprokop@stats.uwo.ca

Key Words: *point processes, smoothing*

In a series of papers, Brillinger has demonstrated the usefulness of nonparametric intensity function estimation in the study of temporal point process data. An alternative approach is local likelihood estimation using a flexible conditional intensity model. Local polynomial approximations are used for the logarithm of the conditional intensity and, as in Hjort and Jones (1996), we define a local kernel-smoothed log likelihood function. Parameter estimation is done via maximizing the localized version of the log likelihood. We show that, for the class of self-exciting processes, the local constant cross-intensity function estimator is not equivalent to Brillinger's cross-intensity estimator. Simulation is employed to compare the quality of the intensity function estimators.

269 Can We Measure America's Position and Progress? Challenges in Developing National Indicators ▲

Social Statistics Section

Tuesday, August 10, 2:00 pm–3:50 pm

Can We Measure America's Position and Progress? Challenges in Developing Key National Indicators

◆ Martha Farnsworth Riche, Farnsworth Riche Associates; ◆ Christopher W. Hoenig, General Accounting Office; ◆ Robert Groves, University of Michigan; ◆ Enrico Giovannini, Organization for Economic Cooperation and Development; ◆ Geoffrey F. Lee, Australian Bureau of Statistics

Australian Bureau of Statistics, Room 4N520, ABS House, Locked Bag 10, Belconnen, Canberra, ACT 2616 Australia

geoff.lee@abs.gov.au

In today's increasingly complex national and global environment, there is a growing need for high-quality, accessible, and accurate information to help the public, business and civic leaders, and the government make critical choices. A large and growing amount of dialogue and activity has taken place throughout the United States and around the world on key national, state and local indicators and how to develop them in a manner that is truly useful to communities and society. To advance this dialogue, the U.S. General Accounting Office, in cooperation with the National Academies, hosted a February 2003 Forum which provided the impetus to the Key National Indicators Initiative. This panel will discuss challenges and progress to date at the local, national, and international level.

270 Integrating the Advanced Placement Statistics Exam into the Higher Education Curriculum

Section on Statistical Education

Tuesday, August 10, 2:00 pm–3:50 pm

Integrating the Advanced Placement Statistics Exam into the Higher Education Curriculum

◆ Allison Clark, Advanced Placement Program, College Board; ◆ Brad Hartlaub, Kenyon College; ◆ Jessica Utts, University of California, Davis; ◆ Calvin Williams, Clemson University

CALVINW@clemson.edu

Key Words: *Advanced Placement, higher education, K-12 statistics education, introductory statistics course*

The purpose of this panel is to present current thinking behind the curriculum development of the Advanced Placement Statistics course and exam, AP exam-scoring procedures, research on AP Statistics student performance, and data on AP Statistics exam credit and placement policies at a variety of different higher education institution types. Over 58,000 students took the AP Statistics exam in 2003 (total 2003 AP examinees was over 1 million). AP Statistics student performance is often not adequately assessed due to the lack of a clear department and academic path once the student has entered college. Given that an introductory statistics course is required in many fields such as the social sciences, health sciences, and business, it is important that such departments, along with math and statistics, consider AP Statistics exam scores. The goals of the panel include: provide greater familiarity with the AP Statistics exam; discuss how the AP Statistics exam can be better utilized to bridge the gap between statistics education in high school and college; and discuss how AP Statistics can help foster student interest in the study of statistics into their post-secondary academic career.

271 Missing Data and Density Estimation Methods ⌘

General Methodology

Tuesday, August 10, 2:00 pm–3:50 pm

Nonimputational Technique for Parameter Estimation on Missing Data

◆ Sergey S. Tarima, University of Kentucky; Yuriy G. Dmitriev, Tomsk State University

University of Kentucky, 817 Patterson Office Tower, Dept. of Statistics, Lexington, KY 40506-0027

stari@ms.uky.edu

Key Words: *parameter estimation, missing data, nonimputational technique for missing data*

In the paper a new nonimputational approach to parameter estimation on missing data is proposed. This approach combines parameters' estimates obtained from different parts of a sample defined by corresponded missing data patterns. Improvement of the suggested estimators achieved by involving the effect of mutual correlation between estimators derived from different parts of a sample. New estimators are asymptotically unbiased and provide asymptotically smaller variances. This approach is developed for the case of an ignorable missing data mechanism but it also can be applied for a some nonignorable missing data situations. Applying the suggested methodology to a dropout situation with missing completely at random a well-known Kaplan-Meier estimator is derived.

Inference on Three-parameter Lognormal Distribution Based on Progressively Censored Data

◆ Indrani Basak, Pennsylvania State University, Altoona; Prasanta Basak, Pennsylvania State University, Altoona; Narayanaswamy Balakrishnan, McMaster University

Pennsylvania State University, Altoona, 151 CLRC, 3000 Ivyside Park, Altoona, PA 16601

i8b@psu.edu

Key Words: *EM algorithm, progressive censoring, maximum likelihood estimators, asymptotic distribution, missing information principle*

Some work has been done in the past on the three-parameter lognormal distribution estimation based on complete and censored samples. We develop the inferential methods based on Type II progressively censored samples from a three-parameter lognormal distribution. In particular, we use EM algorithm as well as some other numerical methods to determine maximum likelihood (ML) estimates of parameters. We also consider the interval estimation based on large-sample theory and determine the actual probability coverages in case of small samples. The asymptotic variances and covariances of the ML estimates are computed by using missing information principle. The methodology developed is then illustrated using a numerical example.

Random-coefficients Model with Missing Data

◆ Yongyun Shin, University of Michigan; Stephen W. Raudenbush, University of Michigan

University of Michigan, Dept. of Statistics, 439 West Hall 550 East University, Ann Arbor, MI 48109

choil@umich.edu

Key Words: *random coefficient, maximum likelihood, missing data, hierarchical data, EM algorithm, Fisher scoring*

A random-coefficients model is estimated with two-level hierarchical data when level-1 slopes vary randomly across level-2 units. We suggest the estimation method for a random-coefficients model when both individual and cluster level data are subject to missingness with a general missing pattern. Our method considers two-level data where the observations within each cluster are dependent. We maximize the observed data likelihood via a mixture of EM algorithm and Fisher scoring to obtain the maximum likelihood estimates (MLE) of the parameters using all available data. The key assumptions are that the data at the within-cluster or cluster level, or both, are missing at random; that parameter spaces for the complete data model and missing data mechanism are distinct; and that the data subject to missingness are multivariate normal conditional on all observed data. In order to translate back to the random-coefficients model of interest, we suggest two methods: (1) direct estimation of the parameters from the joint MLE by a subset regression on a disjoint subset of the complete data; and (2) multiple model-based imputation based on the joint MLE.

Tests of Independence for Incomplete Contingency Tables

Kenneth J. Koehler, Iowa State University; ◆ Shin-Soo Kang, KwanDong University

KwanDong University, Kang-won Do, South Korea

sskang@iastate.edu

Key Words: *fractional imputation, multiple imputation, Wald statistics*

Imputation procedures such as fully efficient fractional imputation (FEFI) or multiple imputation (MI) can be used to construct complete contingency tables from samples with partially classified responses. Variances of FEFI estimators of population proportions are derived, and Wald and deviance tests of independence are proposed. Simulation results, when data are missing completely at random, reveal that FEFI provides more efficient estimates of population proportions than either MI, based on data augmentation, or complete case analysis, which simply ignores partially classified cases. Neither FEFI nor MI provide more powerful tests of independence than complete case analysis. Comparisons are also made with maximum likelihood estimation of population proportions, using both complete and partially classified cases, and corresponding tests of independence.

Controlling Bayes False Discovery Rate

Sanat K. Sarkar, Temple University; ◆ Tianhui Zhou, Temple University

Temple University, 7605 B Massey Way, Elkins Park, PA 19027-1009

tzhou@temple.edu

Key Words: *Bayes false nondiscovery rate, normal prior, posterior false discovery rate, posterior nondiscovery rate*

Given a certain prior distribution of the parameters, the Bayes false discovery rate (BFDR) of a multiple testing procedure is defined in this article as the average of the false discovery rate (FDR) of the procedure with respect to this prior. The Bayes false nondiscovery rate (BFNR) is similarly defined. A new multiple testing procedure controlling the BFDR is developed. Its performance is compared to those of the Benjamini-Hochberg procedure, which also controls the BFDR in the independence as well as some positive dependence cases, and the one that controls the Bayesian FDR in terms of BFNR.

On the Density of the Solution to a Random System of Equations

◆ Anthony Almudevar, University of Rochester

University of Rochester, Dept. of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY 14642

anthony_almudevar@urmc.rochester.edu

Key Words: *density approximation, small sample asymptotics*

A general form of the density f of the solution ^{θ} to a random system of equations $\psi(X,\theta) = 0$ was derived by Skovgaard (1990), Jensen and Wood (1998) and Almudevar, Robinson and Field (2000). I will discuss a general method of obtaining higher-order approximations of f suggested by these results.

Determining the Number of Components in a Finite Mixture of Lifetime Distributions under Progressive Censoring

◆ Prasanta Basak, Pennsylvania State University, Altoona; Indrani Basak, Pennsylvania State University, Altoona

Pennsylvania State University, Altoona, 3000 Ivyside Park, Altoona, PA 16601-3760

fkv@psu.edu

Key Words: *finite mixtures, progressive censoring, maximum likelihood estimates, lifetime data*

A wide range of lifetime data has been found to be represented by a finite Weibull or lognormal mixture models. In fitting finite mixture models, an important but difficult problem is determination of number of components in the mixture. Recently, progressive censoring scheme has been very popular in lifetime data because of its effectiveness concerning time and money. We investigate the problem of determing the number of components in a mixture of lifetime distributions when samples are progressively censored. Once the numbers of components are determined, the parameters are estimated from the resulting finite mixture of distributions. The methodology is applied to the finite mixture of two popular models in lifetime-Weibull and lognormal distributions.

272 Spatial Inference ▲

Section on Bayesian Statistical Science

Tuesday, August 10, 2:00 pm–3:50 pm

Gaussian Process Models for the Sphere

◆ Margaret B. Short, Los Alamos National Laboratory

Los Alamos National Laboratory, 3262 Walnut A, Los Alamos, NM 87544

mbshort@lanl.gov

Key Words: *spatial process, process convolutions, spherical model, Markov chain Monte Carlo, atmospheric science*

Spatial processes are often described with a Gaussian process model, which may be modeled via the convolution of a white noise process and a smoothing kernel. Usually, a Cartesian coordinate system (flat-earth assumption) is employed. We explore the validity of this assumption by defining convolution processes on the sphere and compare various compactly supported kernels against their flat earth counterparts. Ours is a Bayesian approach which uses Markov chain Monte Carlo (MCMC) for estimation. We illustrate using examples from atmospheric science.

Normal Observables with CAR Priors Having Two Types of Neighbor Relations

◆ Brian J. Reich, University of Minnesota; James S. Hodges, University of Minnesota; Brad Carlin, University of Minnesota

University of Minnesota, A460 Mayo Building, 420 Delaware St. SE, Minneapolis, MN 55455

brianr@biostat.umn.edu

Key Words: *Bayesian, spatial, conditional, autoregressive, prior*

Bayesian analyses of areal data often use a conditionally autoregressive (CAR) prior distribution which allows fitted values to be smoothed toward values of neighboring regions. Sometimes it is desirable to have more than one type of neighbor relation in the spatial structure, so the different types of neighbor relations can induce different degrees of smoothing. For example, in the periodontal datasets, the degree of smoothing of neighbor pairs bridging the gap between teeth may be different from the smoothing of pairs that do not bridge such gaps. We develop a two neighbor relation CAR (2NRCAR) model to handle this situation, and present associated theory to help explain the sometimes unusual posterior behavior of the parameters that control the different types of smoothing in this model. We also illustrate use of this model by applying it to an analysis of some periodontal data on attachment loss.

Generalized Hierarchical Multivariate CAR Models for Areal Data

◆ Xiaoping Jin, University of Minnesota; Brad Carlin, University of Minnesota

University of Minnesota, Division of Biostatistics, A460 Mayo Bldg., MMC 303, 420 Delaware St. SE, Minneapolis, MN 55455

xiaopinj@biostat.umn.edu

Key Words: *areal data, conditionally autoregressive (CAR), hierarchical Bayesian model, Markov chain Monte Carlo, multivariate data, spatial statistics*

In the fields of medicine and public health, a common application of spatial areal models is the study of geographical patterns of disease. When we have several measurements recorded at each spatial location (for example, $p>2$ diseases available from the same population groups or regions), we need to consider multivariate spatial areal models, in order to handle the dependence among the multivariate components, as well as the spatial dependence between sites. We propose a new flexible class of generalized multivariate conditional autoregressive (GMCAR) models for areal data, and show how it enriches the existing MCAR class. Our approach differs from earlier ones in that it directly specifies the joint distribution for a multivariate Markov random-filed (MRF) through specification of simpler conditional and marginal models. This in turn leads to a significant reduction in the computational burden in hierarchical spatial random effect modeling, where posterior summaries are computed using Markov chain Monte Carlo. We compare our approach with existing MCAR models in the literature via simulation, using average mean square error.

Process Approximation for Analysis of Large Spatial Datasets

◆ Gangqiang Xia, Duke University; Alan E. Gelfand, Duke University

Duke University, PO Box 90251, ISDS, Old Chem. Bldg., ISDS, Durham, NC 27708

xia@stat.duke.edu

Key Words: *spatial process, kernel mixing, subsampling, MCMC*

We consider a customary spatial model with mean structure, Gaussian spatial process, and nugget. When the number of observations N grows large, likelihood-based inference will become unstable and, eventually, infeasible since it involves computing various forms of a large covariance matrix. If we are to fit a Bayesian model and implement some MCMC algorithm, the large covariance matrix will make repeated calculations very slow. We refer to this computational problem as the “Large N problem.” We will review a number of ways to deal with the “Large N problem.” We propose a discrete kernel mixing approximation model. We adopt a Bayesian framework. We examine the accuracy of our approximation model and compare it with the subsampling model. Examples will be given to illustrate the methods. We will also discuss different representations of the spatial random field and the spatial design for the subsampling strategy.

Urban Sprawl/Effect and Chronic Obstructive Pulmonary Disease: A Bayesian Spatial Analysis with Maps

◆ Gerald Shoultz, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 3402, Hyattsville, MD 20782

bjz9@cdc.gov

Key Words: *spatial analysis, urban sprawl, COPD, chloropleth maps*

Prior research which uses spatial and Bayesian methods for modeling age-specific mortality rates for Chronic Obstructive Pulmonary Disease (COPD) suggest that some urban effect may be an important explanatory variable in such modeling. Indeed, a cursory look at the literature indicates that rates of asthma and other COPD’s tend to be higher in urban areas than other areas. Recent studies indicate a potential association between urban sprawl/development and obesity, with residual effects for such health conditions/diseases as hypertension, cardiovascular disease, etc. This research addresses two issues in the study of sprawl/development: (1) Measurement: How is urban sprawl/development measured? And (2) Impact: How does urban sprawl/development relate to mortality rates of COPD? Census 1990 and 2000 data and other data are used to develop indices of urban sprawl/effect; such indices are used as covariates to model COPD mortality data for White and non-White individuals (by gender) in the 1999-2001 period using Bayesian and Spatial methods. Chloropleth maps of mean COPD rates are presented, as are variation maps showing areas where inferences may not be reliable.

On Bayesian Assessment of Curvilinear Boundaries

◆ Sudipto Banerjee, University of Minnesota

University of Minnesota, A460 Mayo Building, MMC-303, 420 Delaware St. SE, Minneapolis, MN 55455

sudiptob@biostat.umn.edu

Key Words: *spatial models, spatial gradients, boundary assessments, Bayesian inference, Markov chain Monte Carlo, Gaussian process*

Boundary analysis concerns the detection and analysis of zones of abrupt change in spatial maps. Its importance in understanding scientific phenomena has been widely recognized in genetics and ecology dating back to Womble. Although contour-mapping visually diplays lines separating such zones, these current methods are based upon rather ad hoc deterministic algorithms. This talk focuses upon a framework to carry out formal statistical inference for statistically assessing whether curvilinear boundaries (or segments thereof) delineate regions with differing responses. Line integrals of gradient processes, arising out of parent Gaussian processes, are employed to compute average gradients along boundaries. Using posterior samples from typical Markov chain Monte Carlo output from fitted spatial models, posterior distributions of analytically intractable Gaussian line integrals are computed enabling formal statistical inference. Examples of the methodology will be presented through simulated as well as real data.

Predicting Gaussian Fields with Unknown Covariance Structure

◆ Elizabeth C. Shamseldin, University of North Carolina, Chapel Hill; Richard L. Smith, University of North Carolina

University of North Carolina, Chapel Hill, Dept. of Statistics, CB 3260, 117 New West, Chapel Hill, NC 27599-326

shamseld@email.unc.edu

Key Words: *spatial model, Bayesian analysis, kriging, MCMC, spatial covariance structure, covariance estimate error*

Spatial models are concerned with the underlying covariance structure between a collection of measurements. The theory can be applied to a number of areas including environmental applications such as monitor readings of an atmospheric pollutant such as particulate matter. Traditional spatial systems are modelled as Gaussian random fields whose covariances are some function of the distance between two monitor sites (for example, the exponential or Matern models). “Kriging” is a method of spatial interpolation that uses the covariances to construct optimal estimators of the random field at unobserved locations. However, traditional interval estimates using kriging assume the spatial covariance structure is known, ignoring the possible error in estimating the parameters of a spatial model. Bayesian methods provide one possible resolution of this difficulty, but in general it is unknown whether a Bayesian prediction interval has the correct frequentist coverage probability. We consider traditional kriging methods, Bayesian MCMC methods, and analytic approximations to Bayesian methods to calculate prediction intervals for spatial interpolation.

273 Missing Data and Dropout in the Analysis of Longitudinal Data

Biometrics Section, Section on Statistics in Epidemiology

Tuesday, August 10, 2:00 pm–3:50 pm

Do Mixed Model and Multiple Imputations Work Together for Longitudinal Data Analysis with Missing Values?

◆ Hrishikesh Chakraborty, RTI International; Yaping Cai, University of North Carolina, Chapel Hill; Peg Burchinal, University of North Carolina, Chapel Hill

RTI International, 3040 Cornwallis Rd., Durham, NC 27709

hchakraborty@rti.org

Key Words: *multiple imputations, mixed models, missing values*

Multiple imputations are a simulation-based approach to filling in missing values so that complete-data analysis methods may be employed. In contrast, mixed models are used to analyze data with missing dependent variable values. Studies frequently generate missing values for the dependent and independent variables. Including a large number of independent variables with missing values in mixed model analysis will exclude significant number of data vectors from analysis which creates a major problem of drawing meaningful conclusions. We used multiple imputations to impute missing values for all independent variables in a longitudinal study setting and used the complete data matrix to fit mixed model independently for each of the imputed datasets. We present a method to combine multiple estimates and inferential statistics generated from multiply imputed datasets using the same mixed model with random intercepts and slopes. We account for the within and between imputations variability in a mixed model setting and incorporating several variance-covariance components. We conclude that in some situations it is desirable to use both methods together to draw conclusion.

A Two-part Model for Longitudinal Data with Missing Values

◆ Yeonjoo Yi, Tulane University; Leann Myers, Tulane University

Tulane University, 1440 Canal St. Suite 2001, New Orleans, LA 70112

yyi@tulane.edu

Key Words: *semicontinuous variable, longitudinal data, two-part model, generalized estimating equations (GEE), missing values, simulation study*

Semicontinuous variables have a proportion of responses which are a single value (often zero) and the remaining responses which follow a continuous, often skewed, distribution. In a two-part model, the first level models the probability that the semicontinuous variable takes on its point mass value, and the second level models the distribution of the variable given that it is not at its point mass. The two parts are then combined into a single model. We extend this two-part model approach to longitudinal settings with missing observations at both levels. We use generalized estimating equations regression methods (GEE) with a logit link to predict who will respond at different time points. Using only the nonzero responses, GEE methods using a normal link are used to predict the mean level of response at different time points. The two parts are then combined to obtain a single prediction model. We will present epidemiological examples of this two-part model for longitudinal studies by GEE with missing values. Limitations are also discussed.

Analysis of Longitudinal Data with Nonmonotone, Nonignorable Missing Data

◆ Lin Wang, University of Wisconsin, Madison; Mari Palta, University of Wisconsin, Madison; Jun Shao, University of Wisconsin

University of Wisconsin, Madison, 7102 Longmeadow Rd., Dept. of Statistics, Madison, WI 53717

wangl@stat.wisc.edu

Key Words: *longitudinal, missing data, IPW*

Nonmonotone missing outcomes present a difficult problem for longitudinal data analysis. When the missing probability depends on outcomes at previous time points, those outcomes could also be missing, and hence the missingness is nonignorable. Selection model approaches leading to weighted analyses have been proposed. Some of these exclude observations to make missingness monotone or assign individual level weights to subjects with complete observations only. We show that such methods can be quite inefficient. We extend Robins' method (1997), to the longitudinal case and apply separate inverse probability weights to each observation within an individual. We then use GEE with independent working correlation structure to obtain estimates. We also consider application of a regression-based, pattern mixture approach, where the regression parameters are estimated from a conditional mean function without specification of the missing mechanism.

Markov Transition Models for Binary Longitudinal Data with Missing Values

◆ Xiaowei Yang, BayesSoft, Inc.; Kun Nie, BayesSoft, Inc.; Steven Shoptaw, BayesSoft, Inc.; Qing J. Zhang, BayesSoft, Inc.

BayesSoft, Inc., 3641 Midvale Ave. #207, Los Angeles, CA 90034

XYang@BayesSoft.com

Key Words: *missing data, longitudinal data analysis, Markov chain, transition model*

In longitudinal studies, binary repeated measures are commonly met. Statistical analyses for such datasets are often plagued by problems with missing values due to nonresponses or withdrawal. For example, due to the chaotic nature of substance abuse research, many participants would miss their clinic visits or drop out of the studies prematurely. When all measures are observed, Markov transition models provide a good analytical solution where repeated measures on each subject are viewed as samples from a Markov chain with transition probabilities: $P(I,j) = prob(Y(t)=j|Y(t-1)=i)$, where I and j equals to 1 or 0 indicating recent "use" or "no-use," Y(t) and Y(t-1) represent two measures collected on time t and t-1 (t=1, ., T). This modeling strategy brings convenience to study dynamic changes of binary variable through time, which simply involves a modified logistic regression model. By estimating transition probabilities for each treatment condition, we can compare their treatment effects. We propose a method to extend the above Markov transition models.

A Method for Imputing Missing Correlated Binary Responses

◆ Sandra Hall, University of Kansas Medical Center; Jasjit Ahluwalia, University of Kansas Medical Center; Matthew Mayo, University of Kansas Medical Center

University of Kansas Medical Center, 3901 Rainbow Blvd. MS1008, Kansas City, KS 66160

shall@kumc.edu

Key Words: *longitudinal, generalized linear mixed model*

Missing data is a common problem in medical research. Many methods have been used to handle missing data at a single time point. However, a number of studies in medical research are longitudinal, resulting in correlated outcomes at numerous time points. We are imputing missing correlated binary data arising from a longitudinal study. We propose a method based on generalized linear mixed model to impute missing data and compare it to a traditional imputation strategy using commercially available software. These two methods are compared and contrasted, and recommendations are made. Each method is then applied to a dataset from a smoking cessation intervention study. The response was a binary correlated outcome, in that it is the repeated measurement of whether the subject had quit smoking or not over time.

Mixed Effects Models for Truncated Longitudinal Outcome and Nonignorable Missing Data

◆ Sujuan Gao, Indiana University School of Medicine

Indiana University School of Medicine, Division of Biostatistics, 1050 Wishard Blvd., RG4101, Indianapolis, IN 46202

sgao@iupui.edu

Key Words: *mixed effect models, truncation, nonignorable missing, EM algorithm*

Mixed effects models are often used for estimating fixed effects and variance components in continuous longitudinal outcomes. Hughes proposed an EM based estimation approach for mixed effects models when the outcomes are truncated. We consider the situation when the longitudinal outcomes are also subject to nonignorable missing in addition to truncation. A selection model approach is proposed to model the joint likelihood of observed data by modeling the probability of missing data to depend on the unobserved outcomes. Sensitivity analyses will be conducted to examine the effect of the missing data assumption on the parameter estimates. Data from neuropsychological testing of dementia will be used to illustrate the proposed approach.

274 Microarray

Biopharmaceutical Section

Tuesday, August 10, 2:00 pm–3:50 pm

At What Scale Should Microarray Data Be Analyzed?

◆ Shuguang Huang, Eli Lilly and Company; Kerry G. Bemis, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

huang_shuguang@lilly.com

Key Words: *microarray, t-test, robustness, power, transformation, normality*

The motivation for transforming microarray data in one way or another is usually to satisfy the model assumptions such as normality and/or homoscedasticity. Generally, two types of strategies are often applied to microarray data depending on the analysis need. One strategy, such as correlation analysis, considers all the gene intensities on the array simultaneously; the other, such as gene-by-gene ANOVA, analyzes each gene individually. We investigate the distributional properties of the Affymetrix genechip signal data under the two scenarios, with the focus on the impact of analyzing the data at an inappropriate scale. When all the genes on the array are considered together (pooled), the dependent relationship between the expression and its variation level can be satisfactorily removed by Box-Cox-type of transformation. When genes are analyzed individually, the distributional properties of the intensities are shown to be gene-dependent. Derivation and simulation show that some loss of power is incurred when a wrong scale is used, but due to the robustness of the t-test, the loss is rather acceptable when the fold-change is not very big.

Issues in the Application of Microarray Data in a Clinical Trial

◆ Bret J. Musser, Merck & Co., Inc.

Merck & Co., Inc., 126 E Lincoln Ave, RY34-A316, Rahway, NJ 07065

bret_musser@merck.com

Key Words: *microarray, clinical trial, crossover, variance components*

The clinical application of microarray data has typically been in an observational setting. This presentation will discuss uses of microarray data in a phase I clinical trial. There are a number of specific issues that arise in a clinical trial setting for the use of microarray data. In a clinical trial, there are limits to the types and quantities of samples that can be taken from healthy volunteers. Invasive tissue sampling limits the number of volunteers and samples per volunteer, thus careful attention must be paid to the sources of error, including errors in sample collection and handling, and the typical inter- and intra-subject variation. The impact of these sources of variability on the power of the study will be discussed. I will present an overview of a proposed experiment and then focus on the design of the sampling scheme for the study. Real and/or simulated data are shown to justify the proposed design.

FDR-based Methods for Determining Replicate Size For Microarray Experiments

◆ Vishwanath Iyer, Bristol-Myers Squibb Company; Sanat K. Sarkar, Temple University

Bristol-Myers Squibb Company, 5 Research Pkwy., Wallingford, CT 06492

vishwanath_i@yahoo.com

Key Words: *replicate size, FDR, microarray, FNR, sample size*

One issue that has received considerable attention recently has been the determination of the number of DNA array replicates that one needs to use to be "reasonably certain" that the differences one observes between control and treatments are "real." The challenge has been to strike a balance between the cost of expensive replicates and some measure of statistical judiciousness. The exploratory nature of these experiments allows us the leeway of considering methods based on the less strict criterion of controlling the False Discovery Rate (FDR) instead of the more conservative Familywise Error Rate (FWER). Previously, we derived an exact formula for FDR and its conjugate quantity, the False Negative Rate (FNR) and described a method to select a replicate size based on these two criteria under certain stringent assumptions. We compared different multiplicity adjustment methods, and their effect on replicate size. We propose an approximation to the exact formula that significantly decreases computation time. We explore relaxing the normality assumptions from the previous paper.

Bayesian Average FDR and FNR in Single-step Procedure with Application to Microarray Experiments

◆ Jie Chen, Merck & Co., Inc.; Sanat K. Sarkar, Temple University

Merck & Co., Inc., 770 Sumneytown Pike, PO Box 4, WP37T2-2, West Point, PA 19486

jie_chen@merck.com

Key Words: *average FDR, average FNR, average Bayes error rate, hierarchical mixture priors, microarray experiments, gene expression*

The original definition of the FDR can be understood as the frequentist risk of false rejections conditional on the unknown parameter, while the Bayesian posterior FDR is conditioned on the data, a particular realization of an experiment. From a Bayesian point of view, it seems natural to take into account the uncertainty in both the parameter and the data. We propose the average FDR (AFDR) and average FNR (AFNR) approaches in which the frequentist risks of false rejections and false nonrejections are averaged out with respect to some prior distribution of parameter. A linear combination of the AFDR and AFNR, called the Average Bayes Error Rate (ABER), is considered as an overall risk. Some useful formulas for the AFDR, AFNR and ABER are developed for normal samples with hierarchical mixture priors. The idea of finding threshold values minimizing the ABER is illustrated using a gene expression data and a simulation study.

Standardization, False Positives, and False Negatives in Microarray Gene Expression

◆ Borko D. Jovanovic, Northwestern University Medical School; Irene B. Helenowski, Northwestern University; Marlos A.G. Viana, University of Illinois, Chicago; Alfred W. Rademaker, Northwestern University Medical School; Raymond C. Bergan, Northwestern University Medical School

Northwestern University Medical School, 680 N. Lake Shore Dr., suite 1102, Chicago, IL 60611

borko@northwestern.edu

Key Words: *Monte Carlo, simulations, log-normal, outliers*

We compare error rates stemming from using ratio (1/k)< R=Y/X< k in gene screening for several commonly used methods for standardizing gene expression in microarrays. The underlying distribution of gene expression is assummed to be log-normal, with mean and SD taken from a published study. Computer simulations are run in S-PLUS and R.

Integrative Analysis of Genomics, Proteomics, Metabonomics, and Blood Chemistry Data

◆ Raymond L. Lam, GlaxoSmithKline; Lei Zhu, GlaxoSmithKline; Amit Bhattacharyya, GlaxoSmithKline; Kwan R. Lee, GlaxoSmithKline; Alan Menius, GlaxoSmithKline ; Paul McAllister, GlaxoSmithKline

GlaxoSmithKline, 709 Swedeland Rd., King of Prussia, PA 19406

raymond.l.lam@gsk.com

Key Words: *data integration, integrative analysis, metabonomics, proteomics, systems biology, transcriptomics*

Systems biology is an integrated approach for studying complex biological systems. It involves combining and analysing data from multiple experimental platforms including blood chemistry, mRNA, proteins, metabolites collected from both preclinical and clinical samples. Recent technological advances make it possible to measure biological activity at the molecular level enabling the generation of large volumes of data representing gene expression, protein, and metabolite abundance. Identifying biomarkers specific and common to different drugs and diseases and understanding the association of these biomarkers will result in a better understanding of drug efficacy and drug safety. There are several challenges in integrative analysis of data from several platforms. The number of measurements (variables) is much larger than the number of subjects. Wide differences in the number of measurements per platform provides a challenge for balancing the platform contributions to the analysis. Issues of multicolinearity and multiple testing abound. We present statistical issues and results related to integration of various platforms.

Empirical Bayesian Analysis of Variance Component Models for Microarray Data

◆ Sheng Feng, North Carolina State University; Russell D. Wolfinger, SAS Institute Inc.; Tzi-Ming Chu, SAS Institute Inc.; Greg Gibson, North Carolina State University; Lisa McGraw, Cornell University

North Carolina State University, Dept. of Statistics, Raleigh, NC 27607

sfeng@unity.ncsu.edu

Key Words: *empirical Bayes, variance component model, mixed model, microarray data analysis, shrinkage estimators*

A gene by gene mixed model analysis (called “single gene analysis”) is a useful statistical method for assessing significance for microarray gene differential expression. While a large amount of data are collected in a typical microarray experiment, the sample size for each gene is usually relatively small, which could limit the statistical power of this analysis. We introduce an empirical Bayesian approach for general variance component models applied to microarray data. The power lost because of small sample size is regained by integrating prior information on variance components estimated from all genes. The approach starts with the single gene analysis. The estimated variance components from each gene are transformed to the “ANOVA components.” This transformation makes it possible that the prior density of each “ANOVA component” is estimated independently. For every gene, the posterior density of each “ANOVA component” can be easily derived based on previous work. The modes of the posterior distribution are inversely transformed to compute the posterior estimates of the variance components.

275 Sampling and Simulation ▲ ⊞

Business and Economics Statistics Section, Section on Survey Research Methods

Tuesday, August 10, 2:00 pm–3:50 pm

A Better Estimate of the Number of Valid Signatures on a Petition

◆ Mary M. Whiteside, University of Texas, Arlington; Mark E. Eakin, University of Texas, Arlington

University of Texas, Arlington, Box 19437, Dept. of Information Systems and Ops. Management, Arlington, TX 76019

whiteside@uta.edu

Key Words: *sampling, signatures, petition, Goodman, nonlinear, replicates*

The 2003 election of Governor Arnold Schwarznegger in California, the current attempt to recall Dallas Mayor Laura Miller, and the 2002 Austin City Council elections illustrate the need for a reliable sampling method for estimating the number of valid signatures on a petition. Signatures may be invalid for several reasons: not a registered voter, incorrect address, replicate, etc. The problem is interesting because replicated signatures must be estimated differently than signatures invalid for other reasons. Goodman’s seminal work and subsequent Goodman-type estimators first estimate the total number of valid signatures on a petition, ignoring initially whether or not they are replicated. Then the problem reduces to estimating the number of classes in a finite population. We first estimate the number of unique, original signatures, and then the proportion that are valid. The result is a nonbiased nonlinear estimator with smaller variance than the

Goodman-type statistics for the case where the proportion of duplicated signatures is the same for valid and invalid signatures.

Minimum Sample Sizes with Rare Events in Stratified Designs

Eric Falk, Ernst & Young LLP; ◆ Joo M. Kim, Ernst & Young LLP

Ernst & Young LLP, 1225 Connecticut Ave., NW, Washington, DC 20036

joomi.kim@ey.com

Key Words: *stratified sampling, rare events, ratio estimation, simulation, normality*

Sampling in a rare event setting with stratified designs has many statistical issues: empty strata, expectation of finding enough of the rare event in each stratum, accuracy of estimates, accuracy of variance calculations, and normality assumptions for confidence intervals. Sampling sales and use tax data involves such a setting, where a large population is sampled to find a rare event (taxable amount) and a small ratio is calculated. Most states require minimum stratum sample sizes of anywhere between 250 to 300 records. We study when this minimum requirement is adequate, insufficient, or in excess of the number necessary to obtain reasonable estimates.

The Accuracy of Multiyear Estimates vs. Annual Estimates for R&D

Wendy Rotz, Ernst & Young LLP; ◆ Jinhee Yang, Ernst & Young LLP

Ernst & Young LLP, 1225 Connecticut Ave. NW, Washington, DC 20036

jinhee.yang@ey.com

Key Words: *sampling, tax, simulation, business*

Sampling is used to estimate a corporation's tax credit for qualifying research and development (R&D) expenditures. Studies encompass multiple year periods, but tax is reported on an annual basis and qualification for the credit is based on annual qualifying expenditures exceeding a base amount. When sampling multiple years, the estimated overall qualifying percentage is applied to each year's expenditures to obtain annual qualifying amounts. However, an IRS concern is that a year with qualifying expenditures below the base amount may be allowed an R&D credit because that year's low expenditures are masked by other highly qualifying years. The effect of a low year is unclear because a low year would also lower the overall qualifying percentage for the time frame, causing less credit in highly qualifying years, which may balance credit erroneously taken in a low year. This paper studies estimation of the R&D tax credit in multiple year studies with variable annual qualifying percentages. We explore whether an overall estimated credit for a multiyear time period can be reasonably estimated from an overall qualifying percentage, or whether separate annual estimates are necessary.

A Generalized Maximum Entropy Approach to Estimating Missing Categorical Dependent Variable Values

◆ Lawrence C. Marsh, University of Notre Dame; Amos Golan, American University

University of Notre Dame, Dept. of Economics and Econometrics, 436 Flanner Hall, Notre Dame, IN 46556-5611

Lawrence.C.Marsh.1@nd.edu

Key Words: *incomplete discrete response, biased sample selection, logistic/logit models, eigenvalues and eigenvectors, singular value decomposition, missing-at-random nonresponse*

Recent research using maximum likelihood methods in estimating missing categorical dependent variable values has produced promising but highly unstable results. Entropy and informational econometric methods offer an approach to the problem that is specifically designed to deal with ill-posed inverse problems to produce more stable and reliable estimates. The research literature on missing categorical response models generally assumes that the responses are missing at random after controlling for all relevant observable data. Our method corrects for systematic bias that may remain after controlling for observables. With relatively mild assumptions we are able to restore the full joint probabilities for both the complete and the incomplete responses.

A Comparison of Random Number Generators Used in Business—II

◆ Archana Joshee, Ernst & Young LLP; Wendy Rotz, Ernst & Young LLP; Eric Falk, Ernst & Young LLP

Ernst & Young LLP, 1225 Connecticut Ave. NW, Washington, DC 20036

archana.joshee@ey.com

Key Words: *DIEHARD, NIST Statistical Test Suite, Random Number Generator, Excel, S-Plus, SAS*

This paper is a continuation of a study done by Wendy Rotz et al. presented in the 2001 JSM on the Comparison of Random Number Generators Used in Business. Microsoft revised Excel's Random Number Generator based on the research done in this 2001 study. How does Excel's new Random Number Generator perform compared to the version tested earlier? Does the RAND function in combination with the STREAMINIT function in SAS 8.2 work better than the RANUNI function? The DIEHARD battery of tests as well as the tests from the National Institute of Standards and Technology (NIST) are used to assess random number generators used in business.

Broadening the Scope of the Bootstrap in Complex Problems

◆ Manuel A. Dominguez, ITAM; Victor Aguirre, ITAM

ITAM, Rio Hondo 1, Tizapan - San Angel, Mexico City, 01000 Mexico

madt@itam.mx

Key Words: *resampling, Monte-Carlo-based inference, percentile bootstrap method, pivotal statistics, t-ratio*

The avaliability of computers has allowed the increase of complexity of models in statistics reach levels never seen before. Computer intensive methods and estimation methods based on simulations to estimate the coefficients of stochastic differential equations are two examples just to name a few. Bootstrap methods are becoming increasingly familiar in statistics. Since these methods are applied by means of Monte Carlo experiments, the computational cost involved in its applicaton to methods like the ones mentioned before could become prohibitive. Therefore, results for reducing this computational cost should be welcome. We propose an approach to the bootstrap that allows us to apply accurate inferential procedures with very small Monte Carlo experiments. Also, the use of this approach permits the determination of the smallest number of bootstrap replications needed to achieve a desired accuracy level. The approach is very general and it applies to the iid or dependent situation. The assumptions required are those that ensure convergence in distribution of bootstrap, there is no need of extra assumptions to insure convergence of moments.

Estimating Variability in Biodiversity—Ecosystem Function Experiments with Resampling Methods

◆ Kevin Gross, North Carolina State University; Bradley J. Cardinale, University of Wisconsin, Madison

North Carolina State University, Dept. of Statistics, CB 8203, Raleigh, NC 27608

gross@stat.ncsu.edu

Key Words: *biodiversity, ecology, bootstrap*

In ecology, several high-profile experiments have been conducted recently to elucidate the effects of biodiversity loss on ecosystem function. In these experiments, communities are assembled by random selection of species from an experimental species pool, and relationships between the number of species in a community and ecosystem function are sought. From a statistical perspective, shared dependence of the experimental communities on the species pool induces strong correlations among community responses. To generalize the results of these experiments, it is necessary to understand how this correlation structure affects the sampling variability in the observed biodiversity—ecosystem function patterns. Here, we propose three resampling methods (bootstrap, jackknife, and infinitesimal jackknife) to estimate this sampling variability, and study their performance in simulations. Our results suggest that these basic re-sampling methods tend to overestimate sampling variability, and that opportunity remains for the development of a superior resampling method.

276 Data and Dimension Reduction ⌘

General Methodology

Tuesday, August 10, 2:00 pm–3:50 pm

Sufficient Dimension Reduction via Hellinger Integral

◆ Xiangrong Yin, University of Georgia

University of Georgia, 204 Statistics Bldg., Dept. of Statistics, Athens, GA 30602

xryin@stat.uga.edu

Key Words: *dimension reduction subspace, central subspace, regression graphics*

Sufficient dimension reduction has been widely used in statistics, particularly in regression. We propose a general sufficient dimension reduction method that uses Hellinger integral that makes no assumption on the model or the distribution of the predictors. While this method preserves the population properties for the log likelihood methods by Yin and Cook (2003), it may simplify and improve the computations. Consistent as well as asymptotic normality of the estimator is derived. Illustrative examples are presented.

Robust Dimension Reduction Based on CANCOR

◆ Jianhui Zhou, University of Illinois, Urbana-Champaign

University of Illinois, Urbana-Champaign, Dept. of Statistics, 101 Illini Hall, 725 S. Wright St., Champaign, IL 61820

zhou@uiuc.edu

Key Words: *canonical correlation, dimension reduction, MCD estimator, permutation test, robustness*

The canonical correlation (CANCOR) method for dimension reduction in a regression setting is based on the first and second moments of the data, and therefore sensitive to outliers. We study a weighted canonical correlation (WCANCOR) method where each observation is weighted based on its Mahalanobis distance to the center of the predictors. Robust estimates of the location and scatter such as the minimum covariance determinant (MCD) estimator of Rousseeuw (1984) can be used to compute the Mahalanobis distance. To determine the number of significant dimensions in the WCANCOR method, we consider a weighted permutation test. A comparison between the weighted CANCOR and the weighted SIR methods is also made to see the relative merits of each method.

Nonlinear Manifold Learning

◆ Alan J. Izenman, Temple University

Temple University, Dept. of Statistics, Speakman Hall (006-00), 1810 N. 13th St., Philadelphia, PA 19122-6083

alan@temple.edu

Key Words: *multivariate, high dimensionality, PCA, machine-learning, nonlinearity, kernels*

Classical multivariate analysis has focused mainly on linear methods for dimensionality reduction. Examples include principal components analysis, canonical variate analysis, and discriminant analysis. New techniques are now being introduced into the statistical literature and the machine-learning literature that attempt to generalize these linear methods for dimensionality reduction to nonlinear methods. One of the main aspects of such generalizations is that there are different strategies for viewing the presence of nonlinearity in high-dimensional space. Thus, we have recently seen a number of versions of "nonlinear" PCA, including polynomial PCA, principal curves and surfaces, auto-associative multilayer neural networks, smallest additive principal components, and kernel PCA. Other techniques for finding nonlinear structure in data are Isomap, local linear embedding, and Laplacian eigenmaps. We describe several of these methods and their connections to each other.

Comparison of Orthogonal and Oblique Factor Analysis Rotations in Factor Recovery

◆ Holmes Finch, Ball State University

Ball State University, Dept. of Educational Psychology, Muncie, IN 47306

whfinch@bsu.edu

Key Words: *factor analysis, orthogonal rotation, oblique rotation, factor recovery*

Factor analysis (FA) is a popular methodological approach for reducing dimensionality in a set of data and understanding latent factors expressed in manifest variables. Interpretation of FA solutions involves the use of factor loadings, which indicate the nature of the relationship between observed variables and the latent traits that they ostensibly measure. To make these factors more interpretable, the loadings are transformed using a process known as rotation. There are two general approaches to factor loading rotation, orthogonal and oblique, with the latter allowing for correlated factors and the former not. Practitioners will find a variety of recommendations as to which type of rotation to use when. The goal of the current study is to compare the performance of oblique and orthogonal rotations in terms of factor recovery, with principal axis factoring. Previous research with nonlinear factor analysis indicates that they might perform very similarly, regardless of the between factors correlation. A Monte Carlo simulation is used, manipulating the number of variables, sample size, level of correlation among factors, and the number of factors.

A New Approach to Data Dimension Reduction and Variable Selection

◆ Shiying Wu, RTI International; Jun Liu, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

swu@rti.org

Key Words: *principal components, statistical testing*

Variable selection and data dimension reduction is a commonly encounter problem in many scientific disciplines. When analyzing the full scan chromatograms from breath samples of asthmatic children, we applied an innovative approach to identify dozens VOCs, among hundreds of potential targets, that might be triggers of asthma attack. In our new approach, the independent variable was regressed on the rest of the principle components after the least important principal components were first being removed. The regression equation in principal components was translated back to the equation in the original variables. The approach takes the advantage of PCA and provides results that are directly interpretable. We attempt to provide a formal statistical inference framework within which formal statistical testing can be performed for variable selection. We will also attempt to address the multiple test problem encountered in this type of problems.

Detection of Modal Clusters in High-dimensional Data

◆ Surajit Ray, University of North Carolina, Chapel Hill; Bruce G. Lindsay, Pennsylvania State University

University of North Carolina, Chapel Hill, Biostatistics, 3125 MCGV Bldg., CB# 7420, Chapel Hill, NC 27599

surajit@unc.edu

Key Words: *modes, bimodal, dimension reduction, modal cluster, mixture*

Fitting a mixture model offers a primary data reduction through the number, location, and shape of its components, but in more complex settings we would like to know more about how the components interact to describe an overall pattern of density. In terms of subgroup structure for a two component mixture, we might create two competing hypotheses: either there are a homogeneous groups with normal shape, or one aggregated group having a more complex shape. We are interested determining how the aggregation of mixture components defines a modal cluster. This paper develops new tools for finding modal clusters, ones that are useful especially in high dimensions. For a unimodal density, concepts like skewness and kurtosis, heavy and light tails are used to describe shape. When the density is multimodal however, emphasis usually shifts to the number and location of modes. This is because the modes are often symptomatic of underlying population structures. Our investigation into the topography of a mixture of K=2 components D-variate normals yielded several interesting results.

A Forward Approach to Estimating Effective Dimension Reduction Space in High-dimensional Regression

◆ Peng Zeng, Purdue University; Yu Zhu, Purdue University

Purdue University, Statistics Dept., 150 N. University St., W. Lafayette, IN 47907

pzeng@stat.purdue.edu

Key Words: *dimension reduction, e.d.r subspace, SIR, SAVE, regression*

Dimension reduction is crucial to the success of high dimensional regression. Suppose Y is the response and x is the d-dimensional input vector. It is reasonable to assume that Y depends on x through k linear combinations, i.e., Y=f(b_1 x,.,b_k x,e), where e is the random error and f the unknown link function. The space spanned by b_1, b_2, ., b_k is called the effective dimension reduction (e.d.r) space. Two approaches exist to recover the e.d.r space, namely the forward approach and the inverse approach. The former involves the fitting of f, thus can become infeasible even when d is moderate. The latter has led to several methods such as SIR and SAVE, which successfully avoid the fitting of f. Due to their inverse fashion, some disadvantages exist. We propose a novel method based on the forward approach. Our method can also avoid the fitting of f (thus it is link-free) and the curse of dimensionality, and is more transparent than SIR and SAVE. Theoretical results have been established under the model assumption. Simulation results using synthetic and real data will be reported to compare our method with SIR and SAVE.

277 After the Survey: Analysis and Public Use of Government Surveys ⟁

Social Statistics Section, Section on Survey Research Methods, Section on Government Statistics

Tuesday, August 10, 2:00 pm–3:50 pm

Immigrant-Nonimmigrant Differences in Canadian Healthcare Utilization

◆ Evelyn Park, Statistics Canada; Monica Boyd, Statistics Canada

Statistics Canada, Main Bldg., Room 1710, Ottawa, ON K1A 0T6 Canada

evelyn.park@statcan.ca

Key Words: *immigrants, health care utilization, enabling factors, health factors, health survey*

We study immigrant-nonimmigrant differences in three types of health care utilization: medical (physicians and nurses); dentists; and other health care practitioners. Using 118,000 respondents from the Canadian Community Health Survey (CCHS 2000/01), our model includes age, nativity including duration in Canada, and enabling and health-related factors. We estimate the odds of accessing each type of health care services, and the frequency of consultations for health service users, using logistic regression, general linear models, and bootstrapping techniques. Consistent with the "healthy immigrant effect" and a nationally funded medical system, immigrants are similar to the Canadian-born in their utilization of medical services and in frequency of use. However, they are less likely to use dental services, which are not nationally funded, and they use other types of care less frequently at the beginning of their stay and more frequently after 20 years residency. Enabling and health-related factors affect health care utilization, but their inclusion only marginally alters immigrant-to-Canadian-born utilization of health care services.

Using National Health Interview Survey Data to Detect Signals of Natural or Intentional Outbreaks

◆ Abera Wouhib, Centers for Disease Control and Prevention; Paul D. Williams, Centers for Disease Control and Prevention; Joe Fred Gonzalez, Jr., Centers for Disease Control and Prevention; Myron Katzoff, National Center for Health Statistics

Centers for Disease Control and Prevention, NCHS, 3311 Toledo Rd. Room # 3124, Hyattsville, MD 20782

awouhib@cdc.gov

Key Words: *symptomatic data, bioterrorism*

The National Health Interview Survey (NHIS) is a household survey which has been conducted by the National Center for Health Statistics (NCHS) for four decades. As part of the NHIS, information on head or chest colds, stomach or intestinal illnesses with vomiting and diarrhea are collected. These symptoms are commonly associated with colds and flu, are emerging diseases. However, these may also be initial symptoms for victims of bioterrorism. Since 1997, the NCHS has been collecting data on these common symptoms of illness using a two-week reference period for sampled adults and sampled children in the NHIS. Prevalence estimates for these data could possibly be used as a baseline to detect natural or intentional outbreaks. National Prevalence estimates were computed for each two-week period for sampled adults and children. By pooling the data collected over a five-year period, prevalence estimates were calculated by geographic region for all persons and by gender. Also, sample sizes that are necessary to produce reliable estimates with a specified level of precision for each two-week period were computed.

Enhanced Outreach Activities Involving Data from the National Center for Health Statistics

Robert Weinzimer, National Center for Health Statistics; ◆ Mary Moien, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 7115, Hyattsville, MD 20782

MMoien@cdc.gov

Key Words: *data dissemination, government statistics, NCHS*

The National Center for Health Statistics, CDC, has a long reputation for disseminating timely data from its surveys to a broad spectrum of users. Data are disseminated in many forms, including reports, journal articles, press releases, tabulations, and microdata. Special data analyses are now found on the NCHS web site. Recently, an Outreach Team was established to investigate additional ways to generate use and awareness of NCHS data. The activities of the team will be highlighted. Projects included, among others to be discussed, an electronic database of organizations; building a simple desktop publishing capability for on-demand production of fact sheets and an electronic newsletter; and conducting special briefings, visits, and workshops to highlight data from NCHS.

Growth in Mathematics Achievement during Middle and High School: Analysis With Classification and Regression Trees

◆ Xin Ma, University of Kentucky

University of Kentucky, 335 Dickey Hall, Lexington, KY 40506-0017

xin.ma@uky.edu

Key Words: *classification and regression tree, data analysis, mathematics achievement, measure of change*

Using a recently developed statistical technique often referred to as classification and regression trees (CART), the present study classified students into groups with differential rates of growth in mathematics achievement during middle and high school based on their individual and family variables and characterized students who grew fast and slow in mathematics achievement. Data analysis of the Longitudinal Study of American Youth (LSAY) showed that age (being younger in the same grade cohort) was critically important for fast growth in mathematics achievement. While younger white and Asian students grew at the best rate in mathematics achievement, older white and Asian students with low parental socioeconomic status (SES) grew at the worst rate. Hispanic, Black, and other students were sandwiched in between. One in three socially disadvantaged students in the sample overcame the negative impacts of low family SES and large family size to grow at the second best rates in mathematics achievement, while one in seven socially advantaged students in the sample failed to take advantage of high family SES and small family size. Males and females shared the fast end of growth in mathematics.

Do NCTM Reform Practices Matter? Predicting Precollege Student Mathematics Achievement Using NAEP 2000 Data

◆ Mack C. Shelley, II, Iowa State University; Sarah T. Lubienski, Iowa State University

Iowa State University, Dept. of Statistics, 323 Snedecor Hall, Ames, IA 50011-1210

mshelley@iastate.edu

Key Words: *mathematics education, NAEP data, HLM models, socioeconomic status, gender, race*

Two-level hierarchical linear models are used to estimate the impact of National Council of Teachers of Mathematics reform instructional practices for students nested within schools, for five student mathematics achievement strands (number/operations, geometry, measurement, data analysis, and algebra/functions), controlling for socioeconomic status (SES), race, and gender at both levels, using National Assessment of Educational Progress data adjusted for “plausible values” through model-based imputation and item response theory. Initial level 1 results show that higher student SES leads to higher mean math achievement scores, White and Asian/Pacific Islander students on average have higher mean math achievement scores than other students, and female students have lower mean math achievement scores than male students. At level 2, math achievement is enhanced by higher school SES; the effect of school race is more ambiguous. No reform teaching activity is significant. This study can help policymakers and educators create equitable classrooms, which ultimately will enhance life chances for minority, low-SES, and female students entering higher education and employment.

The Relationship between Wages and Alcohol Dependence

◆ Richard R. Bryant, University of Missouri, Rolla; V. A. Samaranayake, University of Missouri-Rolla

University of Missouri, Rolla, Economics and Finance Dept., G-9 Harris Hall, Rolla, MO 65401

rrbryant@umr.edu

Key Words: *wages, alcohol dependence, NLSY, longitudinal model*

The relationship between an individual’s wage and a measure of alcohol dependence is estimated using data from the National Longitudinal Survey of Youth. Previous research on wages, hours worked, and employment, using this data, has focused on the relationship between these productivity measures and frequency and quantity of alcohol use variables. An emerging contention among some researchers is that while alcohol consumption measures may not affect labor productivity, measures such alcohol dependency that reflect problematic drinking do. A time-series cross-sectional model is used to estimate the relationship between hourly wages of individuals reported in two time periods and standard wage-determining characteristics as well as measures of current and past alcohol dependence constructed in the spirit of criteria used in the Diagnostic and Statistical Manual of Mental Disorders, third and fourth edition. Also, the relationship between wages and other indicators shown in the literature to be related to alcohol dependence or abuse, such as early onset of drinking, family history of alcoholism, and “binge” drinking is investigated.

278 Sample Designs △

Section on Survey Research Methods

Tuesday, August 10, 2:00 pm–3:50 pm

Optimal Sample Allocation for Design-consistent Regression

◆ Hui Zheng, Harvard Medical School; Alan Zaslavsky, Harvard Medical School; John L. Adams, RAND Corporation

Harvard Medical School, 180 Longwood Ave., Dept. of Health Care Policy, Boston, MA 02115

zheng@hcp.med.harvard.edu

Key Words: *descriptive population quantity, survey, measurement error*

We consider optimal sampling rates when the anticipated analysis is survey-weighted linear regression and the estimands of interest are combinations of regression coefficients from one or more models. Methods are first developed assuming that exact design information is available in the sampling frame and then generalized to situations in which some design variables are available only as aggregates for groups of potential subjects, or from inaccurate or old data. We also consider design for estimation of combinations of coefficients from more than one model. A further generalization allows for flexible combinations of coefficients chosen to improve estimation of one effect while controlling for another. We illustrate the potential gains from using these methods with simulated continuous variables. Potential applications include estimation of means for several sets of overlapping domains, or improving estimates for subpopulations such as minority races by disproportionate sampling of geographic areas.

Optimal Stratification Based on Nonparametric Regression Methods

◆ Enrico Fabrizi, Università di Bologna; Carlo Trivisano, Università di Bologna

Università di Bologna, Via Belle Arti 41, Bologna, 40126 Italy
fabrizi@stat.unibo.it

Key Words: *optimal stratification, nonparametric regression, finite population*

In the design of surveys, stratification is introduced for a number of different purposes, among which the gain in efficiency for design based estimators of population parameters. A stratification providing the most efficient stratified mean estimator is therefore called "optimal." A landmark in the literature about optimal stratification is the Dalenius method. We propose to generalize and extend the Dalenius method to situations in which the target variable (or a proxy) is available for a sample and a large set of covariates is known for each unit in the population. This situation may arise in repeated surveys or when stratifying first-level units in multistage designs. We propose to replace the linear regression which form the kernel of the Dalenius method with more flexible non parametric regression techniques such as MARS, generalized additive models, and regression trees. All evaluations and comparisons are carried out with reference to a large simulation exercise in which different hypotheses on the target variable, the covariates, and the relation between them are considered. Some real-life example is also considered.

A Study of an Optimization Method Used for the Planning of a Complex Multivariate Survey

◆ Anders Holmberg, Statistics Sweden

Statistics Sweden, Dept. of Research & Methodology, Örebro, SE-70189 Sweden

Anders.Holmberg@scb.se

Key Words: *auxiliary information, nonlinear optimization, probability proportional-to-size sampling, sample allocation*

A very important part in the planning phase of a survey is the choice of sampling design. We present preliminary studies of a method that aims to find an optimal sampling design when the survey has a multivariate purpose and when the survey population is stratified. To solve a classical planning problem, i.e., finding a survey strategy where the precision of the estimators is as high as possible for a given cost, a nonlinear optimization problem is formulated. The optimization problem is to determine element inclusion probabilities and the stratum sample allocation, which minimize a function of estimator variances under certain restrictions of precision. We illustrate the method and study its behavior when the restrictions are made on several variables, across or within strata.

Sample Selection by Powers of Size when Needing Estimates at Multiple Levels

◆ Pedro J. Saavedra, ORC Macro International Inc.

ORC Macro International Inc., 11785 Beltsville Dr., Calverton, MD 20705

Pedro.J.Saavedra@orcmacro.com

Key Words: *sampling, PPS, institutions, variance, simulations, establishments*

Institution and establishment surveys often require two kinds of estimates: one may be the percentage of units meeting a particular characteristic, and the other may be the percentage of clients served or volume sold associated with units meeting the characteristic. For example, a survey of schools may need to estimate the percent of schools that implement a given program, and the percentage of students attending such schools. For the first estimate, one should assign the same probability of selection to every school, whereas for the second one should sample with probabilities proportional to the enrollment. When both are needed, some researchers have sampled with probabilities proportional to the square root of enrollment. The question is what power to use to minimize the larger of the two coefficients of variation. Using a national database of schools, five powers of the enrollment were used and estimates were obtained for several simulated and actual categorical variables varying in magnitude and in their correlation to the enrollment. The coefficients of variation were obtained through multiple simulations and the maximum of the two CVs compared.

Obtaining Stratum Breaks in Skewed Populations Using a Simple Method

◆ Patricia M. Gunning, Dublin City University; Jane M. Horgan, Dublin City University; Gary Keogh, Dublin City University

Dublin City University, School of Computing, Dublin 9, Ireland

pgunning@computing.dcu.ie

Key Words: *efficiency, geometric progression, optimum allocation, stratification*

Dalenius (1950) derived equations for determining boundaries when stratifying continuous variables so as to minimize the variance of the estimates with optimum allocation. No exact solution of these equations exists, and implementation of optimum stratification is usually based on approximation methods. This paper derives a new algorithm for the construction of stratum boundaries in positively skewed populations, which is easier to implement than current methods. The algorithm is based on an observation by Cochran (1961) that, with near optimum boundaries, the coefficients of variation are often found to be approximately the same in all strata. We show that, when the populations are positively skewed, the stratum breaks may be obtained using the geometric progression. Tests on four real populations show that the suggested method of stratum construction compares favorably with the commonly used cumulative root frequency approximation method of Dalenius and Hodges (1957), and the method of Lavallée and Hidiroglou (1988), which is specifically designed for skewed populations, in terms of the precision of the estimator of the mean and total.

Balanced Sampling Excluding Adjacent Units

◆ James Wright, Bucknell University

Bucknell University, 380 Olin Science, Lewisburg, PA 17837

jwright@bucknell.edu

Key Words: *balanced sampling, circularly ordered populations, finite population sampling, generator blocks*

Hedayat, Rao, and Stufken (1988) first introduced balanced sampling designs for the exclusion of contiguous units. Under the assumption that contiguous units of the population provided similar information, finite one-dimensional populations arranged in a circular order were considered. Sampling plans that excluded the selection of contiguous units within a given sample, while maintaining a constant second-order inclusion probability for noncontiguous units, were investigated. Stufken (1993) extended the general concept to balanced sampling avoiding adjacent units from circularly ordered populations. Under such a plan units that are within a distance of alpha of each other are considered to be adjacent. While the early papers developed the foundation for sampling one-dimensional populations avoiding adjacent units under circularly ordered populations, it was not until recently that significant advancements were made in the direct identification of such plans. Through the work of Wright and Stufken (2004), search algorithms have been developed that have successfully identified balanced sampling plans excluding adjacent units for circularly ordered populations.

Assessing Complex Sample Designs via Design Effect Decompositions

◆ Inho Park, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

inhopark@westat.com

Key Words: *clustering, unequal weighting, stratification, PPS sampling, simulation*

The well-known design effect decomposition model proposed by Kish (1987) is often used to evaluate the efficiency of two features, clustering and unequal weighting. For complex sample designs involving other features such as stratification and PPS sampling of clusters, the model lumps all the features other than unequal weighting under the design effect component due to clustering. To provide separate evaluations of features, we propose a decomposition model extending Park et al. (2003). We also compare the Kish and proposed models and their respective design effect component models through simulation.

279 Government Surveys ▲

Section on Survey Research Methods, Social Statistics Section, Section on Government Statistics

Tuesday, August 10, 2:00 pm–3:50 pm

Canadian Labour Force Survey

◆ Guy Laflamme, Statistics Canada

Statistics Canada, 120 Parkdale, Ottawa, ON K1A 0T6 Canada

guy.laflamme@statcan.ca

Key Words: *area frame, stratification, address register, labour force survey*

Every 10 years, following the decennial census, the sample design of the Canadian Labour Force Survey is revisited. Because of limited resources, limited research was done for this edition. Regardless, many innovations will be implemented. The presentation will focus on two of these: the use of the Address Register to reduce collection costs and improvement made to the stratification strategy.

Evaluation of Unique Aspects of the Sample Design for the National Compensation Survey

◆ Lawrence R. Ernst, Bureau of Labor Statistics; Christopher J. Guciardo, Bureau of Labor Statistics; Yoel Izsak, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160, Washington, DC 20212-0001

ernst.lawrence@bls.gov

Key Words: *rotating panel design, PSUs, PPS sampling*

The National Compensation Survey (NCS), conducted by the Bureau of Labor Statistics, uses a rotating panel design, with three stages of selection used in selecting each annual sample panel, namely: geographic area PSUs, which are only reselected once a decade; establishments selected from industry strata; and occupations selected separately from each sample establishment. PPS sampling is used at each stage. Certain aspects of the NCS design are atypical and perhaps unique to this survey program, two of which are studied in this paper. First, although for most of the frame one fifth of the sample establishments are replaced each year, certainty establishments are selected only once every five years using smaller sampling intervals than used to select the remainder of the sample. The second issue is that many of the PSUs are clustered and the sample establishments selected across the cluster as a whole, rather than in each PSU independently. The paper discusses these aspects of the design, explains the reasons for their use, and evaluates their impact through empirical studies.

Use of Overlap Maximization in the Redesign of the National Compensation Survey

◆ Steven P. Paben, Bureau of Labor Statistics; Lawrence R. Ernst, Bureau of Labor Statistics; Yoel Izsak, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3160, Washington, DC 20212

Paben.Steven@bls.gov

Key Words: *sample redesign, overlap maximization, transportation theory*

Among the key products produced by the National Compensation Survey (NCS), which is conducted by the Bureau of Labor Statistics, are locality wage surveys. Consequently, a sample redesign of the NCS is necessary each decade following the release of new metropolitan area definitions. NCS uses a rotating panel design, with three stages of selection, namely: geographic areas; establishments, and occupations. We apply an overlap maximization procedure derived by Causey, Cox, and Ernst (1985) in the selection of noncertainty metropolitan areas in the NCS redesign. This procedure uses transportation theory to obtain an optimal overlap. Despite the optimality of this procedure, it has rarely, if ever, been used in production because it often produces transportation problems that are impractically large. For our NCS application, however, the transportation problems will be of manageable size. We present the expected overlap results for this procedure and compare them to the results using other overlap procedures, such as Perkins (1970) and Ohlsson (1996). We explain how we handle in the overlap procedure areas whose definition has changed in the new design.

Is My Panel Survey Still Representative? A New Method for Detecting whether Two Subsamples Represent the Same Universe—The German Socio-Economic Panel Study Experience

◆ Jan Goebel, DIW Berlin; Joachim R. Frick, DIW Berlin; Edna Schechtman, Ben Gurion University; Gert G. Wagner, DIW Berlin; Shlomo Yitzhaki, Hebrew University and Central Bureau of Statistics

DIW Berlin, Koenigin-Luise-Strasse 5, Berlin, 14195 Germany

jgoebel@diw.de

Key Words: *panel studies, Gini decomposition*

A most important shortcoming of panel surveys is bias arising from selective attrition. Based on data of the German Socio-Economic Panel Study (SOEP) we analyze potential artifacts (level, structure, inequality of income) by comparing results for two independently drawn panel subsamples started in 1984 and 2000, respectively. Both samples are carried on using the same set of follow-up rules. We apply ANOGI (Analysis Of Gini) techniques, the equivalent of ANOVA (Analysis Of Variance) performed with Gini coefficient. The decomposition followed is presented in Yitzhaki (1994). We rearrange, reinterpret, and use the decomposition in the comparison of subpopulations from which the different samples were drawn. Taking into account indicators for income—and for control purposes those for education and satisfaction as well—significant differences between these two samples with respect to income inequality are found in the first year, which start to fade away in wave two and di appear in wave 3. We find credible indication for these differences to be driven by changes in response behavior of short-term panel members rather than by attrition among members of the longer-running sample.

Challenges in the Design and Implementation of the Adult Education Literacy Study

◆ Thomas R. Krenzke, Westat; Leyla Mohadjer, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

tomkenzke@westat.com

Key Words: *sample design, sampling frame, adult education, literacy assessment*

The Adult Education Literacy (AEL) study, sponsored by the Office of Vocational and Adult Education (OVAE), and the National Center for Education Statistics (NCES), is the first nationally representative sample that included a literacy assessment administered to adults enrolled in adult education programs. The AEL sample design and selection procedures required careful attention to a number of constraints present in such a survey. The study was designed to satisfy the strict requirements for achieving the target sample sizes for English and Spanish assessments, as well the counts of adults in the ethnic groups. The survey required creating a number of sampling frames needed at various stages of sampling. In addition, the design had to take into consideration such issues as: the absence of sampling frame sources; lack of ethnicity data at the program level; small enrollments in some adult education locations; maintaining a probability-based assessment booklet distribution; administering within-class sampling without professional interviewers; varying classification definitions of English as a Second Language, Adult Basic Education, and Adult Secondary Education courses; and more.

Group Quarters Sampling and Estimation Plans for 2005 American Community Survey

◆ Mark E. Asiala, U.S. Census Bureau

U.S. Census Bureau, 2555 Robinson Place, Waldorf, MD 20602-2061

mark.e.asiala@census.gov

Key Words: *group quarters, sampling, estimation, American Community Survey (ACS), GQ*

The American Community Survey (ACS) will include group quarters (GQ) population in its sample along with the household population when it is at full implementation in 2005. This paper will detail the plans on how the sample of GQs for 2005 was selected and will describe the plans for the estimation operations required to process that sample. Some of the more challenging issues regarding weighting the GQ data for ACS will be presented.

280 Assessing Students, Instructors, Techniques, Exams, and Courses

Section on Statistical Education

Tuesday, August 10, 2:00 pm–3:50 pm

A Model for Attitudes Toward Statistics, Reasoning Abilities, and Course Performance

◆ Dirk Tempelaar, University of Maastricht

University of Maastricht, UM, FdEWB, KE, PO Box 616, Faculty of Economics & Business, Maastricht, International 6200 MD The Netherlands

D.Tempelaar@ke.unimaas.nl

Key Words: *educational statistics, assessment, higher education, statistical reasoning, attitudes towards statistics*

The main contribution of this paper is the empirical testing of a structural equation model with the purpose of advancing our knowledge about the link between the outcome of a course in statistics on the one hand, and attitudes and beliefs as well as prior reasoning abilities on the other. To this purpose measurement models for reasoning abilities, using the Statistical Reasoning Assessment instrument, and for attitudes and beliefs, using the Survey of Attitudes Towards Statistics, are developed that are used in a structural equation model in conjunction with constructs combining the results of various quizzes and tests.

Toward Reform in Introductory Statistics—An Unexplored Domain

◆ Rossi A. Hassad, Mercy College

Mercy College, c/o 1980 Unionport Rd., Apt. E 31, Bronx, NY 10462

Rhassad@mercy.edu

Key Words: *statistics, psychology, attitude, teaching, learning, reform*

There is strong theoretical and empirical evidence that instructors' attitudes toward teaching and learning can inhibit adoption and maintenance of "best practices" and lead to learning difficulties which are widespread among students of introductory statistics. Such difficulties persist at the college level amidst reform aimed at promoting statistical literacy, which is fast becoming a core competency for students in the behavioral sciences. Pioneer statistics educators and researchers insist on identifying instructor training and preparation as the primary barrier to achieving desired learning outcomes. Yet there has been no formal and focused attempt by the reform movement to determine what beliefs and attitudes possessed by instructors of introductory statistics may have been engendered by their academic training and preparation, and the possible influence of these characteristics on teaching approaches and learning outcomes. This paper will outline and discuss from a qualitative perspective, a framework (guided by the Theory of Planned Behavior and the Concerns Based Adoption Model) for exploring and measuring instructors' attitudes in the context of innovation.

An Assessment of Computer-based Learning in Teaching Statistics

◆ Paul J. Fields, Brigham Young University; Patti B. Collings, Brigham Young University

Brigham Young University, 222 TMCB, Provo, UT 84602

pjfields@stat.byu.edu

Key Words: *computer-based learning, teaching statistics*

In an initial assessment of the impact of using computer-based learning (CBL) to teach statistics, four of 40 lessons were provided via CBL in an introductory undergraduate statistics course. The students did not attend classes on the days they viewed the computer-based lessons. After each computer-based lesson the students completed an online opinion survey of multiple choice and open-ended questions. Based on examination scores, there was no evidence of a difference in the students' level of mastery of statistics concepts using computer-based lessons compared to classroom-based instruction. However, CBL was overwhelmingly preferred because of the greater control it provided the student in the learning process. The students reported that the greatest benefit of CBL was the flexibility it provided them to fit the lessons into their weekly schedules thereby providing the students the opportunity to better manage their time. It is suggested that better student time management could lead to better overall student learning.

Instructor Moderated Group Work: Does it Work?

◆ William L. Harkness, Pennsylvania State University; Patricia Buchanan, Pennsylvania State University

Pennsylvania State University, Dept. of Statistics, 318 Thomas Bldg., University Park, PA 16802

wlh@stat.psu.edu

Key Words: *group moderation, ill-structured problem-solving, effect of scoring rubrics, online group communication*

An experiment was conducted to investigate the relative effects of two approaches to peer online collaboration (peer controlled and externatally structured and moderated) on the achievements of small groups during well-structured and ill-structured problem solving tasks, in a large (280+ students) introductory statistics class. A randomized post-test only experimental design was used to investigate which of the two approaches is more effective in promoting college students' group achievement for each type of problem. Two distinctly different scoring rubrics were used: one developed by the instructor and the second by an instructional designer. About 35 groups were randomly assigned to each approach. We describe the "moderation" process and compare the outcomes as determined by the two scoring rubrics.

On the Effect of the Ordering of Questions on Exam Scores

◆ Juergen Symanzik, Utah State University; Palyne Gaenir, CyberGnostics, Inc.; Robert Erbacher, Utah State University; Natascha Vukasinovic, Monsanto Animal AG Genomics

Utah State University, Dept. of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322-3900

symanzik@math.usu.edu

Key Words: *teaching, electronic textbook, CyberStats, grades, student performance*

The effect of the ordering of questions in sample surveys has been well investigated. However, there has been no study so far that examined the effect of the ordering of questions on exam scores. We will discuss the effect of the ordering of questions on the exam scores for an introductory long-distance statistics course. Since we were working with the electronic textbook CyberStats, which records exact submission times of student answers to exam and homework questions, additional questions can be answered: Do students usually answer questions in sequential order (independent from the questions) or do they tend to jump from one question to another while leaving some types of (apparently more difficult) questions until the very end? How many times do students change their answers, and how often do they change previously correct answers to incorrect answers? Finally, in which order and when do students work on homework assignments—throughout the week or just before the submission deadline?

Maximizing Learning of Statistics in Distance Courses

◆ Steven M. LaLonde, Rochester Institute of Technology

Rochester Institute of Technology, 98 Lomb Memorial Dr., Rochester, NY 14623-5604

smleqa@rit.edu

Key Words: *distance, online, education, evaluation, graduate, applied*

The Center for Quality and Applied Statistics at the Rochester Institute of Technology has long been known for producing applied statisticians that have an immediate impact on their company's research and manufacturing quality goals. In years past, many of the part-time graduate students were full-time employees at companies such as Kodak and Xerox, but changes in local employment situation have required the Center to look elsewhere for students. The Center has responded by accepting more full-time students and by offering the master's degree program in distance format. Enrollment in distance sections of courses now represents over half of the course hour equivalents. This paper illustrates some of the ways in which distance learning is maximized with sychronous and asynchronous discussions, team projects, courseware, and professionally prepared lectures delivered on tape and CD. Evaluation in the distance format is also addressed with a discussion of take home tests, proctored tests, projects, and grading for participation in discussions.

Different Types of Students in the Italian Universities: Is There a Need for Different Policies and Governance Interventions?

◆ Luigi Biggeri, National University Evaluation Committee; Matilde Bini, University of Florence

National University Evaluation Committee, P.le Kennedy, 20, Rome, 00144 Italy

bini@ds.unifi.it

Key Words: *correspondence analysis, forward search, robustness*

The Italian university system is undergoing some radical changes that will affect almost every aspect of its organization. The changes were motivated by several needs—for example, to lower the extremely high drop out rate, to lower the number of years spent by students to obtain their degree, to increase the probability to get the first job, and to improve the quality of teaching. The performance of the teaching activity surely depends, among other factors, on teaching service and on characteristics and study conditions of the students (quality of teaching, classrooms availability, laboratories, libraries, location of the offered services, mark obtained by students in the secondary school certificate, full-time or part-time study condition, living conditions of the students, etc.). The aim of this study is to verify if there are clear different types of students so that it is necessary to implement difference policies and governance interventions to improve the capability of the students to attend in an efficient way their course programs, to obtain better academic results, and to increase the probability of getting the first job.

281 Deming Lecture ◬ ⌘

Deming Lectureship Committee

Tuesday, August 10, 4:00 pm–5:50 pm

Deming and Bell Labs

◆ Colin L. Mallows, Avaya Labs Research

Avaya Labs Research, 233 Mt. Airy Rd., Basking Ridge, NJ 07920

colinm@research.avayalabs.com

Deming's early interactions with the Bell System (he worked summers at the Western Electric plant at Hawthorne, near Chicago, and later studied with Walter Shewhart) were important in the development of his philosophy of management. How relevant are his ideas to the management of research?

282 ASA Presidential Address and Awards

ASA

Tuesday, August 10, 8:00 pm–9:30 pm

Presentation of Awards

◆ Robert L. Mason, Southwest Research Institute

Southwest Research Institute, 6220 Culebra Rd., San Antonio, TX 78238-5166

RMason@swri.edu

- Presentation of Certificates of Appreciation for Retiring Editors; Presented by Robert L. Mason
- Presentation of Samuel S. Wilks Memorial Medal; Mark P. Becker, Chair, Samuel S. Wilks Memorial Medal Committee
- Presentation of Goffried E. Noether Award; Regina Y. Liu, Chair, Noether Awards Committee
- Presentation of Statistics in Chemistry Award; Raymond L. Lam, Chair, SPES Committee on Chemometrics
- Presentation of Award of Outstanding Statistical Application: Paul S. Albert, Chair, Award of Outstanding Statistical Application Committee
- Presentation of W.J. Youden Award in Interlaboratory Testing; Philip R. Scinto, Chair, W.J. Youden Award in Interlaboratory Testing Committee
- Presentation of Edward C. Bryant Scholarship Award; Jean D. Opsomer, Chair, Edward C. Bryant Scholarship Award Committee
- Presentation of Gertrude M. Cox Scholarship in Statistics Award; Holly B. Shulman, Chair, Subcommittee to Committee on Women in Statistics and the Caucus for Women in Statistics
- Presentation of SPAIG Award; Robert R. Starbuck, Chair, Statistical Partnerships among Academe, Industry, and Government Award Committee

Bayesians, Frequentists, and Scientists

◆ Bradley Efron, Stanford University

Stanford University, Sequoia Hall, 390 Serra Mall, Dept. of Statistics, Stanford, CA 94305-4065

brad@stat.stanford.edu

Broadly speaking, 19th century statistics was Bayesian while the 20th century was frequentist, at least from the point of view of most scientific practitioners. Here in the 21st century scientists are bringing statisticians much bigger problems to solve, often comprising millions of data points and thousands of parameters. Which statistical philosophy will dominate practice? My guess, which I will try to back up with some recent examples, is that a combination of Bayesian and frequentist ideas will be needed to deal with our increasingly intense scientific environment.

Presentation of Founders Awards and New ASA Fellows

◆ Robert L. Mason, Southwest Research Institute

Southwest Research Institute, 6220 Culebra Rd., San Antonio, TX 78238-5166

RMason@swri.edu

- Presentation of Founders Award; Robert L. Mason, Chair, Founders Award Committee
- Presentation of New ASA Fellows; Javier Rojo, Chair, Committee on Fellows

283 Competing Risks Analysis ▲⌘

Biopharmaceutical Section

Wednesday, August 11, 8:30 am–10:20 am

Cumulative Incidence Regression

◆ Jason Fine, University of Wisconsin, Madison

University of Wisconsin, Madison, Dept. of Statistics and Biostatistics, 600 Highland Ave., Room K6-420, CSC, Madison, WI 53792

fine@biostat.wisc.edu

With explanatory covariates, the standard analysis for competing risks data involves modeling the cause-specific hazard functions via a proportional hazards assumption. Unfortunately, the cause-specific hazard function does not have a direct interpretation in terms of survival probabilities for the particular failure type. In recent years, many clinicians have begun using the cumulative incidence function, the marginal failure probabilities for a particular cause in the presence of dependent risks. This quantity is intuitively appealing and more easily explained to the nonstatistician. It is especially relevant in decision analyses in which failure probabilities are needed to evaluate treatment utility. Previously, researchers have considered methods which combine estimates of cause-specific hazard functions under proportional hazards assumptions. However, these methods do not allow the analyst to assess the net effect of a covariate on the cumulative incidence function. We will discuss an alternative semiparametric modeling strategy in which the cumulative incidence is modeled directly, as one ordinarily models the survival function.

Regression Modeling of Competing Risk Data Based on Pseudo-values of the Cumulative Incidence Function

◆ John P. Klein, Medical College of Wisconsin; Per K. Andersen, University of Copenhagen

Medical College of Wisconsin, Division of Biostatistics, 8701 Watertown Plank Rd., Milwaukee, WI 53226

klein@mcw.edu

Key Words: *regression models, general linear models, jackknife, cumulative incidence*

Typically, regression models for competing risk outcomes are based on proportional hazards models for the crude hazard rate of a competing risk. These estimates often do not agree with impressions drawn from plots of cumulative incidence functions for each level of a risk factor. We present a technique that models the cumulative incidence functions directly. The method is based on the pseudo-values from a jackknife statistic constructed from the cumulative incidence curve. These pseudo-values are used in a generalized estimating equation to obtain estimates of model parameters. We study the properties of this estimator and apply the technique to a study of the effect of alternative donors on relapse for patients given a bone marrow transplant for leukemia.

Competing Risks Methods in the Analysis of Clinical Trials for Early-stage Cancer

◆ James J. Dignam, University of Chicago

University of Chicago, 5841 South Maryland Ave., MC 2007, Dept. of Health Studies, Chicago, IL 60637

jdignam@health.bsd.uchicago.edu

Key Words: *Gray's test, cause-specific hazard, cumulative incidence function, subdistribution hazard, logrank test*

In cancer clinical trials, competing risks are frequently encountered. For example, women treated surgically for breast cancer may experience any of the following as a first failure event: recurrence near the removed tumor, recurrence at other anatomic sites, occurrence of a tumor in the opposite breast, occurrence of other cancer types, or death from noncancer causes prior to these events. Cause-specific hazards and cumulative incidence functions are typically used to summarize results by event type. Tests for event-specific differences by treatment group may be based on comparison of cause-specific hazards or quantities associated with the cumulative incidence function, such as the subdistribution hazard or the distance between cumulative incidence function curves. Results of these tests may differ depending on the relative magnitude of different event hazards, and it may be unclear as to which test to apply. Depending on the questions of principal interest, tests on either or both metrics may be appropriate. We present practical examples from cancer clinical trials and simulation study results to illustrate data summary and inferential tools for competing risk observations.

284 Future Directions for Statistical Computing Software

Section on Statistical Computing, Section on Quality and Productivity, Section on Statistical Graphics

Wednesday, August 11, 8:30 am–10:20 am

Some Issues in the Design of Statistical Languages

◆ Luke Tierney, University of Iowa

University of Iowa, Statistics and Actuarial Science, 241 Schaeffer Hall, Iowa City, IA 52242

luke@stat.uiowa.edu

Key Words: *statistical languages*

Languages for statistical computing are now a mainstay of data analysis, statistical practice, and statistical research. Many of the languages in current use were designed to facilitate carrying out relatively simple analyses on, by current standards, relatively small sets of data. These languages are now being challenged by the need for increasingly computationally intensive methods and to deal with larger and more complex forms of data and meta data. This talk will discuss some of the ways in which choices of language design can help current and future languages for statistical computing to meet these challenges.

Embedded Analytics

◆ Leland Wilkinson, SPSS Inc.

SPSS Inc., 233 South Wacker Dr., Chicago, IL 60606

leland@spss.com

Key Words: *statistical computing, data-mining*

Statistical Computing is entering a new age. The last two decades of the last century saw computing systems move from the mainframe to the desktop and from the desktop to the Web. During the same period, there was consolidation of statistical software companies. Surviving statistical software companies are likely to continue serving the needs of both moderately trained and expert statistical users, adding new procedures after they appear in technical journals. Two major trends have emerged in the first few years of this century, however. First, the open-source software movement has spread to statistics and should continue to grow, particularly internationally. The R project is the most noteworthy evidence of this trend. Second, analytic methods are becoming embedded in operating systems and enterprise software. Early indications of this trend are the inclusion of data-mining and statistical systems in Microsoft's Analysis Services software and Oracle's Darwin. Embedded statistical software should lead to more users of statistical methods, but these users will not always be aware they are doing so.

Future Directions for Programming with Data

◆ John Chambers, Bell Laboratories

Bell Laboratories, Lucent Technologies, 700 Mountain Ave., Rm. 2C-282, New Providence, NJ 07974-0636

jmc@bell-labs.com

It is commonplace, but nevertheless accurate, to note that changes in computing hardware and software have radically expanded the resources available for statistical data analysis. In particular, where once analysts might have felt required to implement accessible versions of tools from scratch, now both statistical and general data-related software is available in staggering variety. The same changes, however, have altered the data processes to which data analysis must respond. This talk addresses questions of how current resources can be used effectively, and of how future directions for programming with data may improve our response.

285 Global Players in Global Challenges ▲ ⊞

Business and Economics Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

The Role of Multinational Companies in Developing Countries

◆ Rob van Tulder, Erasmus University

Erasmus University, Rotterdam School of Management, Dept. of Business-Society Management, Room F4-57, PO BOX 1738, Rotterdam, NL 3000 DR The Netherlands

rtulder@fbk.eur.nl

The authors will illustrate the role of multinational companies (MNCs) in developing countries. Based on an industrial analysis, they will evaluate the impact of MNCs on the level of development of, among other countries, China and India, not only in terms of technology spillovers and improved access to world markets, but also in terms of the possible role MNCs should or do play in speeding up development processes.

National Innovation Systems in the Era of Globalization

◆ Cees van Beers, Technical University of Delft

Technical University of Delft, Faculty of Technology, Policy and Management, Jaffalaan 5 NL 2628 BX Delft, The Netherlands

c.p.vanbeers@tdm.tudelft.nl

The authors will focus on the issue of national innovation systems in a globalizing world. In line with the findings of Michael Porter, it is clear that one of the main determinants of national economic growth is technological innovation. However, it is observed that countries differ in (1) the presence of sizable home multinationals, (2) the dominance of host multinationals, and (3) the absorptive capacity needed to benefit from foreign technologies. Hence, there is a strong need for understanding national innovation processes, in order to formulate tailored national innovation policies.

Employment Effects of Globalization

◆ Douglas van den Berghe, United Nations Conference on Trade and Development

United Nations Conference on Trade and Development, Div. on Investment, Technology & Enterprise Development, Palais des Nations, Geneva 10 CH-1211 Switzerland

douglas.vandenberghe@unctad.org

In the last decade, the World Investment Reports, compiled by the UN, have become one of the standards in monitoring the impact of globalization. Although many aspects have been considered, one of the most interesting aspects is employment. First of all, employment remains quite immobile in practice (even though economic regions like the EU ensure labor mobility), and secondly, various motives for internationalization have different implications for employment at the national level. This paper will give an overview of the research in this area.

286 Multilevel Modeling with Survey Data ▲ ⌘

Section on Survey Research Methods, Social Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

Challenges in Bayesian Data Analysis for Survey Data

◆ Andrew Gelman, Columbia University; John B. Carlin, Royal Children's Hospital and University of Melbourne

Columbia University, Dept. of Statistics, 23 Mathematics Bldg., New York, NY 10027

gelman@stat.columbia.edu

Key Words: *multilevel modeling, probability of inclusion, nonresponse, clustering, survey weighting, post-stratification*

The general principles of Bayesian data analysis imply that models for survey responses should be constructed conditional on all variables that affect the probability of inclusion and nonresponse, which are also the variables used in survey weighting and clustering. However, such models can quickly become very complicated, with potentially thousands of post-stratification cells. It is then a challenge to develop general families of multilevel probability models that yield reasonable Bayesian inferences. We discuss in the context of several ongoing public health and social surveys.

Estimation of a Finite Population Mean—an Empirical Restricted Best Prediction Approach

◆ Jiming Jiang, University of California, Davis; Partha Lahiri, University of Maryland

University of California, Davis, Dept. of Statistics, 360 Kerr Hall, One Shields Ave., Davis, CA 95616

jiang@wald.ucdavis.edu

Key Words: *design consistency, empirical restricted best predictor, finite population mean, mean squared prediction error, small-area estimation*

We propose a model-based restricted best (RB) predictor of a finite population mean that minimizes the mean square prediction error (MSPE) among the class of all predictors that depend on the sample only through the traditional design-unbiased estimator of the finite population mean. We then consider an empirical restricted best (ERB) predictor derived from the RB that does not require the knowledge of an explicit model for the unobserved units of the finite population. The proposed ERB enjoys a robust design-consistency property irrespective of the assumed model, and it applies to a wide class of models, including linear and generalized linear mixed models. The case of unknown model parameters is undertaken and a second-order accurate model-based MSPE of the proposed ERB predictor is obtained to measure the precision of the proposed predictor. The method will find applications in small-area estimation.

Bayesian Hierarchical Modeling of NRI Survey Data

◆ Michael D. Larsen, Iowa State University

Iowa State University, Dept. of Statistics, Snedecor Hall 216A, Ames, IA 50011

larsen@iastate.edu

Key Words: *multilevel modeling, survey weights, missing data, imputation, shrinkage estimation, Bayesian modeling*

The USDA Natural Resource Conservation Service (NRCS) conducted the National Resource Inventory (NRI) every five years from 1982 to 1997. The survey design was changed in 2000 to an annual supplemented panel design. The two-stage stratified area sample survey selects land segments as the primary sampling units, then points within the segments as the secondary sampling units. Information is gathered on land use and the presence of water, trees, Rd.s, and structures on nonfederal lands using a combination of fly-over photography and site visits. The results are used in erosion modeling and monitoring Conservation Reserve Program (CRP) lands and wetlands. Efforts to model the hierarchical geographical structure of the data and to account for temporal and spatial relationships will be reported. Use of previous waves of data collection and covariate information will be discussed. This is joint ongoing work with colleagues at Iowa State University.

287 Statistical Geoinformatic Surveillance and Security

Section on Statistical Graphics, Environmental and Ecological Statistics

Wednesday, August 11, 8:30 am–10:20 am

Geographical and Network Hotspot Detection, Prioritization, and Early Warning

◆ Charles Taillie, Pennsylvania State University; Ganapati P. Patil, Pennsylvania State University

Pennsylvania State University, 421 Thomas Bldg., University Park, PA 16802

taillie@stat.psu.edu

Key Words: *geosurveillance, hotspot rating, multicriteria prioritization, partially ordered sets, spatial scan statistic, typology of space-time hotspots*

Hotspot means something unusual-an anomaly, aberration, outbreak, elevated cluster, etc. The responsible factors may be natural, accidental, or intentional. This presentation describes the upper level set (ULS) scan statistic for hotspot detection across geographic regions and across networks. The method is computationally efficient and can identify clusters of arbitrary shape including those that may not be adequately captured by the traditional circle-based scan statistic. The ULS statistic extends to the space-time domain where, because it allows for arbitrarily shaped space-time hotspots, it can characterize the temporal evolution of a spatial hotspot. This leads to the "typology of space-time hotspots." Changing patterns of urban poverty in different metropolitan areas of the U.S. are examples of such typologies. We also describe methods for multi-criteria prioritization of identified hotspots, employing the notion of linear extensions of partially ordered sets. These innovative tools are relevant to syndromic surveillance, cyber surveillance, crop surveillance, invasive species surveillance, drinking water system surveillance, network surveillance, among others.

Coupling Biological Impairment of Freshwater Streams and Human Dimensions

◆ Denice Wardrop, Pennsylvania State University; Wayne Myers, Pennsylvania State University; Patil Ganapati, Pennsylvania State University; Charles Taillie, Pennsylvania State University

Pennsylvania State University, Institutes of the Environment, 216 Walker Bldg., University Park, PA 16802

dhw110@psu.edu

Key Words: *biological integrity, covariate adjustment, critical stream subnetworks, hotspot detection, scan statistic, upper-level set scan statistic*

Human impacts on freshwater streams are well documented and have generated frequent remedial responses, but the goal of restoring their chemical, physical, and biological integrity has not been fully attained. Thus, improvement of the ecological condition of the waters of the U.S. continues to be a broadly pursued goal. The current state of the science identifies several issues: (1) need for a hierarchical process to identify critical stream segments with progressively greater specificity; and (2) need for improvement in the articulation of landscape properties relevant to aquatic resource condition. The presentation describes a new analysis of the stream systems of Pennsylvania that: (1) identifies critical stream subnetworks having low biological integrity, by application of the network version of the upper level set (ULS) scan statistic; and 2) reassesses these subnetworks after taking into account potential explanatory factors. This is accomplished by adjusting expected responses for these factors based on statewide data and re-applying the ULS scan statistic. Three categories of factors are evaluated: descriptive, structural, and network-related.

Mid-Atlantic Watersheds Classification and Prioritization for Protection and Restoration

◆ Wayne Myers, Pennsylvania State University; Mary McKenney-Easterling, Pennsylvania State University; Kristen Hychka, Pennsylvania State University; Bronson Griscom, Canaan Valley Institute; Joseph Bishop, Pennsylvania State University; Gian Rocco, Pennsylvania State University; Robert Brooks, Pennsylvania State University; George Constantz, Canaan Valley Institute; Ganapati P. Patil, Pennsylvania State University; Charles Taillie, Pennsylvania State University

Pennsylvania State University, 124 Land & Water Research Building, University Park, PA 16802

wlm@psu.edu

Key Words: *watersheds, classification, clustering, geographic information systems, environment*

The purpose is to characterize watersheds in a region regarding vulnerability and degradation of water quality due to human impact based on available spatial information. Available information is of five general types: (1) physical and physiographic conformation, (2) soil factors, (3) climatic factors, (4) land-cover/land-use, and (5) prior records of sampling at selected locations for water quality and biological indicators. The strategy is to develop cluster-based classes of watersheds that are expected to have similar responses to anthropogenic stressors, without using metrics that are directly influenced by local human activity. Watersheds in these classes are then analyzed for degree of human influence as indicated by land-cover/land-use and demographics. More sparse data on water quality and biological indicators at stream sampling locations provides a basis for determining the degradation response to human-induced stressors in each class along with potential for remediation. Statistical innovation takes the form of clustering by stages and coupling the clustering process with visualization and expert interaction through the medium of geographic information systems (GIS).

Biostochastics for Remote Sensing of the Environment with Applications

◆ Bo Ranneby, Swedish University of Agricultural Sciences; J. Yu, Swedish University of Agricultural Sciences; M. Ekstrom, Swedish University of Agricultural Sciences

Swedish University of Agricultural Sciences, Centre of Biostochastics, Umea, SE-90183 Sweden

Bo.Ranneby@sekon.slu.se

Key Words: *Gibbs sampler, probabilistic classifiers, quality assessment, resampling, wavelet transforms*

Questions related to environmental health call for a new generation of statistical tools. We describe some statistical methods that are applicable under nonstandard conditions that frequently arise in environmental problems and in remote sensing applications. Error rates of traditional remote sensing classification methods are usually quite high but can be improved by (1) denoising the images using wavelet transforms, (2) using the Gibbs sampler for reclassification, and (3) applying new classification algorithms such as probabilistic classifiers. These classifiers can be used for quality assessments at the pixel level. Resampling of blocks is used for accuracy assessments. An example is determining the total volume of dead wood over some region. Estimates are obtained by summing the predicted volumes over all pixels. A nonparametric method that estimates the variance and the distribution for estimators of this kind has been developed; the approach allows for spatial dependence and for nonstationarity. Consistent variance estimators for totals and functions of totals, e.g. ratio-estimators, together with convergence rates are provided.

288 Adaptive Estimating Equation Approaches for Correlated Data ▲ ⌘

WNAR, Section on Statistics in Epidemiology, ENAR

Wednesday, August 11, 8:30 am–10:20 am

Repeated Measures Analysis: The Secondary Stochastic Processes

◆ Patrick J. Heagerty, University of Washington

University of Washington, Box 357232, Seattle, WA 98195

heagerty@u.washington.edu

Key Words: *longitudinal data, attrition, time-varying covariates*

The analysis of repeated measures requires careful attention to several secondary stochastic processes in addition to the primary response process. For example, it is well known that drop-out and stochastic time-varying covariates can require modification to standard inferential methods otherwise biased estimates will result. In addition, birth processes, death processes, treatment crossover, and measurement timing processes can impact the validity of statistical summaries. This talk will overview several of the key secondary stochastic processes and comment on validity or modification of both likelihood-based and moment-based methods of analysis when repeated measures are subject to these secondary processes.

RKHS Algorithms for Estimation in Random Fields

◆ Christopher G. Small, University of Waterloo

University of Waterloo, Dept. of Statistics & Actuarial Sc., Waterloo, ON N2L 3G1 Canada

cgsmall@uwaterloo.ca

Key Words: *random field, correlated data, estimating function*

In both geostatistical and biomedical applications, data arise which require the estimation of moments when the data are highly correlated. For random fields, which are sampled at finitely many prespecified landmarks, the usual approach is to estimate the mean function of the process using a weighted least-squares estimate whose form is determined in part by inverting the covariance matrix of the sampled process. However, when the covariance matrix is a function of the parameters and the sampling design saturates the domain of observation, this solution becomes computationally problematic. I will propose an RKHS algorithm with linear convergence to the required parameter estimates.

Consistency and Inconsistency of Maximum Quasi-likelihood Estimators

◆ Bing Li, Pennsylvania State University

Pennsylvania State University, 326 Thomas Bldg., Dept. of Statistics, University Park, PA 16802

bing@stat.psu.edu

Key Words: *quasi-likelihood, generalized estimating equations, consistency, path integral*

It has long been speculated that, if a parametric class of estimating equations forms a conservative vector field, then, under some conditions, the maximum point of the potential function should be a consistent estimator of the parameter. This is part of the reason for preferring a maximum quasi-likelihood estimator to other solutions of the quasi-likelihood equation. However, such sufficient conditions have not been established except in special cases. I will discuss two sets of reasonably general sufficient conditions for a maximum quasi-likelihood estimator to be consistent. I will also demonstrate that, if these conditions are violated, it is possible for a maximum quasi-likelihood estimator to be inconsistent. These results will then be applied to study nonconservative estimating equations and generalized estimating equations. In particular, I will discuss what types of path integral of a nonconservative estimating equation with give rise to consistent maximum.

Semiparametric and Nonparametric Models for Correlated Data

◆ Annie Qu, Oregon State University; Runze Li, Pennsylvania State University

Oregon State University, Dept. of Statistics, 44 Kidder Hall, Corvallis, OR 97331

qu@stat.orst.edu

Key Words: *longitudinal data, smoothing spline, goodness-of-fit, nonparametric regression, GEE, quadratic inference function*

Estimating equation approaches is useful for correlated data because the likelihood function is often unknown or intractable. However, estimating equation approaches lacks objective functions for selecting the correct root in multiple root problems, and likelihood-type functions to produce inference functions. A general description is given of the quadratic inference function approach (Qu, et al. 2000), a semiparametric framework defined by a set of mean zero estimating functions, but differing from the standard estimating function approach in that there are more equations than unknown parameters. The quadratic inference function method provides efficient and robust estimation of parameters in longitudinal data settings, and inference functions for testing. Further, an efficient estimator using a nonparametric regression spline is developed, and a goodness-of-fit test is introduced. The asymptotic chi-squared test is useful for testing whether coefficients in nonparametric regression are time-varying or time-invariant.

289 Statistical Methods for Reproductive Health Data ▲ ⌘

Section on Statistics in Epidemiology, Biopharmaceutical Section

Wednesday, August 11, 8:30 am–10:20 am

Assessing the Effects of Reproductive Hormone Profiles on Bone Mineral Density Using Functional Two-stage Mixed Models

◆ Daowen Zhang, North Carolina State University

North Carolina State University, Dept. of Statistics, Raleigh, NC 27696-8203

dzhang2@stat.ncsu.edu

Key Words: *longitudinal covariates, mixed effects models, nonparametric regression, periodic smoothing splines, REML*

In the Study of Women's Health Across the Nation (SWAN), total hip bone mineral density (BMD) was measured together with daily levels of creatinine-adjusted follicle stimulating hormone (FSH) in urine over one menstrual cycle on more than 600 women. It is of scientific interest to investigate the effect of the FSH time profile on the total hip BMD, adjusting for age and body mass index. The statistical analysis is challenged by the complex data structure and tremendous among- and within-woman variation. We propose a measurement error partial functional linear model, where FSH measures are modeled using functional mixed effects models and the effect of the FSH time profile on BMD is modeled using a partial functional linear model by treating the unobserved true woman-specific FSH time profile as a functional covariate. We develop a two-stage estimation procedure using periodic smoothing splines. A key feature of our approach is that estimation at both stages could be conveniently cast into a unified mixed model framework. A simple test for constant functional covariate effect is also proposed. The proposed method is evaluated using simulation studies and applied to the SWAN data.

Order-restricted Bayesian Inference with Applications to Human Fertility Studies

◆ David B. Dunson, National Institute of Environmental Health Sciences

National Institute of Environmental Health Sciences, MD A3-03, PO Box 12233, Biostatistics Branch, Research Triangle Park, NC 27709

dunson1@niehs.nih.gov

Key Words: *Bayesian methods, latent variables, parameter restrictions, variable selection priors, stochastic search, nonlinear modeling*

In many biomedical studies, there is interest in the association between an ordered categorical predictor and a discrete outcome, and it is plausible to assume a priori that the mean is nondecreasing across levels of the categorical predictor. Such order restrictions can greatly improve estimation efficiency and power to detect an association. Motivated by fertility applications, this article proposes a general Bayesian approach for inferences on ordered trends in count, binary, and aggregated binary data. A class of latent count models is proposed, which accommodates a broad variety of data structures while facilitating parameter interpretation and computation. Order restrictions are incorporated through a mixture prior which has restricted support and assigns positive probability to the null hypothesis of no association. This prior is conditionally conjugate after data augmentation, resulting in simple and efficient posterior computation. The methods are applied to studies of predictors of the day-specific probabilities of pregnancy in relation to the estimated day of ovulation.

Piecewise Constant Cross-ratio Estimation for Bivariate Survival Data: Association Between Ages at Onset of a Marker Event and Menopause

◆ Xihong Lin, University of Michigan; Bin Nan, University of Michigan; Lynda Lisabeth, University of Michigan; Sioban Harlow, University of Michigan

University of Michigan, Dept. of Biostatistics, 2420 Washington Heights, School of Public Health, Ann Arbor, MI 48109-2029

xlin@umich.edu

Key Words: *bivariate failure time, Clayton model, cross-ratio function, marker event, menopausal transition, reproductive health*

This article provides an analysis for the association between ages at onset of a marker and menopause using the Tremin Trust data, a long-term cohort study following women through the whole reproductive life span. Although various marker events based on bleeding pattern change have been proposed for early and late transition to menopause clinically, formal statistical analysis for evaluating their associations with age at onset of menopause appears to be complicated. Such an analysis is challenged by the fact that both a marker event and menopause are subject to right-censoring and their association depends on the age at the marker event. We propose to model their association using a cross-ratio model, which is assumed to be a piecewise constant function of the marker event time. Two estimation procedures using the direct and sequential two-stage methods are proposed, and are extended to allow for covariates in marginal survival functions. The proposed methods are applied to the analysis of the Tremin Trust data, and their performance is evaluated using simulations.

290 Careers in Statistics: Past, Present, and Future ⌘

Cmte on Career Development, Section on Statistical Education, Social Statistics Section, Section on Government Statistics

Wednesday, August 11, 8:30 am–10:20 am

Careers in Statistics: Past, Present, and Future

◆ Joseph Waksberg, Westat; ◆ Daniel H. Mowrey, Eli Lilly and Company; ◆ Michael W. Horrigan, Bureau of Labor Statistics; ◆ Tracy Gmoser, Smith Hanley Associates; ◆ John L. Eltinge, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 1950/
PSB, Washington, DC 20212

Eltinge_J@bls.gov

Key Words: *employment, projections*

Over the past 50 years, statistics has grown from a minor subfield of mathematics to a substantial career field whose professionals are widely valued in research, government, and industry. The statistical community today continues to transform itself in response to technical advancement and social change. For those choosing a career, our profession offers intellectual challenges and opportunities for making meaningful contributions to society. We trace the historical path of the statistics profession and examine some of the influential factors—both people and circumstances—that helped shape it. We then present a professional and demographic profile of today's statistical community, emphasizing the main forces for change within and outside of our field. With the help of statistical projections, we look ahead to the possibilities and challenges for the statistical community in the next 50 years. Do we represent a unified discipline? Are we becoming increasingly specialized and diversified as a professional community? We consider expected changes, explore desired changes, and suggest some initiatives that may help carry our profession to new heights.

291 Analysis in the Presence of Nonrandomized Interventions in Randomized Clinical Trials ▲

Biopharmaceutical Section

Wednesday, August 11, 8:30 am–10:20 am

Post-randomization Change in Cancer Treatment: Statistical Analysis of Treatment Effect

◆ Peiling Yang, U.S. Food and Drug Administration; Ning Li, U.S. Food and Drug Administration; Shenghui Tang, U.S. Food and Drug Administration; Yong-Cheng Wang, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1451 Rockville Pike, HFD-710,
Rockville, MD 20852

yangp@cder.fda.gov

Key Words: *noncompliance, survival analysis*

In randomized studies of cancer treatments, oncologists may change treatments prior to assessing the primary endpoint, thus potentially obscuring treatment effects from study treatments. Reasons for change in therapy include drug toxicity, tumor progression, and changes in surrogate markers. This presentation will focus on estimating treatment effect when patients switch treatments from the assigned to the alternative after randomization in oncology clinical trials. We will explore the prognostic factors that may be correlated with treatment noncompliance and discuss issues in statistical analysis from the regulatory perspective.

Post-randomization Change in Cancer Treatment: Statistical Analysis of Treatment Effect

◆ Shenghui Tang, U.S. Food and Drug Administration; Ning Li, U.S. Food and Drug Administration; Peiling Yang, U.S. Food and Drug Administration; Yong-Cheng Wang, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1451 Rockville Pike, HFD-710,
Rockville, MD 20852

Shenghui.Tang@fda.hhs.gov

Key Words: *longitudinal data analysis, survival analysis*

Treatment effect in a randomized oncology study may be obscured due to changes in study arms prior to assessing the primary endpoint. The following approaches will be discussed to adjust the treatment effect due to treatment changes: (1) Longitudinal data up to the treatment change will be used in analysis while informative reasons for the treatment change will be incorporated through a joint model using a linear mixed effects model for repeated measurements and a proportional hazards model with competing risks for the treatment cross-over. (2) Time-to-event data after the cross-over will be modeled using measures at the switching time point to estimate treatment effect.

A Statistical Framework for Characterizing Informed Cross-over in Phase III Clinical Trial

◆ Lue Ping Zhao, Fred Hutchinson Cancer Research Center

Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N,
Seattle, WA 98109

lzhao@fhcrc.org

Key Words: *clinical trial, cross-over, informed cross-over, ICO, likelihood, failure time*

To accelerate the introduction of novel therapy for public consumption, FDA has recently adopted a protocol for conditional approval of a therapeutic agent based on surrogate endpoints, prior to the completion of the Phase III clinical trial. While this protocol is welcomed by industries for accelerating approval process, this conditional approval may be perceived as the "FDA approval" by lay public, resulting in so-called informed cross-over (ICO), i.e., participants in the control arm of the trial may choose to switch from the control arm to the treatment arm, in accordance to their individual prognostic factors at that time, despite the fact that the results with the surrogate endpoint are inconclusive. Analyzing observations from such a trial with ICO, if the percentage of switched subjects is substantial, could be excessively conservative, if the usual intention-to-treat analysis is adopted. Alternative approaches need to be developed for appropriate evaluation of such clinical trial data.

Comparing Different Statistical Approaches in Analysis of Oncology Trial Data with Patient Crossover

◆ Tailiang Xie, Imclone System, Inc.; Ning Li, U.S. Food and Drug Administration

Imclone System, Inc., 33 Chubb Dr., Somerville, NJ 08876

tailiang.xie@imclone.com

Key Words: *treatment crossover, survival data, causal inference, cancer clinical trial*

Treatment crossover from an active control therapy to an experimental therapy occurs commonly in randomized oncology clinical trials, where continuation of therapy in active control arm when disease progression was observed is considered unethical. When treatment crossover occurred, treatment effect of the experimental therapy was confounded and assessment of treatment benefit became very difficult. If the experimental therapy had a beneficial treatment effect, treatment crossover from active control would result in an underestimation of treatment difference, whereas if the experimental therapy had some harmful effects, treatment crossover from active control would result in an over-estimation of treatment difference. Except for many naïve approaches, there is no established method to handle clinical data with treatment crossover. We will discuss commonly used approaches for analyzing survival data in which treatment crossover had happened. We will also propose several alternative approaches. These approaches will be compared via extensive simulations and real examples.

292 Measuring Sexual Orientation for Policy and Research

Section on Government Statistics, Cmte on Gay and Lesbian Concerns in Statistics, Social Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

Asking the Right Questions: Making a Case for Sexual Orientation Data

◆ M.V. Lee Badgett, University of Massachusetts

University of Massachusetts, Dept. of Economics, Thompson Hall, Amherst, MA 01003

lbadgett@econs.umass.edu

Currently, very little information is collected on sexual orientation in the nationally representative surveys that guide much of the investigation of social, economic, and health policy. Asking questions on sexual orientation will help to fulfill the mission of such surveys to measure outcomes both for the population as a whole and population subgroups where a policy role is evident. In many cases, the stated purposes and current uses of survey data may even be seen to require the collection of personal characteristics such as sexual orientation. This paper will first outline the particular areas of research and policy-making that are at stake: families, social service provision, health services, economic development, employment, youth, and business decisions. It will give extended examples of the use of data on sexual orientation in family policy-making and in policies related to youth and education. Finally, the paper will outline the other areas where data collection issues on sexual orientation are most pressing.

Trends in Sexual Orientation Data Collection, Analysis, and Reporting

◆ Randall Sell, Columbia University

Columbia University, 350 Bleecker St., New York, NY 10014

rls39@columbia.edu

Sexual orientation variables are increasingly being included in major information systems and databases. These data are being collected for various purposes including understanding and monitoring social structures, health, wealth, violence, and discrimination, and are being used to develop programs and policy. As part of a workgroup funded by the Ford Foundation, a database is being created to provide detailed information on major information systems and databases that have collected, currently collect, or plan to collect sexual orientation data. This paper will examine trends in data collection, including the types of information systems and databases that include sexual orientation variables (health, economic, marketing, etc), methods of data collection, wording of sexual orientation questions, placement of questions in survey instruments, sample sizes, and response rates. The paper will also discuss how the data from these surveys have been analyzed, reported, and used to design programs or develop public policy. Finally, recommendations will be presented concerning the inclusion of sexual orientation variables in additional information systems and databases.

Benchmarking Census Same-sex Unmarried Partner Data with Other GLBT Survey Data

◆ Gary J. Gates, Urban Institute

Urban Institute, 2100 M St., NW, Washington, DC 20037

ggates@ui.urban.org

The same-sex unmarried partner data from Census 2000 provide an important source of information about the demographic and economic characteristics of gay and lesbian couples in the United States. Research suggests that relatively high levels of under-reporting could affect the accuracy and validity of these data. Understanding the extent of Census 2000 undercounts and the nature of any biases created by an undercount is crucial to careful and credible analyses of Census 2000 data. Two recent surveys, the 1997/1998 Urban Men's Health Study (UMHS) and the 2003 California Statewide GLBT Survey (CSGLBTS) collect what investigators believe are representative samples of GLBT individuals. Each of these surveys utilizes methodologies designed to obtain a representative sample of their populations within their jurisdictions. Each also includes basic demographic information about sex, race/ethnicity, and household composition. This paper uses the UMHS and the CSGLBTS to estimate partnering rates in the gay and lesbian population in the various jurisdictions surveyed in each dataset and then derives estimates of size and demographic characteristics.

Measuring Sexual Orientation in a Highly Marginalized Population

◆ Rebecca Young, Barnard College and NDRI, Inc.

Barnard College and NDRI, Inc., 71 W. 23rd St., 8th Floor, New York, NY 10010

young@ndri.org

Key Words: *sexual orientation, lesbian, gay, measures, minority*

This paper will examine the properties of a new multidimensional measure of "sexual minority status" developed for use in health research with drug-involved, largely poor and ethnic minority women. Prior research has suggested that sexual minority status affects health, health-related behaviors, and the medical and drug treatment service experiences of women drug users in important ways. Unfortunately, classification of sexual minority status can be distorted by such technical issues as question order effects or over-reliance on single-dimension assessments (i.e., measuring behavior only or identity only), as well as social issues such as uneven distribution of risks associated with disclosure of sexual minority status. The discussion will draw upon data from several large, multicity studies of drug users, in-depth qualitative research with sexual minority women drug users, and pilot interviews conducted with 120 women drug users of various sexual orientations.

Measuring Adolescent Sexual Orientation on Health Surveys

◆ Kerith J. Conron, Harvard University School of Public Health

Harvard University School of Public Health, 51 Montebello Rd., #3, Jamaica Plain, MA 02130

kconron@hsph.harvard.edu

As an increasing number of federal and state administered adolescent health surveys include measures of sexual orientation, the need for standardized variables with established validity and reliability increases. This presentation will describe findings from 25 cognitive interviews with Boston-area youth ages 14-19 regarding interpretation of sexual orientation variables included on the Massachusetts Youth Risk Behavior Survey and the Growing Up Today Study (G.U.T.S). Transcribed interview text will be sorted into categories delineated by inclusion and exclusion criteria. The consistency and degree of variation in interpretation of questions and responses between participants will be compared across participants. One or more closed-ended sexual orientation question will be recommended for future inclusion on adolescent health surveys.

293 National Health Interview Survey (NHIS) Mortality Linkage and Public Use File ▲ ⌘

Section on Government Statistics, Section on Health Policy Statistics

Wednesday, August 11, 8:30 am–10:20 am

National Health Interview Survey Linked Mortality File

◆ Christine S. Cox, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 6435, Hyattsville, MD 21774

Ccox@cdc.gov

Key Words: *mortality, National Health Interview Survey, record linkage, National Death Index*

Periodically, the National Center for Health Statistics (NCHS) links the National Health Interview Survey (NHIS) with potential mortality information from the National Death Index (NDI) using a standard record linkage procedure. The resulting linked file includes death information for every NHIS respondent with a potential linked death certificate, along with an indication of the strength of the link. More detailed death certificate information is available on a restricted basis in the Research Data Center (RDC) in Hyattsville, Maryland. The present linkage covers survey years 1986 through 1999 with deaths from NDI years 1986-2001. The wealth of health, social, and economic data from the NHIS combined with mortality and cause of death information provides a unique picture of the factors influencing mortality across a wide spectrum of the population in the U.S.

Methodological Improvements in NHIS-NDI Record Linkage

◆ Bryan Sayer, Social & Scientific Systems, Inc.

Social & Scientific Systems, Inc., 8757 Georgia Ave, STE 1200, Silver Spring, MD 20910-3739

bsayer@s-3.com

Key Words: *record linkage, probabilistic matching, mortality, National Health Interview Survey, National Death Index*

A number of improvements in the record linkage methodology are implemented in the 2003 linkage. These include an expanded use of alternate submission records, more classes of potential matches reflecting a greater refinement of matching characteristics, and improved measures of matching probability. Alternate submission records improve the chance of selecting the correct certificate in the selection process, but may increase the risk of a false positive in the scoring process. We also add several additional scoring measures, including the total possible score, the ratio of the obtained score to the possible score, and model-based estimates of the probability of a match and the probability of a nonmatch. These additional estimates may be used to facilitate an analysis of the sensitivity of score cut-offs in determining mortality status.

Weighting and Sample Design Issues in the NHIS-NDI Linked File

◆ Wenxing Zha, National Center for Health Statistics; Myron Katzoff, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

wzha@cdc.gov

Key Words: *mortality, National Health Interview Survey, weighting, sample design, record linkage, National Death Index*

Both refusal of permission to link the NHIS with the NDI (and other administrative files) and the conservative results of the record linkage change the nature of the sample from the original sample design. In addition, many users aggregate multiple years of data to accumulate enough cases for analysis. However, the NHIS underwent a change in sample design between 1994 and 1995, and stacking the 1995-2000 years with the 1986-1994 years may yield incorrect variance estimates. Lack of permission to match, insufficient data to match, and missed matches result in record nonresponse. We compare observed mortality from the NHIS-NDI linkage to projected mortality rates from period and cohort life tables, and consider the bias in estimates.

The Mortality Impact by Race and Sex of Using Alternate Cutpoints for Determining a Match

◆ Gloria Wheatcroft, National Center for Health Statistics; Jennifer Parker, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

gkw2@cdc.gov

Key Words: *mortality, National Health Interview Survey, record linkage, National Death Index*

The addition of multiple measures of match quality has provided users with new alternatives for determining whether a specific death certificate belongs to the respondent. Consequently, it is now possible to vary the cutpoints for determining a match and perform a sensitivity analysis of the impact on mortality. We examine the changes in mortality by basic demographic characteristics including age, race, sex, and country of birth using a range of cutpoints. In addition, we compare estimates obtained from individual survey years with those resulting from combining the microlevel data.

294 Environmental Applications of Mathematical Geology ▲

International Association of Mathematicial Geology (IAMG), Section on Statistics and the Environment

Wednesday, August 11, 8:30 am–10:20 am

Revisiting the Join-count Statistic to Assess Spatial Association without Assuming First-order Homogeneity

Sandor Kabos, Eotvos Lorand University; ◆ Ferenc Csillag, University of Toronto

University of Toronto, Dept. of Geography, 3359 Mississauga Rd., Mississauga, ON L5L1C6 Canada

fcs@geog.utoronto.ca

Key Words: *spatial statistics, testing for spatial randomness, stationarity, environmental applications*

Characterization of spatial pattern is an increasingly frequent task in geographical, political, epidemiological, ecological analysis. A widely used classical piece of spatial statistics for the assessment of spatial association of nominal data, such as colors on a map, is the ‘ordinary’ join-count statistic introduced by O. Moran. Since it is based on counting color-neighbor frequencies, it can reject the null hypothesis (of spatial randomness) due to deviations in the first-order parameters (the probabilities of the colors), or the second-order parameters (the probabilities of color neighbors, or autocorrelation), or both. We have developed a generalized join-count statistic (H.Moran), which specifically tests the second-order properties assuming that the (potentially heterogeneous) probabilities of colors are known. The variance calculation for the test statistic accounts for the covariance induced by the connecting edges in the planar graph of sites over which the variables are observed. Using stochastic simulation experiments we show how O. Moran can generate misleading results and we also confirm that the asymptotic Gaussian approximation holds.

Spatial-temporal-frequency Analysis of Run-off and Groundwater in the Greater Toronto Area, Canada

◆ Qiuming Cheng, York University/China University of Geosciences

York University/China University of Geosciences, 4700 Keele St., Toronto, ON M3J1P3 Canada

qiuming@yorku.ca

Key Words: *spatial analysis, water systems, statistical analysis, hydrological modeling, glacier landscape, GIS*

It is essential to understand the interactions among precipitation, groundwater recharge/discharge, and surface water run-off for environmental planning and water resources management in an area. Modeling landscape with high-resolution digital elevation model (DEM) in a geographic information system can associate rainfall, run-off, and groundwater systems and model their interactions in the watershed context using various statistical models. It has been statistically demonstrated that topological, hydrological, and geological characteristics of drainage basins in the Oak Ridges Moraine Area have significant influences on locations of springs and flowing wells, response of river flow to rainfall events, and the short-term and long-term dependency of run-off. Base flow component was separated from the total river flow using an integrated spatial-frequency analysis model to reflect the flow components supplied by ground water discharge. The remaining surface run-off was further associated to rainfall event through a surface run-off SCS model with the input parameters estimated using a nonlinear multivariate regression.

Estimation of Background and Threshold in Applied Geochemistry

◆ Robert G. Garrett, Geological Survey of Canada; Peter Filzmoser, Vienna University of Technology; Clemens Reimann, Geological Survey of Norway

Geological Survey of Canada, 601 Booth St., Ottawa, ON K1A 0E8 Canada

garrett@NRCan.gc.ca

Key Words: *geochemistry, outliers, background, multivariate, threshold*

Geochemical methods are used in geological and environmental sciences for both mineral exploration and studies of anthropogenic impacts on the environment. The Earth's surface chemistry is heterogeneous: different rocks, surface environments, and various natural and anthropogenic processes lead to different statistical populations. A task is to identify these populations and estimate their properties so that individuals likely not belonging to them can be identified. Univariate approaches include graphical and formal methods for threshold, i.e., boundary of background, selection. Most geochemical data are multivariate and graphical methods involve subjective decisions and formal methods involve various assumptions. A procedure that identifies the point of greatest difference between the empirical and theoretical distributions of Mahalanobis distances has been developed that leads to the identification of a background dataset that permits the identification of outliers, i.e. individuals falling beyond the threshold. The procedure is demonstrated with data from an ecogeochemistry survey in Northern Europe.

Spatial Distribution of Metals in the Environment Around a Copper Smelter in Western Quebec, Canada

◆ Graeme Bonham-Carter, Geological Survey of Canada

Geological Survey of Canada, 601 Booth St., Ottawa, ON K1A 0E8 Canada

gbonhamc@nrcan.gc.ca

Key Words: *atmospheric pollution, smelter emissions, nonlinear model*

Concentrations of metals in snow, peat, soil, lake water, and lake sediment judged from regional sample surveys show a roughly circular anomaly pattern around the smelter at Rouyn-Noranda, Quebec. Concentrations near the smelter are up to three orders of magnitude greater than regional background levels. The decrease in concentration away from the smelter is approximately exponential, and the distance at which values are indistinguishable from background is about 65 km +/- 5 km. The spatial patterns are anisotropic due to wind effects, but weakly so. Deposition rates of metal calculated for snow and peat integrated over the area of the smelter-centred anomaly allow estimates to of the amounts of metal deposited from the smelter, as distinct from metal levels due to natural sources in background. Deposition tonnage can be compared with total emission tonnage by metal reported from the smelter, and the difference estimates the amount of metal transported in the atmosphere beyond the local anomaly. Deposition rates of smelter metal from the atmosphere beyond this distance are so small due to dilution that detection on the ground is difficult, except perhaps with isotope data.

Conditional Independence Testing in the Weights-of-Evidence Method

◆ Frederik P. Agterberg, Geological Survey of Canada

Geological Survey of Canada, 601 Booth St., Ottawa, ON K1A 0E8 Canada

frits@rogers.com

Key Words: *weights-of-evidence, conditional independence, Bayes's rule*

Weights-of-Evidence (WofE) modeling is a statistical technique for correlating a pattern of n points for discrete events with p map patterns. Originally based on a medical expert system, WofE was developed for mineral potential mapping using Geographic Information Systems. With software including the WofE Arcview extension freely available on the internet, there now are many users including geoscientists concerned with environmental applications. Today, nearly half of WofE users are biologists interested, for example, in wild life habitats. The indicator map patterns used in practical applications frequently are reduced to binary form. Bayes's rule is used to estimate positive or negative weight for presence or absence, and zero weight is assigned to missing data. The contrast representing the difference between positive and negative weight often is maximized to create a suitable binary map layer. The weights for p map layers are added assuming their conditional independence with respect to the point pattern. The sum of all posterior probabilities for arbitrarily small unit cells in the study region then should be equal to n.

295 Student Paper Award Session in Health Policy Statistics

Section on Government Statistics, Section on Health Policy Statistics

Wednesday, August 11, 8:30 am–10:20 am

Covariate Adjustment in Clinical Trials with Nonignorable Missing Data and Noncompliance

◆ Douglas Levy, Harvard University; Alistair J. O'Malley, Harvard Medical School; Sharon-Lise T. Normand, Harvard Medical School

Harvard University, Dept. of Health Care Policy, 180 Longwood Ave., Boston, MA 02115

delevy@fas.harvard.edu

Key Words: *covariate adjustment, instrumental variables, missing data, noncompliance, randomized clinical trials*

Estimating causal effects in psychiatric clinical trials is often complicated by treatment noncompliance and missing outcomes. While new estimators have recently been proposed to address these problems, they do not allow for inclusion of continuous covariates. We propose estimators that adjust for continuous covariates in addition to noncompliance and missing data. Using simulations, we compare mean squared errors for the new estimators with those of previously established estimators. We then illustrate our findings in a study examining the efficacy of clozapine vs. haloperidol in the treatment of refractory schizophrenia. For data with continuous or binary outcomes in the presence of noncompliance, nonignorable missing data, and a covariate effect, the new estimators generally performed better than the previously established estimators. In the clozapine trial, the new estimators gave point and interval estimates similar to established estimators. We recommend the new estimators as they are unbiased even when outcomes are not missing at random and they are more efficient than established estimators in the presence of covariate effects under the widest array of circumstances.

Handling Missing Longitudinal Covariates in Child-development Study: A Functional Multiple Imputation Approach

◆ Yulei He, University of Michigan; Trivellore E. Raghunathan, University of Michigan

University of Michigan, Dept. of Biostatistics, School of Public Health, Ann Arbor, MI 48109

yuleih@umich.edu

Key Words: *children's health development, functional mixed model, longitudinal data, missing covariate, multiple imputation, PSID*

In longitudinal studies, information from the same set of subjects are collected repeatedly over time and such information can be used as covariates for subsequent analysis. However, repeatedly measured covariates are often subject to missing data, which imposes a serious difficulty to further analysis. For missing time-dependent covariate problem in longitudinal studies, the commonly used complete-case, available-case, and last value carried forward methods may give misleading results. We propose a multiple imputation approach based on functional mixed model, which characterize the covariate trends by nonparametric functions. Gibbs sampling algorithm is used to impute the missing values from their posterior predictive distributions. We compare various methods by Monte Carlo simulations and find that the proposed multiple imputation methods performs the best. We apply our method to data from the panel study of income dynamics (PSID). Our post-imputation analysis results suggest that family's economic status at the critical child developmental period may have a significant impact on the children's health development.

Inclusion of Sampling Weights in Mixed-Effects Models: A Comparison of Two Methods

◆ Marianne Bertolet, Carnegie Mellon University; Joel Greenhouse, Carnegie Mellon University; Howard Seltman, Carnegie Mellon University; Kelly Kelleher, The Ohio State University

Carnegie Mellon University, Dept. of Statistics, Baker Hall, Pittsburgh, PA 15213

mhb2@stat.cmu.edu

Key Words: *mixed-effects models, sampling weights*

Hierarchical models can provide valuable information when analyzing survey data. Because many health policy decisions are based on data collected by surveys, this type of research contributes to health policy research. As examples of health policy applications, Pfeffermann et al. (1998) analyzed a survey of psychiatric morbidity which looked at the relationship between demographic characteristics on the prevalence of psychiatric diseases. As another example, Korn and Graubard (2003) analyzed data from the Hispanic Health and Nutrition Examination Survey to analyze height-for-age percentiles in the Mexican-American community. For the survey analysis to provide meaningful results, the sampling weights must be properly incorporated. This research will help explain the differences between the methods, show with simulations the strengths and weaknesses of the methods, and provide software to make the methods more accessible to health policy researchers.

*** To view the abstract for **Comparison of Fixed and Random-effects Methods for Predicting Cancer Incidence in Iowa Counties Using SEER Data**, please see the Online Program at *www.amstat.org/meetings/jsm/2004/onlineprogram/*.

Statistical Model Comparisons for Prediction of Mental Health and Substance Abuse Cost in the Veterans Health Administration

◆ Maria E. Montez, Boston University; Cindy L. Christiansen, Boston University; Susan Loveland, Boston University; Susan L. Ettner, University of California, Los Angeles; Priti Shokeen, Boston University; Amy K. Rosen, Boston University

Boston University, Biostatistics Dept. & CHQOER, VA Medical Center (152), 200 Springs Rd., Bedford, MA 01730

montez@bu.edu

Key Words: *generalized linear models, logged dependent variable, retransformation, risk adjustment*

Previous research shows that risk-adjustment (R-A) systems underestimate total health care costs for individuals with Mental Health & Substance Abuse (MH/SA) disorders. We examine how predictive ability of a R-A system depends on the statistical model chosen. The sample consists of 914,225 MH/SA patients. We regress four generalized linear models on untransformed cost, defined by distributional assumption and link function used: (1) Gaussian identity; (2) Gaussian log; (3) Gamma log; and (4) Gamma square root. A fifth model assumes a log-normal distribution. Regressors include age, sex, and 32 Adjusted

Diagnostic Groups. Robust regression is used to account for variations in services within regions. Each model's predictive ability is evaluated using the root mean square error (RMSE) and mean absolute prediction error (MAPE). Predictive ratios (PRs) are calculated for 12 MH/SA categories. Models' performances do not differ when looking at the RMSE and MAPE. RMSEs range from 10036.3 to 10086.7 and MAPEs range from 2815.7 to 3076.9. The GLM model where a Gaussian distribution and a log link is assumed, has more PRs closer to 1.0 than the other four models.

296 Analyses of Call History to Improve Telephone Surveys—General ▲ ⌘

Section on Survey Research Methods, Social Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

Gaining Efficiencies through Flexible Calling Rules in Large RDD National Survey

◆ Paul J. Lavrakas, Nielsen Media Research; Charles D. Shuttles, Nielsen Media Research; Jennie Lai, Nielsen Media Research; Jeffrey A. Stec, Intecap, Inc.

Nielsen Media Research, 770 Broadway, 14th Floor, New York, NY 10003-9595

paul.lavrakas@nielsenmedia.com

Key Words: *RDD surveys, response rates, advance letters*

Each year Nielsen Media Research (NMR) uses a RDD frame to sample respondent households for its national mixed mode diary surveys of television viewing in the United States. In particular, four times each year (January/February, March/April, June/July, and October/November) NMR conducts a survey of television viewing that begins with a RDD stage. The frame used is all possible 100-banks of telephone number area code/prefix/suffix combinations with at least one number listed in the bank. In each of these surveys upwards of one million RDD telephone numbers are randomly split into four replicates; collectively the four replicates are used for the survey's sampling pool. Following fixed calling rules, which space up to 15 call attempts per telephone number over a 14-16 day field period, NMR makes approximately eight million dialings during each survey. Callbacks are spaced across different days of the week and times of day. This effort leads to far better than average response rates compared to most commercial survey efforts. In the Oct/Nov 2003 survey, for example, an AAPOR Response Rate 1 of 43% and an AAPOR Response Rate 3 of 47% were achieved. This paper will present analyses using NMR's 2003 calling data to illuminate where efficiencies are likely to be gained by changing the fixed calling rules NMR now uses to more flexible calling rules. Past research along these lines suggested that efficiencies may be gained, while keeping the total number of dialings fairly constant, if fewer than 15 call attempts were made to certain call history patterns and more than 15 call attempts were made to certain other call history patterns. We will investigate these findings and present means by which efficiencies may be gained using a more flexible set of calling rules.

Using Call History in a Panel Survey to Better Inform Current Interview Scheduling

◆ Charles C. Mason, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Mass. Ave. N.E. Room 3216, Washington, DC 20212

Mason_C@bls.gov

Key Words: *household surveys, RDD, telephone surveys*

The Bureau of Labor Statistics Point of Purchase survey collects a frame of retail establishments where consumers purchase goods and services. The frame is used to select outlet samples for the items covered by the Consumer Price Index. Data for the survey is collected using a random digit dialing (RDD) computer-assisted telephone interview (CATI) panel survey where households are identified and surveyed initially through an RDD survey. Four quarterly follow-up surveys are conducted using CATI. This paper examines the call history for the RDD component of the survey and reports on an attempt to build a model of optimal call back times for the quarterly panel component of the survey.

Targeted Call-scheduling for Demographic Subgroups

◆ Ellen C. Mecklenburg, NORC, University of Chicago

NORC, University of Chicago, 55 East Monroe Suite 4800, Chicago, IL 60603

mecklenburg-ellen@norc.net

Key Words: *telephone surveys, call-scheduling, random digit dialing*

In Random Digit Dial (RDD) telephone surveys, significant survey resources are used in attempting to reach and obtain cooperative respondents. Effective choice of call times can produce significant savings. In the CDC-sponsored REACH 2010 Risk Factor Survey, many of the targeted respondents are from specific geographic areas and racial/ethnic populations; this presents the additional challenge of optimizing the call times for these particular groups of respondents. The relationship between the time of the initial call and outcome of the call is investigated in terms of sample characteristics. Specifically, we investigate whether the relationship between time of call (by time of day and day of the week) and its outcome varies by exchange-level demographic characteristics. Call outcome is defined in terms of contact rate and cooperation rate. A model is proposed and evaluated.

297 Hot Streaks: Some Recent Evidence ▲

Section on Statistics in Sports

Wednesday, August 11, 8:30 am–10:20 am

Bowlers' Hot Hands

◆ Gary Smith, Pomona College

Pomona College, 425 North College Ave., Claremont, CA 91711

gsmith@pomona.edu

Key Words: *hot hands, bowling, self-efficacy*

Gilovich, Vallone, and Tversky's (1985) analysis of basketball data debunked the common perception that players sometimes have "hot hands." However, their basketball data do not control for several confounding influences. Our analysis of professional bowling indicates that, for many bowlers, the probability of rolling a strike is not independent of previous outcomes and the number of strikes rolled varies more across games than can be explained by chance alone. For example, most bowlers have a higher strike proportion after j consecutive strikes than after j consecutive nonstrikes, and this difference becomes more pronounced as j increases from one to four.

An Argument for the Existence of Streak Shooting in Basketball

◆ Mark Hallahan, College of the Holy Cross

College of the Holy Cross, Dept. of Psychology, Worcester, MA 01610-2395

mhallaha@holycross.edu

Key Words: *streak shooting*

This paper questions the widespread belief that perceptions of streak shooting in basketball are illusory. Limitations of the existing research on this topic will be discussed. Past research that has failed to find statistically significant dependence within shot sequences has had low statistical power and has not adequately considered the possibility of self-correcting sequences (i.e., negative dependence). Three sources of shot data, the NBA long-distance shootout (1994-2003), Gilovich, Vallone, and Tversky's (1985, study 4) college basketball sample, and a new sample of recreational players, raise further questions about the assumption that shot sequences resemble a random process. These data indicate that (a) there are stable individual differences in the degree and direction of dependence in individuals' shot sequences, and (b) distributions of shooters have shapes that differ from a normal distribution of error for a population with no dependence among shots. These data support a model in which there are distinct types of individual shooters, including both streaky and self-correcting shooters, as well, if not better, than a model of shot sequences as entirely random.

Do Golf Professionals Streak?

◆ Russell D. Clark, III, University of North Texas

University of North Texas, Psychology Dept., PO Box 311280, Denton, TX 76203-1280

rdclark@unt.edu

Key Words: *golf, hot-hand, streakiness*

A study investigated streakiness among professional golfers. Hole-to-hole scores within tournaments from a random group of 35 players on the 1997 PGA Tour were analyzed. There was no evidence for a positive relationship between the outcomes of successive holes. Players were just as likely to score par or better following an above par hole as make a par or better following a par or better hole. These results are consistent with those found for individual players in baseball and basketball.

Streakiness in Team Performance

◆ James Albert, Bowling Green State University

Bowling Green State University, Dept. of Mathematics and Statistics, Bowling Green, OH 43403

albert@bgnet.bgsu.edu

Key Words: *hot-hand, baseball, random-effects*

In the 2002 baseball season, the Oakland Athletics had a 20-game winning streak that was one of the longest in baseball history. A standard p value calculation would suggest that this was indeed an unusual sports accomplishment. A random effects model is constructed for baseball competition and various streaky statistics are simulated from the predictive distribution from this model. Using this model, we see that streaks of length 20 or longer are possible and we should expect to see a streak similar to that of Oakland in every 25 years of baseball competition.

298 Bayesian Methods in Environment, Forestry, and Wildlife Science

Section on Bayesian Statistical Science

Wednesday, August 11, 8:30 am–10:20 am

Bayesian Estimation of Wildlife Population Size for Small Sample Data

◆ Chong He, University of Missouri, Columbia

University of Missouri, Columbia, Columbia, MO 65211

chong@stat.missouri.edu

Key Words: *Bayesian estimation, population size, small sample, capture-recapture, noninformative prior*

We consider the problem of estimating the population size using capture-recapture data in a Bayesian framework. The model M_t has prodigious application in wildlife management, ecology, software liability, epidemiological study, census undercount, and other research areas. Previous studies showed that Bayesian inference can have a sensitive dependence upon the hyper-parameters of the prior distribution. Four commonly used noninformative priors are onsidered. The optimal choice of noninformative priors depends on the number of sampling occasions. Some guidelines on the choice of noninformative priors are provided based on the simulation results. Propriety of applying improper noninformative prior is discussed. Simulation studies are developed to inspect the frequentist performance of Bayesian point and interval estimates with different noninformative priors under various population size, capture probabilities, and the number of sampling occasions. The simulation results show that the Bayesian approach can provide more accurate estimates of the population size than the MLE for small sample. Two real-data examples are given to illustrate the method.

Selection of Spatial Correlation Structures with Bayesian Model-averaging with Application to Loblolly Pines

◆ Edward L. Boone, University of North Carolina, Wilmington; Bronson Bullock, North Carolina State University

University of North Carolina, Wilmington, Dept. of Mathematics and Statistics, 601 S. College Rd., Wilmington, NC 28403

boonee@uncw.edu

Key Words: *spatial statistics, Bayesian model-averaging, forestry*

Many applications of statistical methods for data that are spatially correlated require the researcher to specify the correlation structure of the data. This can be a difficult task due to the fact that there are many candidate structures. Some spatial correlation structures depend on the distance between the observed data points while others rely on neighborhood structures. Bayesian methods that systematically determine the “best” correlation structure from a predefined class of structures are proposed. Bayes factors, highest probability models, and Bayesian model-averaging are employed to determine the “best” correlation structure and to average across these structures to create a nonparametric alternative structure. An application is given for a loblolly pine dataset with known coordinates. Tree diameters and heights were measured and an investigation into the spatial dependence between the trees was conducted. Results showed that the most probable model for the spatial correlation structure agreed with allometric trends for loblolly pine.

Bayesian Logistic Modeling in the Classification Analysis of Water Quality

◆ Huizi Zhang, Virginia Polytechnic Institute and State University; Eric P. Smith, Virginia Polytechnic Institute and State University; Keying Ye, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, 406 A Hutcheson Hall, Dept. of Statistics, Blacksburg, VA 24061-0439

huzhang@vt.edu

Key Words: *Bayesian logistic modeling, water quality, MCMC method*

In water quality studies, logistic regression is used to predict toxic/nontoxic conditions of water systems given relatively available chemical and/or environmental information. When data display hierarchical structure, Bayesian logistic modeling provides as a natural alternative to either a grand model or several separate models for estimating location-specific parameters while taking into account the correlations among the data. Using matched data on sediment toxicity and chemical concentrations from the National Oceanographic and Atmospheric Administration from different coastal locations in the United States, we developed a predictive model for classifying water quality. Dimensionality reduction on the original 22 metal contamination and polycyclic aromatic hydrocarbon contamination variables was performed using Principal Components Analysis. Empirical Bayesian methods were used in developing the prior distributions and Markov chain Monte Carlo (MCMC) method was used to obtain predictions. Furthermore, goodness of fit and predictive ability of both ordinary logistical models and Bayesian logistic models were compared and their respective strength/weakness discussed.

Detecting Pattern in Biological Stressor Response Relationships Using Model-based Cluster Analysis

◆ Ilya A. Lipkovich, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

lipkovichia@lilly.com

Key Words: *water quality, regression, Bayesian methods, Markov chain Monte Carlo*

Environmental monitoring of aquatic systems is needed to estimate the quality of the systems, to evaluate standards and to study stressor-response relationships. The monitoring program focuses on the collection of biological, chemical, and physical measures of the system. An important concern is the effect of chemical and physical stressors on the biological community. From a management perspective, interest is on what factors affect the biological community and how the relationship changes within the state. The focus of this paper is on the use of cluster analysis as a tool for finding relationships between a single biological response and environmental stressors. The approach to cluster analysis is based on the use a model-based analysis using a penalized classification likelihood. The mean parameter depends on the relationship as well as the cluster. This approach allows for simultaneous development of regression models and clustering of the regression models. The implementation of the approach is based on Markov chain Monte Carlo Model Comparison. This method is applied to the analysis of a dataset describing stressors/response relationship in one of the Ohio eco-regions.

Model Selection in Canonical Variate Analysis Using Bayesian Model Averaging

◆ Robert Noble, Miami University; Eric P. Smith, Virginia Polytechnic Institute and State University; Keying Ye, Virginia Polytechnic Institute and State University

Miami University, Dept. of Mathematics & Statistics, Oxford, OH 45056

noblerb@muohio.edu

Key Words: *Bayesian model averaging, model selection, canonical variate analysis*

The standard methodology when building discriminant analysis models has been to use one of several algorithms to systematically search the model space for a good model. If the number of variables is small then all possible models or best subset procedures may be used, but for datasets with a large number of variables, a stepwise procedure is usually implemented. The stepwise procedure of model selection was designed for its computational efficiency and is not guaranteed to find the “best” model with respect to any optimality criteria. Many times there will be several models that exist that may be competitors of the best model in terms of the selection criterion, but classical model building dictates that a single model be chosen to the exclusion of all others. An alternative to this is to use Bayesian model averaging (BMA), which uses the information from all models based on how well each is supported by the data. The BMA model is weighted average of all possible models where the individual variable configuration.

299 Bayesian Methods for the Social Sciences

Section on Bayesian Statistical Science

Wednesday, August 11, 8:30 am–10:20 am

A Bayesian Network Model of Trademark Dilution

◆ Greg M. Allenby, The Ohio State University

The Ohio State University, 2100 Neil Ave., Columbus, OH 43210

allenby.1@osu.edu

Key Words: *memory, consumer behavior, MCMC, marketing, neural network*

Trademark dilution refers to the harm that may result to a trademark owner when its well-known mark, such as a brand name, is used in an unauthorized manner by another firm in an unrelated product category. Imagine that a consumer, while shopping, came across a product called Kodak bicycles (a product not made by the Eastman Kodak company). If the consumer were to mistakenly infer that the bicycle was manufactured or authorized for manufacture by Eastman Kodak, s/he would be said by legal practitioners to be confused regarding product source. If a significant number of consumers exhibited a similar likelihood of consumer confusion, this would constitute the basis for a standard case of trademark infringement. The Kodak bicycle scenario described here, one of the earliest cases of trademark dilution, was decided in Kodak's favor in 1898 (*Eastman Photographic Materials Co. v. John Griffith Corp.*, 1898), with the bicycle maker prohibited from further use of the Kodak name. We propose a statistical model for measuring the effects of various variables used in trademark litigation on the consumer's ability to recall product categories associated with given brand names.

Nonparametric and Parametric Bayes Models for Item Response Data

◆ Kristin Duncan, The Ohio State University; Steven MacEachern, The Ohio State University

The Ohio State University, 2002 W. Schantz Ave., Kettering, OH 45409

duncan.898@osu.edu

Key Words: *nonparametric Bayes, item response, cross-validation*

Both classical nonparametric models and Bayesian parametric models are commonly used in psychometrics, but Bayesian nonparametric models have yet to receive much attention in this field and offer some improvements over current models. Ling Qin took a first look at the problem in her 1998 dissertation where she used the two-parameter logistic curve as the prior mean for the item response curve. We extend her work by using the three-parameter logistic curve as the prior mean. This allows us to study sensitivity to the prior distribution. Bayesian nonparametric models are computationally intensive, so we compare results with these models to those obtained with their faster classical nonparametric counterparts. Bayesian cross-validation is used to assess the fit of the models.

Bayesian Hierarchical Models for Process Dissociation Procedure in Memory Research

◆ Jun Lu, University of Missouri, Columbia; Dongchu Sun, University of Missouri, Columbia; Paul L. Speckman, University of Missouri, Columbia; Jeff Rouder, University of Missouri, Columbia

University of Missouri, Columbia, 146 Middlebush Hall, Columbia, MO 65211-4100

JL09B@mizzou.edu

Key Words: *Bayesian hierarchical model, probit model, linear mixed model, process dissociation procedure*

Memory serves multiple functions in different ways. Many multiple memory theories consist of two primary components: (1) the "recollection," which refers to a conscious recall of an event in the past, and (2) the "automatic activation," which reflects the unconscious activation of previously encountered events. These two components are typically correlated. Such correlation makes it difficult to estimate the parameters of the automatic and recollective components. Thus, we propose a Bayesian hierarchical model to get stable estimates under a Process Dissociation Procedure (PDP) experiment. The model features two probit links. Each probit link has a form of linear mixed model with additive components that reflects the participant effects and the item effects. We assume correlations across the effects between two linear models. Two sets of prior distributions are proposed to model the correlation: the shared-component prior and the Wishart prior. Bayesian computation can be done via MCMC. Simulation studies are conducted to evaluate the performance of the Bayesian approach. We also provide the derivation of Bayes Factors and an alternative memory experiment design.

Fitting Response Time Models by Adaptive Importance Sampling

◆ Cheryl Niermann, The Ohio State University; Mario Peruggia, The Ohio State University; Trisha Van Zandt, The Ohio State University

The Ohio State University, 227 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210

niermann@stat.ohio-state.edu

Key Words: *Bayesian modeling, MCMC algorithms, mixture distributions, shifted exponential distribution, shifted Weibull distribution*

Due to the growing popularity of Bayesian models in the social sciences, there is a need to develop user-friendly computational tools to perform model-fitting and model assessment tasks. A software environment like BUGS allows a user to perform MCMC estimation on the basis of approximate posterior draws. The user need only specify the Bayesian model in a simple and intuitive manner without having to write specialized computer code to implement the MCMC algorithm. However, the choice of models that can be fit is limited by the list of distributions allowed by BUGS. In our analyses of response time data arising from cognitive experiments we used importance sampling to reweight MCMC output from BUGS and fit models whose likelihoods cannot be handled directly. This talk will describe the models that we fit, will outline the strategies that we used to construct and adapt sequentially the importance sampling densities so as to improve the accuracy of our inferences, and will illustrate the performance of the methodology on simulated and real data.

Modeling Dependence in Response Time Data

◆ Mario Peruggia, The Ohio State University; Peter F. Craigmile, The Ohio State University; Trisha Van Zandt, The Ohio State University

The Ohio State University, Dept. of Statistics, 404 CH, 1958 Neil Ave., Columbus, OH 43210-1247

peruggia@stat.ohio-state.edu

Key Words: *probability inverse transformation, ARMA models, Bayesian models, ex-Gaussian distribution, long-range dependence*

Human response time data are widely used in cognitive psychology to evaluate theories of mental processing. Typically, the data constitute the times taken by a subject to react to a succession of stimuli under varying experimental conditions. The sequential nature of the experiments induces serial dependencies in the data that are often ignored in the analysis. We use data from a variety of experiments to explore in detail the nature of these dependencies and compare several alternative modeling strategies. We employ a Bayesian framework to incorporate known characteristics of RT data into our models. Our approach allows us to investigate explicitly the effects of experimental covariates and of outlying observations on the types of dependencies that might arise.

300 Interviewer Falsification of Survey Data ▲

Section on Survey Research Methods, Social Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

Interviewer Falsification of Survey Data

◆ Alan Price, Office of Research Integrity; ◆ Craig Hill, RTI International; ◆ John Thompson, NORC, University of Chicago

NORC, University of Chicago, 55 E. Monroe St., Suite 4800, Chicago, IL 60611

Thompson-John@norcmail.uchicago.edu

Key Words: *curbstoning, interviewer error, research integrity*

Interviewer falsification has been a topic discussed since the early days of survey research, and there is a small research literature on detection, diagnosis, and remediation because of interviewer falsification. The topic is receiving renewed attention because of a recent ruling by the Office of Research Integrity of the U.S. Department of Health and Human Services. ORI has statutory authority to assure the integrity of scientific activities funded by the Public Health Service. The office has recently deemed interviewer falsification of data as an act of scientific misconduct. This panel will discuss ORI's ruling and a subsequent "summit" that brought together researchers from a number of major survey organizations. The participants at the summit developed a set of guidelines about the detection, diagnosis, and remediation of data falsified by interviewers. "Detection" means the identification of completed interviews that do not contain the actual responses of sample persons, but of someone else. "Diagnosis" means the judgment of how the responses were obtained. "Remediation" means the replacement of the falsified responses with an "acceptable" response.

301 Innovative Ideas for the Statistics Classroom

Section on Statistical Education

Wednesday, August 11, 8:30 am–10:20 am

Innovative Ideas for the Statistics Classroom

◆ Joy Jordan, Lawrence University; ◆ Shonda Kuiper, Grinnell College; ◆ John Holcomb, Cleveland State University; ◆ Julie Legler, St. Olaf College

legler@stolaf.edu

Key Words: *activities, consulting, projects, statistics classroom, writing*

This panel discusses a variety of ideas for the statistics classroom. Joy Jordan presents new writing assignments for an introductory statistics course and discusses the implementation of writing in her classroom. Shonda Kuiper introduces projects (used in an introductory probability and statistics course) that emphasize the relevance of statistics, problem formulation, analysis, and clear communication of the results. Julie Legler describes St. Olaf's January-term Math Practicum, where three teams of five students each work for a month on a real-world problem and present their results to scientists and executives of the company or agency that posed the problem. Finally, John Holcomb presents a statistics project based on messy data. Time is left for floor discussion about the current and potential use of writing, activities, consulting, and projects in statistics courses.

302 Incomplete and Mismeasured Data in Studies of Disease

WNAR, Biometrics Section, Section on Statistics in Epidemiology

Wednesday, August 11, 8:30 am–10:20 am

Left-, Right-, and Interval-bivariate-censored Data: Evaluating Screening Mammography in the Prescence of Lead-time Bias, Length Bias, and Over-detection

◆ Jonathan D. Mahnken, University of Kansas Medical Center; Wenyaw Chan, University of Texas; Daniel Freeman, University of Texas Medical Branch; Jean Freeman, University of Texas Medical Branch

University of Kansas Medical Center, 3901 Rainbow Blvd., Mailstop 1008, Kansas City, KS 66160

jmahnken@kumc.edu

Key Words: *bivariate survival, gamma frailty model, censored data, Monte Carlo integration, screening, mammography*

Of the large clinical trials evaluating screening mammography efficacy, none included women ages 75 and older. Recommendations on an upper age limit at which to discontinue screening are based on indirect evidence and are not consistent. Screening mammography was evaluated using observational data from the SEER-Medicare linked database. Measuring the benefit of screening mammography is difficult due to the impact of lead-time bias, length bias, and over-detection. A new method was developed to address this problem. The underlying conceptual model divided the disease into two stages; pre-clinical (T0) and symptomatic (T1) breast cancer. Treating time in these phases as a pair of dependent bivariate observations, (t0,t1), estimates were derived to describe the distribution of this random vector. Censored information was attained about (t0,t1) from each observation. After adding mild assumptions, the likelihood function was approximated using Monte Carlo integration and used to find MLEs and an estimate of the variance-covariance matrix. This method was also studied using simulations to ensure valid parameter estimates are attainable using this new approach.

Statistical Inference of the Lead Time in Periodic Screening

◆ Dongfeng Wu, Mississippi State University; Gary L. Rosner, University of Texas M.D. Anderson Cancer Center; Lyle Broemeling, University of Texas MD Anderson Cancer Center

Mississippi State University, Dept. of Math and Statistics, Mississippi State, MS 39759

dwu@math.msstate.edu

Key Words: *periodic screening, lead time, early detection, breast cancer, sojourn time, sensitivity*

Periodic screening examinations for aggressive disease have been existed for many years, one of the most important characteristics is lead time, which is the length of time the diagnosis is advanced by screening. The effectiveness of the periodic screening program is directly related to the lead time. We developed a probability model for periodic screening, derived the probability distribution function for the lead time, which is a mixture of a point mass and a continuous distribution. The reason is that the lead time is zero for the interval incident cases. Simulation studies are carried out using the HIP data. It is shown that the proportion of breast cancer patients who truly benefit from the periodic screening varies from 57% to 80% depends on the screening intervals. Though the study is focused on female breast cancer, the method is also applicable to other kinds of chronic disease.

Modeling Risk Reduction in Limited Distribution—Coronary Artery Calcification in the DCCT/EDIC Study

◆ Wanjie Sun, George Washington University; Patricia Cleary, George Washington University

George Washington University, 6110 Executive Blvd., Suite 750, Biostatistic Center, Rockville, MD 20878

wsun@biostat.bsc.gwu.edu

Key Words: *Tobit regression, limited distribution, left-censoring*

Limited distributions are observed in medical studies, where a large percentage of the dependent variable is censored below or above a threshold value, thus generating a mixture distribution. Ordinary least square estimates or logistic regression are inconsistent or inefficient for censored data. The Tobit model was designed for censored normal data. Tobit models the probability of being censored and if uncensored, the variability of the continuous variable simultaneously. In the DCCT/EDIC study, we used the Tobit model to perform a statistical analysis of a highly skewed Coronary Artery Calcification (CAC) dataset with 70% left-censoring of zero CAC scores. Natural log transformation was applied in order to preserve the non-negativeness of CAC scores and to conform to the normal assumption of the Tobit model. The adjusted geometric mean ratio of CAC scores between the two treatment groups was used to assess the group effect for each combination of the effect modifiers. Overall treatment effect was assessed using a likelihood ratio test.

Predicting Hearing Threshold in Nonresponsive Subjects Using a Bayesian Linear Model in the Presence of Left-censored Covariates

◆ Byron J. Gajewski, University of Kansas Medical Center; Judy Widen, University of Kansas Medical Center; Nannette Nicholson, University of Kansas Medical Center

University of Kansas Medical Center, 3901 Rainbow Blvd., Mail Stop 4043, Kansas City, KS 66160

bgajewski@kumc.edu

Key Words: *lognormal, location shift*

This paper considers the prediction of pure tone thresholds for populations with the inability to respond to usual hearing tests. This is important because early intervention of hearing problems in infants, for example, is imperative for future development and quality of life. A standard statistical method in the hearing literature is to define a level of adequate hearing threshold and utilize categorical techniques such as discriminate analysis or logistic regression to diagnose based on a reproduction of a signal sent through the ear that does not require a subject's active response. The problem is that the signal suffers from censoring due to measurement constraints. Thus, the current methods deflate the leverage in the prediction model. We remedy this shortcoming by modeling the independent variables with a normal distribution while accounting for the censoring. Additionally, expanding from a previous hearing threshold model, we provide the audiologist a prediction interval of the actual pure tone threshold rather than a dichotomous prediction of hearing deficiency.

Estimation in the Presence of Nondetectables and Mixed Distributions: A Maximum-likelihood Approach with Left Censoring

◆ Robert E. Johnson, Virginia Commonwealth University; Heather J. Hoffman, Virginia Commonwealth University

Virginia Commonwealth University, Dept. of Biostatistics, Box 980032, 1101 East Marshall St., Richmond, VA 23298-0032

rjohnson@vcu.edu

Key Words: *nondetectable, censored values, zero-inflated distribution*

Methods of estimation for positive-valued distributions—e.g., assays—are well known. When the measures are left-censored, as is the case when values fall below the level of quantitation, estimation requires special techniques. Maximum-likelihood-based methods are used in survival and reliability analysis where the censored values are only known to occur within an interval. For example, subjects who self-report as nonsmokers may have trace amounts of nicotine that are nondetectable or fall below the lower limit of quantitation. In fact, true nonsmokers may have zero-valued levels while others have positive levels due to environmental tobacco smoke ETS exposure. Collecting self-reported information about ETS may not accurately distinguish between these two groups. These data follow a mixture of distributions. Techniques exist for estimation in the presence of zero-inflated data such as zero-inflated Poisson and zero-inflated log-normal. Extensions of these techniques which account for left-censored observations and for more complex mixtures will be discussed. These methods will be compared to imputing the censored values with zero, the lower limit of quantitation, or other values.

303 Longitudinal and Spatial Data Analysis

Section on Statistical Computing

Wednesday, August 11, 8:30 am–10:20 am

Spatial Modeling of fMRI Data Using non-Euclidian Distances

◆ Yulia R. Gel, George Washington University; Rajesh R. Nandy, University of Washington; Sudip Bose, George Washington University; Dietmar Cordes, University of Washington

George Washington University, Dept. of Statistics, 2201 G St., NW, Washington, DC 20052

ygl@gwu.edu

Key Words: *spatio-temporal modeling, fMRI data, non-Euclidian metrics, spatial interpolation*

Our objective is to adapt spatio-temporal methods of geostatistics to human-brain-mapping in order to study spatio-temporal variability of fMRI data. Our current focus is on the spatial pattern of fMRI data. First, we apply a segmentation-based approach to isolate gray matter, white matter, and CSF, which makes each individual segment more homogeneous. Therefore, the assumption of stationarity becomes more plausible. However, segmentation has an essential disadvantage of creating holes in the segmented data, which are most visible in gray matter. Thus the obtained data are not anymore simply connected. Therefore we suggest to use non-Euclidian metrics, e.g., the Manhattan distance. Our main interest is extension of semivariance analysis and kriging methods to data with non-Euclidian distances with emphasis on fMRI data.

Relative Risk Survival Trees with Time-dependent Covariates Using TARGET

◆ Guangzhe Fan, University of Alabama; J. Brian Gray, University of Alabama

University of Alabama, 1120 8th St., Apt. 9, Tuscaloosa, AL 35401

gfan@cba.ua.edu

Key Words: *genetic algorithm, counting process, changepoint, stochastic search, maximum likelihood*

We propose a new tree-structured model for relative risk (Cox) survival analysis with time-dependent covariates. There are two major contributions in this paper: (1) The TARGET approach, a tree-structured genetic algorithm, is used to construct the survival tree models. It finds trees with smaller size and higher accuracy than traditional greedy search methods. (2) We use a novel way to handle time-dependent covariates based on the likelihood function but without any parametric assumption. Our method can also be easily generalized to the martingale residual survival tree model. Publicly available datasets are used to illustrate the improvement of TARGET methods on traditional models.

Piecewise Constant Estimation for Prediction of Survival Outcomes: Applications in Genomics

◆ Annette Molinaro, University of California, Berkeley; Mark van der Laan, University of California, Berkeley

University of California, Berkeley, 908 the Alameda #8, Berkeley, CA 94707

molinaro@stat.berkeley.edu

Key Words: *prediction, survival analysis, model selection, regression trees, loss function, comparative genomic hybridization*

Clinicians and researchers collect a tremendous amount of data on cancer patients in the hopes of finding significant prognostic factors. Medical studies commonly involve thousands of clinical, epidemiological, and genomic measurements collected on each patient, along with a time to the clinical event of interest, such as disease recurrence or death. Over the past several decades there have been numerous attempts to use nonparametric methods with this type of data to find an estimator of outcome. A common approach is to modify classification and regression trees (CART), specifically for right-censored data. This presentation includes a generalization of CART based on a unified strategy for estimator construction, selection, and performance assessment in the presence of censoring. In this approach, the parameter of interest is defined as the risk minimizer for a suitable loss function and candidate estimators are generated with CART using this loss function. Cross-validation is applied to select an optimal estimator among the candidates and to assess the overall performance of the resulting estimator.

Comparison of Two SAS Procedures for Longitudinal Data with Evaluation of SAS Experimental Procedure Multiple Imputations

◆ Zoran Bursac, University of Arkansas for Medical Sciences; Barbara Neas, University of Oklahoma Health Sciences Center

University of Arkansas for Medical Sciences, 4301 W. Markham St., Slot 820, Little Rock, AR 72205

zbursac@uams.edu

Key Words: *incomplete data analysis, imputation methods, longitudinal data*

We compared the performance of SAS procedures MIXED and GENMOD in fitting multiple-population models to longitudinal growth data with missing values. Comparison was done under 40 simulated conditions that included two different percentages of missing data, high and medium mixed correlation structures, four and seven repeated time points, and five levels of data completeness. Completeness levels included complete data, data missing at random (MAR), MAR with imputed values, nonignorable missing data mechanism (NI) and NI with imputed values. Data imputation was performed using SAS experimental procedure MI. We found that PROC MIXED performed with as good or higher accuracy than PROC GENMOD under more than 90% of experimental conditions and that multiple data imputations helped improve the model components only for NI missing data mechanism.

Evaluating Risk of HIV Acquisition from a Cohort of IDUs Using Decision Tree Hazard Ratio Criterion

◆ Hicham Al-Nachawati, Université de Montréal; Julie Bruneau, Université de Montréal; Jean Vincelette, Université de Montréal; François Lamothe, Université de Montréal

Université de Montréal, Centre de Recherche du Centre hospitalier, 1058 Saint-Denis, Montréal, PQ H2X 3J4 Canada

hicham.al-nachawati@UMontreal.CA

Key Words: *decision tree, CHAID, survival analysis, Cox regression, HIV, injection drug users*

Decision tree (Dtree) or automatic interaction detection (AID) has been defined as a method of partitioning a set E, which is successively divided using explicative variables (risk factors, predictors) and referring to dependent variable (response, outcome). The purpose of this paper is to classify data from a cohort of Injection Drug Users (IDUs) in order to identify their risk of infection, behavioral patterns and factors associated with HIV seroconversion. Our results suggest that the Survival Dtree method using Hazard Ratio offers an alternative to standard regression models in order to classify subgroups of IDUs in relation to their risk of HIV acquisition.

Statistical Computing in Neutrino Signal Detection

◆ Fan Lu, University of Wisconsin, Madison; Grace Wahba, University of Wisconsin, Madison; Gary Hill, University of Wisconsin, Madison; Paolo Desiati, University of Wisconsin, Madison

University of Wisconsin, Madison, 1210 W. Dayton St., Dept. of Statistics, Madison, WI 53706

flu@wisc.edu

Key Words: *penalized log-likelihood, nonstandard support vector machine, Kullback-Leiber distance, logit function, optimization, neutrino signal*

Physicists are trying to use the giant device called AMANDA (Antarctic Muon and Neutrino Detector Array) buried deep in the Antarctic ice cap to detect certain neutrino signals within comparatively overwhelming background noise. Distributions of signal and background are generated by an importance sampling procedure that generates events described by multiple feature variables, computable from AMANDA data. Each event is labeled with an importance sampling weight for signal and one for background. The task is to find the most powerful decision boundary at certain significance level to distinguish signal neutrino from background neutrino. Because of the curse of dimensionality, usual Monte Carlo methods are not practical. We first propose a modified penalized log-likelihood approach to estimate the logit function in this scenario, which involves two major optimization steps and the use of KL (Kullback-Leibler) distance criterion for model tuning. Then we may compare this approach with a nonstandard SVM (support vector machine) approach. Simulation studies are presented for both approaches before we show the final results for the neutrino data.

State-space Representation for Smoothing Spline ANOVA Models

◆ Li Qin, University of Pennsylvania

University of Pennsylvania, Dept. of Biostatistics and Epidemiology, 501 Blockley Hall, 423 Guardian Dr., Philadelphia, PA 19104

lqin@cceb.upenn.edu

Key Words: *high-dimensional smoothing, Kalman Filter, smoothing spline, smoothing spline ANOVA, state-space model, tensor product*

Smoothing Spline ANOVA (SS-ANOVA) model is an important tool to estimate a multivariate smooth function. The heavy computational demand prevents its applications to high-dimensional problems and large datasets. We propose a state-space representation for the SS-ANOVA. A fast algorithm for multivariate state-space model is adapted for the SS-ANOVA. Representing a SS-ANOVA model in the state space form not only leads to a computationally efficient algorithm for estimation, but also enables the algorithm to be implemented in an online setting which is of particular importance to our real data application. For simplicity, we present our approach for the two-dimensional setting as the extension to higher dimension is straightforward. A numerical example and an application to the EEG data are used as illustration.

304 Topics in Quality ▲

Section on Quality and Productivity

Wednesday, August 11, 8:30 am–10:20 am

Defects Modeling and Yield Forecasting in Semiconductor Manufacturing

◆ Michael Baron, University of Texas, Dallas; Asya Takken, IBM Microelectronics; Emmanuel Yashchin, IBM Research; Mary Wisniewski, IBM Research Division

University of Texas, Dallas, Dept. of Mathematical Sciences, IBM Research Division, Richardson, TX 75083-0688

mbaron@utdallas.edu

Key Words: *EM algorithm, maximum likelihood, missing data*

Presented model reflects the cause-and-effect connection between observed defects on manufactured chips and their failures. Many factors such as defect type, size, frequency, and location play the role of covariates for yield prediction. Estimation of a large number of parameters is complicated by a substantial portion of uninspected layers and unclassified defects. Special model calibration techniques are proposed. Results of this analysis are used for designing optimal yield-enhancement strategies.

Statistical Detection and Estimation of Differential Item Functioning in Computerized Adaptive

◆ Xin Feng, CTB McGraw-Hill; Zhiliang Ying, Columbia University

CTB McGraw-Hill, 20 Ryan Ranch Rd., Monterey, CA 93940

ajax_feng@hotmail.com

Key Words: *measurement error, item-response theory, computerized adaptive testing, differential item functioning*

WITHDRAWN

Differential item functioning (DIF) is an important issue in large scale standardized testing. DIF refers to the unexpected difference in item performances among groups of equally proficient examinees. Its presence could seriously affect the validity of inferences drawn from a test. We address DIF analysis in the context of computerized adaptive testing (CAT). In a CAT, DIF item may be more consequential and more deteriorative. We propose simultaneous implementations of online calibration and DIF testing. Under any specific parametric IRT model, we can use the (online) estimated latent traits as covariates and fit a nonlinear regression model to each of the two groups. Because of the use of the estimated, not the true ability, the regression fit has to adjust for the covariate "measurement errors." We develop two bias-correction methods using asymptotic expansion and conditional score theory. After correcting the bias caused by measurement error, one can perform a significance test to detect DIF with the parameter estimates for different groups. Extensive simulation studies show that the resulting methods perform well.

The Normality of Flatness Measurements and The Best of k Vendors Based on CPU

◆ Abdelaziz Berrado, Arizona State University; Norma F. Hubele, Arizona State University; Esma S. Gel, Arizona State University

Arizona State University, Dept. of Industrial Engineering, PO Box 875906, Tempe, AZ 85287-5906

abdelaziz.berrado@asu.edu

Key Words: *capability index, vendor selection, coefficient of variation*

Flatness measurements are observations recorded on a quality characteristic with lower specification limit and ideal value of zero and a non-zero, positive upper specification limit. Based on the definition and operational definition of flatness in ANSI, we would anticipate that the underlying distribution of flatness measurements would be positively skewed. The analysis and fitting of 20 industrial datasets from four different processes and three different materials provide evidence that the underlying distribution is, in fact, the symmetric Gaussian distribution. As a result, the capability index CPU, which is merely a linear function of the inverse coefficient of variation, and all the associated theory supporting hypothesis testing is available. A general procedure for testing the equality of the CPUs of k independent samples is suggested.

Alternative Attri-var Quality Control Schemes

◆ Malik B. Malik, University of Maryland, Eastern Shore

University of Maryland, Eastern Shore, 1108 Kiah Hall, Princess Anne, MD 21853

mbmalik@umes.edu

Key Words: *by-attribute and by-variables schemes, inspection sampling plans, sample size performance, matching of plans*

Inspection sampling plans are dominated by the by-attributes, by-variables or attri-var mixed schemes. There are many good properties of inspection plans where these schemes show different levels of satisfying these properties. There are two good properties where attributes and variables schemes behave differently; namely small sample size and high robustness. The by-attribute plans are better in robustness but have a larger sample size than their equivalent by-variable plans and vice versa. We provide new schemes of inspection sampling plans. The levels of sample size and robustness of the new schemes lie in between the two classic by-variable and by-attribute plans. The new schemes will be designed, explained and their equivalence to the present orthodox schemes established. The sample size performance is thoroughly investigated and compared for all old and new equivalent schemes. Their robustness will be discussed at a preliminary level.

Economic Design of Control Charts with Sampling Delay

◆ Christopher A. Carolan, East Carolina University

East Carolina University, 2707-B Richard Dr., Greenville, NC 27834

carolanc@mail.ecu.edu

Key Words: *cost equation, cost-efficient, statistical design, quality*

Many authors have modeled sampling delay in development of a cost equation used to aid in the determination of an optimal (minimal cost) setting for statistical process control. However, these models fail to account for all of the consequences on the cost equation resulting from a sampling delay. A cost equation is therefore derived which accurately reflects the implications of sampling delay with regards to statistical process control. From this equation, optimal values of the process setting can be obtained.

Double Window Acceptance Sampling

◆ Morteza Ebneshahrashoob, California State University, Long Beach; Tangan Gao, California State University, Long Beach; Milton Sobel, University of California, Santa Barbara

California State University, Long Beach, Dept. of Mathematics and Statistics, 1250 Bellflower Blvd., Long Beach, CA 90840

mortezae@csulb.edu

Key Words: *acceptance sampling, sequential method, double window procedure*

We introduce the concept of a sequential double window procedure and compare the resulting acceptance test with both a fixed sample size method as well as with the Wald sequential probability ratio test.

305 Dose Studies ⌘

Biopharmaceutical Section, SSC, Biometrics Section

Wednesday, August 11, 8:30 am–10:20 am

A Group Sequential Approach to Evaluation of Bridging Studies

◆ Chin-Fu Hsiao, National Health Research Institutes

National Health Research Institutes, 128 Yen-Chiu-Yuan Rd. Sec. 2, Taipei, 115 Taiwan, ROC

chinfu@nhri.org.tw

Key Words: *bridging study, extrapolation, group sequential analysis, similarity*

The ICH E5 guideline defines a bridging study as a supplementary study conducted in the new region to provide pharmacodynamic or clinical data on efficacy, safety, dosage, and dose regimen to allow extrapolation of the foreign clinical data to the population of the new region. Therefore, a bridging study is usually conducted in the new region only after the test product has been approved for commercial marketing in the original region due to its proven efficacy and safety. The issue of analysis of clinical data generated by the bridging study conducted in the new region to evaluate the similarity for extrapolation of the foreign clinical data to the population of the new region is the information on efficacy, safety, dosage, and dose regimen of the original region cannot be concurrently obtained from the local bridging studies but available in the trials conducted in the original region. A group sequential approach is therefore proposed to overcome the issue of internal validity. In particular, we use the region as group sequence to enroll the patients from the original region first and then to enroll patients from the new region subsequently.

Comparison of Different Analyses of Data from Sequential Designs with Applications in Drug Interaction Studies

◆ Yonghua Wang, Bristol-Myers Squibb Company; Robert A. Smith, Bristol-Myers Squibb Company; Jingdong Xie, Bristol-Myers Squibb Company

Bristol-Myers Squibb Company, PO Box 4000, Princeton, NJ 08543-4000

yonghua.wang@bms.com

Key Words: *sequential design, drug interaction, mixed-effect model, standard deviation*

The issues on how to analyze data from sequential designs arise from drug interaction studies. Assuming that Drug X has interaction with Drug A in the form of reducing its bioavailability, the interest is to find a way of co-administering both drugs without sacrificing the bioavailability of Drug A. This can be accomplished by increasing the dose level of Drug A, or adding pro-drug for Drug A, etc. One design can be as follows: all subjects will take Drug A alone (Treatment A) for the first period, then these subjects will be randomized into different co-administration treatments (Treatments B, C, D, etc.). Bioavailability of Drug A from Treatments B, C, D, etc. will be compared to that of Treatment A. Our research is mainly focused on the comparison of the following analyses: (1) mixed effect model only including subjects who have data from both treatments (Drug A alone and one of co-administration treatments) in a comparison; (2) mixed effect model with sequential effect (sequences defined as AB, AC, etc.) in the model, including all subjects; and (3) mixed effect model without sequential effect in the model, including all subjects.

A Model-based Approach in the Estimation of the Maximum Tolerated Dose in Phase I Cancer Clinical Trials

◆ Weili He, Merck & Co., Inc.; Jun Liu, Rutgers University; Bruce Binkowitz, Merck & Co., Inc.; Hui Quan, Merck & Co., Inc.

Merck & Co., Inc., 1129 Blvd., Westfield, NJ 07090

weili_he@merck.com

Key Words: *maximum tolerated dose, dose-limiting toxicity, model-based approach, traditional algorithm-based designs, traditional 3+3 designs*

The primary aim of a phase I cancer clinical trial is to determine the maximum tolerated dose (MTD) of a new agent. The MTD is determined at the highest dose level of a therapeutic agent in which the patients have experienced an acceptable level of dose-limiting toxicity (DLT). The standard methods for conducting phase I cancer trials has been the traditional algorithm-based designs. These designs begin with the selection of a starting dose based on animal studies, and the dose escalation usually follows a modified Fibonacci scheme. Due to practical simplicity, the algorithm-based designs are still widely used. Most often, the traditional “3+3” design is used. Our simulation studies show that the “3+3” designs cannot provide reasonable estimates of the MTD due to their intrinsic design limitations. We propose a model-based approach in the estimation of the MTD following a “3+3” design.

Our simulations found that our approach produces less biased estimates of the MTD than the estimates obtained from the "3+3" designs. We conclude that our proposed method provides much improved estimates of the MTD. Further, our model can be easily modified to extend to any traditional A+B designs.

Experimental Design for Combination Studies

◆ Ming T. Tan, University of Maryland and Medicine; Hongbin Fang, University of Maryland and Medicine; Guoliang Tian, University of Maryland and Medicine

University of Maryland and Medicine, Greenbaum Cancer Center, Division of Biostatistics, 22 South Green St., Baltimore, MD 21201

mtan@umm.edu

Drug combination is central to cancer therapy. A statistical experimental design is necessary since even the administration of precisely the same dose to virtually identical animals may result in different responses. Such data variation needs to be controlled in the experimental design and accounted for in the analysis to allow proper inference on the efficacy of combinations. However, methods currently available are derived under some restrictive assumptions. For example, Abdelbasit and Plackett (1982) proposed an optimal experimental design assuming that the dose-response relationship follows a specified linear model. Tallarida et al. (1997) derived a fixed-ratio design and used t-test to detect the simple similar action. However, in reality, we usually do not have enough information on what kind of synergistical effect on the dose-response relationship is before experiment. We first propose a novel (nonparametric) models that does not impose such strong assumption on the possible joint action. We then propose a experimental design for the joint action with nonparametric models based on the uniform measure. This design is optimal in the sense that it reduces the variability in modeling synergy while allocates the mixtures reasonably to extract maximum information on the joint action of the compounds. Based on this experimental design, a robust F-test is proposed to detect the simple similar action of two compounds. The method is illustrated with the study of the combination of two anti-cancer agents: temozolomide and irinotecan.

Comparative Dilution Bioassays: Parallel Line Approach vs. Converting through Standard Curve

◆ Charles Y. Tan, Merck & Co., Inc.

Merck & Co., Inc., PO Box 4, WP37C-305, West Point, PA 19486

charles_tan@merck.com

Key Words: *parallel line assay, bioassay, relative potency*

The biological activity or potency of a complex material is usually measured by dilution assays in which the biochemical responses of the test material are compared to those of a standard material at several dilution levels. A parallel line approach derives the relative potency from the "shift" between the two dilution-response curves. The potency of the test material is simply the relative potency times the known (or nominal) potency of the standard. Another common approach is to convert the responses of the test material through a standard curve (potency-response), adjust the estimated potencies by their dilution factors and then average. The statistical and scientific advantages of the parallel line approach will be argued. Statistical advantages will be detailed for the linear response curve case. Results will be extended to the four-parameter logistic response curve case.

Max-min Confidence Intervals for Differences between Treatments and Control for Monotone Dose-response Curves

◆ Jianan Peng, Acadia University; Charles Lee, Memorial University of Newfoundland; Lin Liu, University of California, San Diego

Acadia University, 12 University Ave., Wolfville, NS B4P 2R6 Canada

jianan.peng@acadiau.ca

Key Words: *simultaneous confidence intervals, Dunnett's procedure*

The comparison of increasing doses of a drug to a zero dose control is of interest in medical and toxicological studies. Usually the dose-response curve is monotone. A simple procedure that modifies Dunnett's procedure is proposed to construct simultaneous confidence intervals for differences between each treatment and the control by utilizing the ordering of the means. The procedure compares well with its predecessor.

Assessing the Joint Action of Multiple Drugs: Synergy, Additivity, and Antagonism

◆ Xiaoli S. Hou, Merck & Co., Inc.; Keith A. Soper, Merck & Co., Inc.

Merck & Co., Inc., Biometrics Research Lab., PO Box 4 WP 37C-305, West Point, PA 19486-0004

xiaoli_hou@merck.com

Key Words: *synergy, additive, antagonism, response additive, concentration additive, interaction*

A new "segmented response surface" approach is proposed for measurement of the joint effect of two or more drugs in vitro: synergy, additivity, or antagonism. Widely used approaches based on response surface methods either parameterize how drugs interact, resulting in poor conclusions when the true interaction is more complex than assumed, or use a nonparametric response surface method that lacks power and often is difficult to summarize. Our approach compares simple dose-response curves such as the response curve for drug "A" alone vs. at fixed combinations of drugs "B," "C," etc, vs. the dose-response predicted by "additivity." Our graphical results are easy to explain and have direct clinical interpretation. Our method easily detects when drugs may be synergistic at some doses but antagonistic at other doses. We show how to construct a single map partitioning the observed dose-space into "smooth" regions of synergy, additive, subadditive, inert, and interference that is consistent with all dose response curves regardless of which compounds are fixed and which allowed to vary. Associated statistical tests are more powerful than competing standard methods in realistic situations.

306 Nonparametric Regression

Section on Nonparametric Statistics

Wednesday, August 11, 8:30 am–10:20 am

Linear Regression in the Presence of Spatially Correlated Errors: A Computer Intensive Approach

◆ Liliana Gonzalez, University of Rhode Island; Leah Passmore, Wake Forest University

University of Rhode Island, CS and Statistics Dept., Tyler Hall, Kingston, RI 02881

liliana@cs.uri.edu

WITHDRAWN

Key Words: *block bootstrap, bootstrap, randomization, jackknife, spatial correlation, regression*

Computer intensive statistical techniques have been used for testing hypotheses concerning linear regression parameters. These methods assume that the model errors are identically and independently distributed and little is known about the effect of correlated errors on performance of these methods. Computer intensive procedures allowing for dependent data are available. In the time series setting, most of these involve dividing the series into either overlapping or nonoverlapping blocks. The blocks are then randomly resampled and joined together to produce a version of the original series. Hall (1985) proposed a similar blocking approach for spatial data. Several variants of this technique have been developed for a number of purposes but the performance of block resampling has not been evaluated in the context of linear regression analysis. This study compares several standard computer intensive procedures and spatial resampling methods (overlapping and nonoverlapping blocks) for significance testing of regression coefficients when model errors are spatially correlated. Simulation experiments were performed using several spatial covariance structures to evaluate the size and power of tests achieved using each method. Recommendations are given concerning practical application of computer intensive techniques for regression models with spatially dependent errors.

Robust Weighted LAD Regression

◆ Avi H. Giloni, Yeshiva University; Jeffrey S. Simonoff, New York University; Bhaskar Sengupta, Yeshiva University

Yeshiva University, Sy Syms School of Business, 500 West 185th St., BH-428, New York, NY 10033

agiloni@ymail.yu.edu

Key Words: *LAD regression, breakdown point, robustness, high breakdown estimator*

Consider a regression problem with a given design matrix, where there is possible contamination in the response variable values. We show that a weighted version of the least absolute deviation (LAD) regression estimator can attain a high breakdown point given the design (as high as 30%). This can be accomplished even in the presence of designs with "leverage" points. Furthermore, we show that the estimator is asymptotically normal, root n consistent, and give the asymptotic variance of the estimator. Several examples illustrating the usefulness of the estimator are provided.

Nonparametric Assessment of Dimension in a General Regression Problem

Pilar Barrios, Fundación Española de Ciencia y TecnologÌa; Santiago Velilla, Universidad Carlos III de Madrid

Universidad Carlos III de Madrid, C/Madrid, 126, Getafe (Madrid), 8903 Spain

santiago.velilla@uc3m.es

Key Words: *dimension reduction subspace, eigenvalues, nonparametric regression estimator, resampling, sliced inverse regression (SIR), smoothing parameter*

A method for assessing the dimension of a general regression problem is presented, based on a nonparametric estimator of the inverse mean subspace kernel matrix. The proposed approach is an alternative to the usual inference techniques based on a preliminary "slicing" of the values of the response variable.

An Optimal Approach of Sufficient Dimension Reduction in Multivariate Regression

◆ Jae Keun Yoo, University of Minnesota, Twin Cities

University of Minnesota, Twin Cities, 1025 29th Ave. SE #F, School of Statistics, Minneapolis, MN 55414

yjkstat@stat.umn.edu

Key Words: *dimension reduction, predictor effects, central subspaces, multivariate regressions, chi-square distribution*

Recently, Cook and Setodji (2003) developed a method for sufficient dimension reduction in multivariate regressions by estimating multivariate central mean space. Their test statistic for the dimension of the multivariate central mean subspace is a weighted sum of independent chi-square random variables. I provide an optimal version of this method in the same context. The test statistic for the optimal version has chi-square distribution. Additionally, the optimal version allows tests of predictor effects. A comparison of the two methods will be provided.

Optimal Sufficient Dimension Reduction in Regressions with Categorical Predictors

◆ Xuerong Wen, University of Minnesota

University of Minnesota, 313 Ford Hall, 224 Church St. SE, School of Statistics, Minneapolis, MN 55108

wenxr@stat.umn.edu

Key Words: *sufficient dimension reduction, categorical predictors, minimum discrepancy approach, testing coordinates hypotheses*

In the context of sufficient dimension reduction, a Minimum Discrepancy Approach (MDA) is introduced for regressions with a mix of continuous and categorical predictors. It is shown that current methods are special cases of this approach and an optimal method, partial inverse regression estimation (PIRE) is proposed. PIRE provides an optimal estimate of the partial central subspace (PCS). An asymptotic chi-squared distribution for rank test is given. A chi-squared test of the conditional independence hypothesis that the response is independent of a selected subset of quantitative predictors given the remaining predictors is obtained.

307 **Design Issues** ▲ ⌘

Biometrics Section, ENAR, Biopharmaceutical Section

Wednesday, August 11, 8:30 am–10:20 am

Goodness of Fit of Relative Survival Models

◆ Janez Stare, Univerza v Ljubljani, Slovenija; Maja Pohar, Univerza v Ljubljani, Slovenija; Robin Henderson, University of Lancaster

Univerza v Ljubljani, Slovenija, Medicinska fakulteta, Vrazov trg 2, Dept. of Biomedical Informatics, Ljubljana, 1000 Slovenia

janez.stare@mf.uni-lj.si

Key Words: *relative survival, regression models, goodness of fit*

Additive regression models in relative survival studies are preferred over multiplicative models by some authors. Such preferences are mainly grounded in practical experience with mostly cancer registries data, where the basic assumption of the additivity of hazards is more likely to be met. Additive models are by definition not suited for data containing subsets of subjects living longer than the general population. Multiplicative models have no such limitations. We study goodness of fit of models based on a recently introduced individual measure of relative survival. We show that such models are more flexible than standard additive models and will satisfactorily fit very different kinds of data.

Rank-based Estimate of Four-parameter Logistic Model with Application in Pharmaceutical Sciences

◆ Kimberly Crimin, Pfizer Inc.; Thomas Vidmar, Pfizer Inc.; Joseph McKean, Western Michigan University

Pfizer Inc., Eastern Point Rd., MS8260-2313, Groton, CT 06340

kimberly.crimin@pfizer.com

Key Words: *rank-based estimate, IC50, four-parameter logistic*

In the pharmaceutical industry, the calculation of inhibitory concentration that results in a response of 50% (IC50), is performed thousands of times every day. The nonlinear model most often used to performed this calculation is a four-parameter logistic, suitably parameterized to estimate the IC50 directly. When performing these calculations in a high-throughput mode, each and every curve cannot be studied in detail. Outliers in the responses are a common problem. A robust estimation procedure to perform this calculation is desirable. We are proposing a rank-based estimate of the four-parameter logistic model that is analogous to least squares, except a norm other than the Euclidean norm is used. We illustrate this robust procedure with several examples from pharmaceutical industry. When no outliers are present in the data, the robust estimate of IC50 is comparable to the least squares estimate and when outliers are present in the data, the robust estimates is more accurate. We also provide a robust goodness-of-fit test and present the results from a small simulation study.

On Selecting among Treatments with Binomial Outcomes

◆ Elena M. Buzaianu, Syracuse University; Pinyuen Chen, Syracuse University

Syracuse University, 215 Carnegie Hall, Syracuse, NY 13210

mbuzaianu@yahoo.com

Key Words: *selection, binomial, least favorable configuration*

Most of the procedures for selecting one of the best among several treatments under the binomial settings are based on applying the normal theory to the discrete data. Consequently, such procedures may not be safely applied when dealing with small sample sizes. This paper considers the LFCs for the exact versions of those procedures. That is, procedures that are based on the binomial distributions instead of their approximations. We begin with the procedure proposed by Sobel and Hyuett, which does not involve a control. Then we consider the single stage procedure by Dunnet and a two-stage procedure by Thall, Simmon, and Ellenberg, where a control is involved. With our results, the redefined procedures can be applied to any sample size.

Comparison of Analyses for Repeated Measurements from Ophthalmic Data

◆ Kuolung Hu, Eli Lilly and Company; Vipin K. Arora, Eli Lilly and Company; David Manner, Southwestern Medical School

Eli Lilly and Company, LCC, Indianapolis, IN 46285

hu_kuolung@lilly.com

Key Words: *generalized estimating equations (GEE), survival analysis, ophthalmology, clinical studies*

This work considers two different model-based approaches for analyzing repeated binary data from longitudinal ophthalmology clinical studies. The two statistical approaches considered are: nonlikelihood (GEE) regression and partial-likelihood model built from the simulated survival times. They are compared in an analysis of the binary response outcome (which can reverse during the study). The ophthalmic data are simulated for a two treatment arm (treated and nontreated group) study with persons with diabetes. The primary endpoint is progression to center-involved Diabetic Macular Edema (DME) (Yes/No) and is observed over time (maximum exposure 36 months with 12 time points including baseline). The probability of dropouts and reversal (of DME outcome) of events are considered in several simulation scenarios. By focusing on the efficiency of study design, we will demonstrate that GEE is superior to the partial-likelihood model in detecting the difference between the two treatment arms. Furthermore, we will also demonstrate that GEE performs asymptotically as good as the partial-likelihood model under the assumption of no reversals.

On Semiparametric Transformation Cure Models

◆ Wenbin Lu, North Carolina State University; Zhiliang Ying, Columbia University

North Carolina State University, 2501 Founder's Dr., 210E Patterson Hall, Raleigh, NC 27695

lu@stat.ncsu.edu

Key Words: *cure rate models, estimating equations, long-term survivors, martingale, transformation models*

Survival models with a cure rate have received much attention in recent years. These models are useful when a proportion of study subjects never experience the event of interest. Such examples can be found in many disciplines, including biomedical sciences, economics, sociology, and engineering science. A general class of semiparametric transformation cure models will be studied for the analysis of survival data with long-term survivors. It combines a logistic regression for the probability of event occurrence with the class of transformation models for the time of occurrence. Included as special cases are the proportional hazards cure model and the proportional odds cure model. Estimating equations are proposed for parameter estimation. It is shown that the resulting estimators are asymptotically normal, with variance-covariance matrix that has a closed form and can be consistently estimated by the usual plug-in method. Simulation studies show that the proposed approach is appropriate for practical use. An application to data from a breast cancer study is given to illustrate the methodology.

Drug-drug Interaction Study of an Antimigraine Agent and Erythromycin

◆ Meiyu Shen, U.S. Food and Drug Administration; Stella G. Machado, U.S. Food and Drug Administration

shenm@cder.fda.gov

Key Words: *antimigraine, PK-PD modeling, drug-drug interaction*

Two studies were used to evaluate the potential elevation of diastolic blood pressure due to the administration of an antimigraine agent and erythromycin. The pharmacokinetic-pharmacodynamic (PK/PD) study of the antimigraine agent alone was a four-period crossover study. Fifteen subjects randomly received placebo plus the antimigraine agent at three of the following doses: 30, 60, 90, and 120 mg. The drug-drug interaction study was also a crossover study, 18 subjects randomly assigned to two groups, the first receiving 80 mg of the antimigraine agent, the second receiving 80 mg of the antimigraine agent plus erythromycin. After a washout period, both groups were then given the alternative treatment. The blood plasma concentration of the antimigraine agent and the diastolic blood pressure (DBP) were simultaneously measured at predetermined times. The pharmacodynamic measurement of interest for each patient was the maximum DPB change from baseline. The change in PK/PD of the antimigraine agent due to coadministration with erythromycin was assessed using general linear models. Plasma concentration of the antimigraine agent was elevated when co-administered with erythromycin.

A Stopping Plan for a Bioequivalence Study

◆ Shu-Ping Lan, Pfizer Inc.

Pfizer Inc., 50 Pequot Ave., MS 6025-A4269, New London, 06320

shirley_lan@groton.pfizer.com

Key Words: *adaptive design, group sequential methods, bioequivalence study, stopping rules*

In designing bioequivalence studies, if the possibility of early termination to claim success is desirable, interim analyses based on group sequential methods can be planned and pre-specified in the protocol. Reasons for the consideration of early termination include unstable estimate of the variance, larger-than-usual sample size required, difficulty in recruitment, ethical reasons, etc. In addition, at the interim analyses, considerations for stopping the trial due to futility can be evaluated based on conditional power calculations. Stopping the trial for futility is at the sponsor's risk and hence need not be pre-specified in the protocol. An example is given for a stopping plan in designing a bioequivalence study, which uses the extension of group sequential methods and conditional power calculations. The sequential boundaries were transformed into "sequential confidence levels," which can be used as the criteria for boundary-crossing in the interim analyses. We also extended the concepts of conditional power to provide evaluations of stopping due to futility.

308 Time Series Analysis and Forecasting ◬

Business and Economics Statistics Section

Wednesday, August 11, 8:30 am–10:20 am

Revenue Management with Correlated Demand Forecasting and Multistage Stochastic Programming

◆ Catalina Stefanescu, London Business School; Kristin Fridgeirsdottir, London Business School; Victor de Miguel, London Business School; Stefanos Zenios, Stanford University

London Business School, Regent's Park, London, International NW1 4SA UK

cstefanescu@london.edu

Key Words: *revenue management, forecasting, EM algorithm, multistage stochastic programming*

A popular approach to airline revenue management is to assign a bid price to each resource in the network. Bid prices are estimates of the expected future revenues, generated in a two-step process. First, a demand forecasting model is estimated from historical demand data. Second, this model is used to generate a scenario tree of future travel demand, and the expected future revenue is maximized with an optimization algorithm. This process is repeated daily to compute bid prices and make booking decisions. We propose a revenue management methodology that has two innovative features. First, we develop a forecasting model that takes into account the correlation between the demand for different products and also the correlation between demand during different time periods. The EM algorithm is used to estimate the model, in order to incorporate censored demand data. Second, we show how to determine bid prices using an optimization algorithm based on a multistage stochastic program. Simulations studies and an application to a real airline booking dataset show the improved performance of our methodology.

El Paso Customs District Cross-border Trade Flows

◆ Roberto Tinajero, University of Texas, El Paso; Thomas M. Fullerton, Jr., University of Texas, El Paso; Richard L. Sprinkle, University of Texas, El Paso

University of Texas, El Paso, Dept. of Economics & Finance, 500 W. University Ave., El Paso, TX 79968

rtinajero@utep.edu

Key Words: *international econometrics, border trade flows, transfer function ARIMA models*

An important issue in applied international economics is the extent to which trade flows adjust to changes in economic activity, relative prices, and exchange rates. While there have been numerous studies regarding merchandise trade elasticities for industrial economies such as the European Union, Japan, and the United States, relatively little empirical work has been completed with respect to analyzing international merchandise trade growth through the United States-Mexico border region. To examine whether border trade flows can be successfully modeled, short-term time series characteristics of cross-border trade flows through El Paso, Texas, are analyzed. A transfer function ARIMA econometric methodology is employed using data from January 1995 to December 2002 to model border trade flows. Empirical results indicate that economic activity in the United States and Mexico along with relative prices adjusted for exchange rate changes play important roles in determining month-to-month fluctuations in border region trade flows.

Forecasting Models Applied to Data Updates Over Time Periods

◆ Martin B. Selzer, National Exchange Carrier Association

National Exchange Carrier Association, 80 South Jefferson Rd., Whippany, New Jersey, NJ 07981

mselzer@neca.org

Key Words: *forecasting, nonlinear models, time-series*

Certain businesses require forecasting updates from an initial value. The updates to a data point can be predicted using models if historical data of such updates is available. At any particular time during the update period, the models give a final prediction of the data point. The National Exchange Carrier Association makes forecasts under such conditions for rural telecommunications companies making updates to revenue and cost data submitted monthly. The current research uses models that assume a first observation fixed and estimate a trajectory to a final value using nonlinear methods. It applies these models to forecasting revenues and costs and compares results to predictions currently made by expert managers.

Absolute and Relative Measures for Evaluating the Forecasting Performance of Time-series Models

◆ Tessema Astatkie, Nova Scotia Agricultural College

Nova Scotia Agricultural College, Dept. of Engineering, PO Box 550, Truro, NS B2N 5E3 Canada

tastatkie@nsac.ns.ca

Key Words: *nonlinear time series, model validation, coefficient of efficiency, index of agreement*

Mean squared error (MSE) and mean absolute error (MAE) of prediction are widely used measures for evaluating the forecasting performance of time-series models. Although these absolute measures are good to compare alternative models, they do not indicate how well the models forecast the true values. Two relative measures, namely coefficient of efficiency (E) and index of agreement (d) are discussed. Also, their modified versions, known as baseline-modified coefficient of efficiency (EM) and baseline-modified index of agreement (dM) are proposed. The values of E and EM range from minus infinity to 1, while those of d and dM range from 0 to 1. For both measures, values closer to 1 are preferred. These measures are illustrated using the benchmark Jokulsa Eystri daily River flow data. The out-of-sample forecasts of one to five days in 1974 by the NAARX, NeTAR and MNITF models, calibrated using the 1972-73 data, are compared in terms of MSE, MAE, E, EM, d, and dM. NeTAR described the system best, and gave better one- and two-day ahead forecasts. MNITF was better for three-day, all three were comparable for four-day, and NAARX and NeTAR were equally better for five-day-ahead forecasts.

Identifying the Cycle of a Macroeconomic Time Series Using Fuzzy Filtering

◆ David E. Giles, University of Victoria; Chad N. Stroomer, University of Victoria

University of Victoria, PO Box 1700, STN CSC, Dept. of Economics, Victoria, BC V8W2Y2 Canada

dgiles@uvic.ca

Key Words: *business cycle, fuzzy sets, Hodrick-Prescott filter, fuzzy logic*

This paper presents a new method for extracting the cycle from an economic time series. This method uses the fuzzy c-means clustering algorithm, drawn from the pattern recognition literature, to identify groups of observations. The time series is modeled over each of these subsamples, and the results are combined using the “degrees of membership” for each data point with each cluster. The result is a totally flexible model that readily captures complex nonlinearities in the data. This type of “fuzzy regression” analysis has been shown by Giles and Draeseke (2003) to be highly effective in a broad range of situations with economic data. The fuzzy filter that we develop here is compared with the well-known Hodrick-Prescott (HP) filter in an extensive Monte Carlo experiment, and the new filter is found to perform as well as, or better than, the HP filter over a wide range of time-series characterstics. The paper also includes some applications with real time series to illustrate the different conclusions that can emerge when the fuzzy filter and the HP filter are each applied to extract the cycle.

Tests for Duration Clustering and Diagnostic Checking of ACD Models Using Kernel-based Spectral Density Estimators

◆ Maria Pacurar, HEC Montréal; Pierre Duchesne, Université de Montréal

maria.pacurar@hec.ca

Key Words: *duration clustering, model adequacy, kernel-based spectral density, time series*

Engle and Russell's autoregressive conditional duration (ACD) models have been widely used to model financial data that arrive at irregular intervals. In the modeling of such data, testing for duration clustering and evaluation procedures of a particular model are important steps. We propose two classes of tests for ACD effects and one class for the adequacy, using kernel-based spectral density estimators. The tests of ACD effects of the first class are obtained by comparing a kernel-based normalized spectral density estimator and the normalized spectral density under the null hypothesis of no ACD effects, using a norm. The second class exploits the one-sided nature of the alternative hypothesis. Tests for the adequacy are obtained by comparing a kernel-based spectral density estimator of the standardized residuals and the null hypothesis of adequacy using a norm. Asymptotic distributions of the test statistics are obtained, which are normal. We present a simulation study illustrating the merits of the proposed procedures and an application with financial data is conducted.

Functional Coefficient Regression Models with Dependent Data

◆ Yanrong Cao, University of Toronto; Haiqun Lin, Yale University; Zhou Wu, University of Cincinnati; Yan Yu, University of Cincinnati

University of Toronto, Joseph L. Rotman School of Management, 105 St. George St., Toronto, ON M5S 3E6 Canada

yanrong.cao03@rotman.utoronto.ca

Key Words: *multivariate time series, functional coefficient regression model, dimension reduction, forecasting, consistency, asymptotics*

A penalized spline (PS) approach is proposed to estimate functional coefficient (FC) regression models for nonlinear time series (TS). FC regression models assume regression coefficients vary with certain lower dimensional covariates, providing appreciable flexibility in capturing the underlying dynamics in data and avoiding the so-called "curse of dimensionality" in multivariate nonparametric TS estimation. One of the appeals of our PS approach lies in the efficiency in estimating coefficient functions via the global smoothing method. In addition, enabled by assigning different penalties accordingly, different smoothness is allowed for different functional coefficients. Penalty terms, selected by minimizing generalized cross-validation scores (GCV), balance the goodness-of-fit and smoothness. The number and location of knots are no longer crucial if the minimum number of knots is reached. Consistency and asymptotic normality of the penalized least squares estimators are obtained. Our PS approach also enables multi-step-ahead forecasting with an explicit model expression in contrast to the local smoothing method. Both simulation examples and a real data application are demonstrated.

309 Coverage ▲

Section on Survey Research Methods

Wednesday, August 11, 8:30 am–10:20 am

Classification of Address Register Coverage Rates—A Field Study

◆ Gavin Thompson, Statistics Canada; Claude Turmelle, Statistics Canada

Statistics Canada, 120 Parkdale, Ottawa, ON K1A 0T6 Canada

gavin.thompson@statcan.ca

Key Words: *Canadian Labour Force Survey, address register, coverage rates, sample frame*

Every 10 years the Canadian Labour Force Survey (CLFS) undergoes a review of its sample design. This time the use of the Canadian Address Register (CAR) was examined to replace the costly listing process. The CLFS has a two-stage design. For each Primary Sampling Unit (PSU) selected a complete listing of all dwellings must be compiled in order to select the final sample. This list is normally compiled by an interviewer touring the entire PSU and noting each dwelling. The use of the CAR would enable us to eliminate this step in some PSUs (type 1), and in others potentially improve the quality of the lists by being used as an aide (type 2). In some PSUs the coverage rates of the CAR is too poor to be used (type 3). A field study in five cities, totaling 55,000 households, was completed in December 2003. This paper will describe the design of the field test, review the overcoverage and undercoverage rates of the CAR and present the results obtained with both methods of listing. Also, we will discuss the methods used for determining which of the three types the 8,000 CLFS selected PSUs fall into.

Building a Sampling Frame from Multiple Lists: Identifying Organizations in the Voluntary and Not-for-profit Sector in Canada

◆ Linda L. Lefebvre, Statistics Canada; Peter G. Wright, Statistics Canada

Statistics Canada, RHC 17-O, 120 Parkdale Ave., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

peter.wright@statcan.ca

Key Words: *hierarchical exact matching, record linkage*

Voluntary and not-for-profit organizations are increasingly being recognized for the important role they play in Canadian society, yet efforts to improve the capacity of these organizations to fulfill their missions are constrained by the lack of information about their characteristics. As part of an initiative by the Canadian government to study the not-for-profit and voluntary sector, Statistics Canada conducted a survey to collect baseline information on the function, size, and roles of organizations in this sector and the benefits that they work to provide. To create a sampling frame, it was necessary to resolve the many-to-many matches of organizations that appeared on source lists of not-for-profit businesses and charities at the federal and provincial levels. Hierarchical exact matching helped to eliminate multiple cases as well as to increase the information available for many of the records on the sampling frame. We will discuss the practical aspects of file preparation, record linkage and hierarchical matching. The discussion will focus primarily on creating sampling frames of smaller businesses.

Methodology for Enumerating U.S. Citizens Living Overseas

◆ Mary Frances E. Zelenak, U.S. Census Bureau; Jon R. Clark, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 7600, Washington, DC 20233-7600

mary.frances.zelenak@census.gov

Key Words: *decennial census, overseas enumeration, 2004 Census Test*

The U.S. Census Bureau is researching ways to include all U.S. citizens living abroad in the decennial census counts. Administrative counts were used in the 1990 and 2000 censuses to enumerate military and civilian federal employees and their dependents living overseas. To develop a methodology and to determine the feasibility, quality, and cost of collecting data from U.S. citizens living overseas, the Census Bureau will conduct an Overseas Enumeration operation in three test countries—France, Kuwait, and Mexico—as part of the 2004 Census Test. During this operation, Americans living in the selected countries will have the opportunity to provide information by mail or internet for inclusion in the census test counts. This paper presents the inherent challenges in, implementation methodology with preliminary results from, and plans for further assessment of the enumeration of U.S. citizens living overseas during the 2004 Census Test.

Re-delivery Operation for Undeliverable as Addressed Housing Units in Census 2000

◆ Felipe Kohn, U.S. Census Bureau

U.S. Census Bureau, Suitland Rd., Washington, DC 20233

felipe.e.kohn@census.gov

Key Words: *re-delivery operation*

The United States Postal Service was the primary vehicle for delivering census questionnaires. Questionnaires that were undeliverable were called undeliverable as addressed (UAA). The Census Bureau re-delivery operation for UAA questionnaires took place in pre-selected ZIP codes and was conducted by specially trained enumerators. These enumerators attempted to re-deliver UAAs in these ZIP codes in order to decrease the number of undeliverable housing units. This paper examines the decrease in undeliverable housing units as a result of the re-delivery operation and uses logistic regression modeling to examine relationships between UAA status and demographic data. For the 2010 Census, the Census Bureau must decide if a re-delivery operation for UAA questionnaires is cost-effective. If a re-delivery operation is used, an optimal plan for designating areas in which to conduct the re-delivery is needed. The research results of this paper will be used by the Census Bureau to address these issues for the 2010 Census.

Methodological Lessons from Census 2000 Coverage Error Measurement

◆ Mary H. Mulry, U.S. Census Bureau

U.S. Census Bureau, 2884 Manorwood Trail, Fort Worth, TX 76109

mary.h.mulry@census.gov

Key Words: *dual system estimator, overcount, nonsampling error, undercount*

This paper synthesizes the methodological lessons from the Census Bureau's program to measure coverage error in Census 2000 from the Accuracy and Coverage Evaluation Survey (A.C.E.). The Census Bureau decided to revise the original A.C.E. dual system estimates (DSEs) because of the discovery of a large number of undetected duplicate enumerations. These revised estimates, known as A.C.E. Revision II, included adjustments for census duplicates as well as measurement errors and showed a net undercount rate of -0.5%, an overcount, in the census count of 281,421,906. Since the A.C.E. and the A.C.E. Revision II estimates each had an evaluation program, there is a wealth of information about the survey design, data collection, data processing, and estimation methodologies used to obtain both sets of estimates. In addition, this information provides guidance for the planning of Census 2010 and for the planning of the program that evaluates the census coverage.

Census Coverage Measurement Design Alternatives

◆ Michael Beaghen, U.S. Census Bureau

U.S. Census Bureau, 6631 Wakefield Dr. #718, Alexandria, VA 22307

michael.a.beaghen@census.gov

Key Words: *duplication, omissions, gross error, any address match, extended search area*

The Census 2000 coverage measurement program was the Accuracy and Coverage Evaluation. This evaluation was flawed because of the ineffectiveness of its interviews in ascertaining when people had a Census Day residence other than the one at which they were being interviewed. As a consequence, the Accuracy and Coverage Evaluation failed to detect a large proportion of the census person duplication and possibly overestimated the number of census omissions. It is imperative that future census coverage measurement accurately measure census person duplication and omissions. A secondary imperative is to measure gross census error in addition to net census error. This paper gives an overview of options being explored for a future coverage measurement design. This overview includes discussion on liberal rostering and any address-matching, an extended search area, integrating an automated search for duplicates, and matching to census records historically viewed as having insufficient information for processing, in addition to discussion on research into interview instrument development and census residency rules.

A Stochastic Process Approach to Estimating Population Size

◆ Emily S. Murphree, Miami University

Miami University, Dept. of Mathematics and Statistics, Oxford, OH 45056

murphres@muohio.edu

Key Words: *population size, renewal theory, sampling with replacement, continuous-time model*

Let K be the unknown size of an animal population and suppose animals are observed at random times S1, S2,... Observation may occur either when an animal is trapped or when an animal visits a bait site and deposits a hair sample on a strand of barbed wire. In the latter case, the visit times S1, S2,... may not be known. The biologist will know W(t), the number of distinct animals observed by time t. Depending on how much other information is available, we propose several different renewal theoretic approaches to estimating K. In an idealized model where animals share the same capture probability and times between observations are exponential, a multinomial model is appropriate. An effort to make the assumptions more realistic will also be discussed.

310 Statistical Genetics Linkage Analysis ▲ ⌘

Biometrics Section, Section on Statistics in Epidemiology

Wednesday, August 11, 8:30 am–10:20 am

Exploring the Adequacy of Gaussian Approximation in the Affected Sib Pair Test

◆ Xiaoling Wu, Johns Hopkins University; Daniel Q. Naiman, Johns Hopkins University

Johns Hopkins University, Dept. of Applied Mathematics and Statistics, Baltimore, MD 21218

xlwu@jhu.edu

Key Words: *importance sampling, affected sib pair test, Gaussian approximation, inheritance vectors, multiple testing, Markov chain*

A standard approach to calculation of critical values for the affected sib pair test is based on: (a) fully informative markers, (b) Haldane map function assumptions leading to a Markov chain model for inheritance vectors, and (c) central limit approximation to averages of sampled inheritance vectors leading to a Ornstein-Uhlenbeck process approximation, (d) renewal theory approximation. Previous work has shown that for equispaced or close to equispaced markers, if the sample size is large, the FBS approximation performs well. However, for a small sample size there is good reason to be suspicious about the use of the central limit approximation. We explore the inadequacies of this central limit approximation and determine sample sizes for which the approximation performs adequately.

Generalized Haplotype Transmission Probabilities in Nuclear Families with an Affected Sib Pair

◆ Fang Xie, Chiron Corporation; Shelley Bull, Samuel Lunenfeld Research Institute; Paul Corey, University of Toronto

Chiron Corporation, 4560 Horton St., M/S U140, Emeryville, CA 94608

fang_xie@chiron.com

Key Words: *haplotype relative risks, TDT, linkage, association, sib pair*

TDT was first proposed to test for linkage in presence of an association by Spielman (1993). It was built on the early work of Falk and Rubinstein (1987) on the concept of haplotype relative risks (HRR). Assuming a single-locus model for the HRR under a recessive mode of inheritance, Ott (1989) showed how the test statistic depends on both the linkage and the association parameters for a biallelic marker. We generalize Ott's work to a general genetic model specified by penetrances and allele frequencies. We consider nuclear families with at least one parent who is heterozygous at a biallelic marker locus with one affect child. The joint distribution of transmitting and nontransmitting a particular marker allele from a heterozygous parent to his affected children is computed. The probability distribution of the TDT is verified to be a special case of this joint distribution. We further extend the model to include nuclear families with an affected sib pair, without the assumption of independent transmission from the parents to their children. A test with two degrees of freedom for linkage can be constructed using the joint haplotype distribution for this type of nuclear families.

On Genetic Linkage/Association Analysis with Incomplete Parental Information

◆ Jing Han, New York University; Yongzhao Shao, New York University

New York University, 650 1st Ave., 5th floor, New York, NY 10016

jh828@nyu.edu

Key Words: *genetic linkage analysis, missing parental genotypes, marker locus, linkage disequilibrium, genetic association test, power analysis*

We study the problem of detecting linkage/association for family-based genetic data. The proposed test is applicable without knowledge of disease penetrance, mating types, the amount of linkage disequilibrium, etc. Simulation shows that it works well even when parental marker genotype information is incomplete.

Testing for Homogeneity in Genetic Linkage Analysis

◆ Yuejiao Fu, University of Waterloo; Jiahua Chen, University of Waterloo; Jack Kalbfleisch, University of Michigan

University of Waterloo, Dept. of Statistics and Actuarial Science, Waterloo, ON N2L 3G1 Canada

y6fu@math.uwaterloo.ca

Key Words: *genetic linkage analysis, hypothesis testing, local asymptotic power, mixture models, modified likelihood, recombinant*

We apply the modified likelihood ratio test to two binomial mixture models arising in genetic linkage analysis. The limiting distribution of the test statistic for both models is shown to be a mixture of chi-squared distributions. A consideration of random family sizes for both models gives similar results. We also explore the power properties under local alternatives. Simulation studies show that the modified likelihood ratio test is more powerful than other methods under a variety of model specifications.

Incorporate Risk Factors into Haseman and Elston Nonparamatric Linkage Analysis Approach

◆ Chi-Hse Teng, Pfizer Inc.

Pfizer Inc., 439 Lexington Cir., Oceanside, CA 92057

chi-hse.teng@pfizer.com

Key Words: *linkage analysis, risk factor, gene-environment interaction*

Penrose proposed the sib-pair method in 1935. Since then, this approach has been extended and improved, such as Haseman and Elston (1972), Holmans and Clayton (1995), and Guo and Elston (2000). There were also efforts to include covariates into the analysis via perform regression between the IBD sharing probabilities and the covariates. We propose a different route to combine linkage information with risk factors information into single statistical model. We prepare the data into a case-control format that many standard statistical techniques can be easily applied to build a single model that includes the IBD sharing, risk factors, and the gene-gene, gene-environment, or environment-environment interactions. Genetic Analysis Workshpp (GAW) 13 simulated data (Problem 2) were used to demonstrate this approach. Logistic regression and survival analysis were performed to identify baseline and slope genes. Systolic blood pressure and high density Lipoprotein are selected phenotypes for the analyses. Several baseline genes, slope genes and risk factors were identified for each phenotype. The model selection criteria and building procedure need to be further improved.

Efficient Study Designs for Identifying SNPs Responsible for a Linkage Signal

◆ Mingyao Li, University of Michigan; Goncalo Abecasis, University of Michigan; Michael Boehnke, University of Michigan

University of Michigan, 1420 Washington Heights, Dept. of Biostatistics, Ann Arbor, MI 48109

myli@umich.edu

Key Words: *linkage disequilibrium, gene localization, identification of disease genes, mapping complex traits*

Once genetic linkage has been identified for a complex disease, the next step is often association analysis, in which single-nucleotide polymorphisms (SNPs) within the linkage region are genotyped and tested for association with the disease. If a SNP shows evidence for association, a key question is to what degree the linkage result can be explained by the associated SNP. To answer this question, we developed a novel statistical method that quantifies the degree of linkage disequilibrium (LD) between the associated SNP and the putative disease locus. We summarized the available data by a simple parametric likelihood of the marker data conditional on the trait data based on disease penetrances and disease-SNP haplotype frequencies, and we proposed two likelihood ratio tests to distinguish the relationship between the associated SNP and the putative disease locus. We further compared different study designs for testing whether the SNP is responsible for a linkage signal. To investigate the performance of our method, we performed simulation studies.

311 Statistical Estimation and Testing ⌘

General Methodology

Wednesday, August 11, 8:30 am–10:20 am

Estimation of the Mean from Normal Distribution with Common Mean and Variance

◆ Greg Cicconetti, Muhlenberg College; Nitis Mukhopadhyay, University of Connecticut

Muhlenberg College, 2400 Chew St., Allentown, PA 18104

cicconet@muhlenberg.edu

Key Words: *uniform minimum unbiased estimator, point estimation, normal distribution, common mean and variance, Fisher information, sufficient statistic*

Assume that observations originate from a normal distribution with common mean and variance, q. The sample mean is an attractive estimator because of its unbiasedness and simplicity. However, if q is close to zero, then it is possible for the sample mean to be negative. This feature is unattractive. The sample variance is also an intuitive candidate for estimation q. Unfortunately, it's Fisher Information is less than that of the entire dataset. That is, the sample variance is not a sufficient statistics for q. Other candidates for estimating q include the sample median and the maximum likelihood estimator of q. Among the class of unbiased estimators, the uniform minimum variance estimator should be preferred. We shall weigh the cost of deriving the UMVUE against using these other naïve estimators of q. Simulations will provide us with some guidance when working in practice.

Estimating the Parameters of the Normal Using Median and Extreme Ranked-set Samples

◆ Hassen A. Muttlak, King Fahd University of Petroleum and Minerals

King Fahd University of Petroleum and Minerals, PO Box 1676, Dhahran, 31261 Sauda Arabia

hamstat@kfupm.edu.sa

Key Words: *simple random sample, ranked-set sample, maximum likelihood estimate, linear unbiased estimate, Fisher information, relative precision*

We propose maximum likelihood estimators (MLEs) as well as linear unbiased estimators (LUEs) of the parameters of the normal distribution, using median ranked-set sampling (MRSS) and extreme ranked set sampling (ERSS). MRSS and ERSS are modifications of ranked set sampling (RSS), which are more practicable and less prone to problems resulting from erroneous ranking. The MLEs of the mean under MRSS are shown to dominate all other estimators, while the mle of the normal standard deviation under ERSS is the most efficient. A similar trend is observed in the LUEs. A modification of ERSS, namely, partial extreme ranked-set sampling (PERSS), is proposed for odd set sizes to generate even-sized samples. The LUE of the normal standard deviation under this modification is shown to be the most efficient of all the LUEs of the same parameter. Among the LUEs considered, the PERSS LUEs are the most efficient when the sample size per cycle is two.

Unbiased Estimation Following a Group Sequential Test for Distributions in the Exponential Family

◆ Aiyi Liu, National Institutes of Health; Jack Hall, University of Rochester Medical Center; Kai F. Yu, National Institutes of Health; Chengqing Wu, National Institutes of Health

National Institutes of Health, DESPR/NICHD/NIH/DHHS, 6100 Executive Blvd., Rockville, MD 20852

liua@mail.nih.gov

Key Words: *clinical trials, completeness, Laplace transform, minimum variance, truncation-adaptation*

We consider unbiased estimation following a group sequential test for distributions in a one-parameter exponential family. We show that, for an estimable parameter function, there exists uniquely an unbiased estimator depending on the sufficient statistic and based on the truncation-adaptation criterion; moreover, this estimator is identical to one based on the Rao-Blackwell theorem. When completeness fails, we show that the uniformly minimum-variance unbiased estimator may not exist or otherwise possess undesirable performance, thus claiming the optimality of the Rao-Blackwell estimator, regardless of the completeness of the sufficient statistics. The results have potential application in clinical trials and other fields.

A Quasi-likelihood-ratio Test for the Equality of k Normal Population Coefficients of Variation

◆ Craig A. Wilson, Takeda Global R&D Center, Inc.

Takeda Global R&D Center, Inc., 475 Half Day Rd., Lincolnshire, IL 60069

cwilson@tgrd.com

Key Words: *coefficient of variation, quasi-likelihood*

The sample coefficient of variation (CV) has long been used as a descriptive measure of the relative variability of different groups or treatments. In recent years, several authors, including Bennett (1976) and Doornbos and Dijkstra (1983), have developed inferential tests for the equality of k normal population CVs. This paper investigates the use of the quasi-likelihood estimation method introduced by Wedderburn (1974), in conjunction with the asymptotic moments of the sample CV, to develop a quasi-likelihood-ratio test for the equality of k normal population CVs. The equivalence of the corresponding Wald and quasi-score tests to tests previously developed by Gupta and Ma (1996) and Feltz and Miller (1996) is demonstrated. The power of the new procedure when data are sampled from normal and right-skewed populations is examined.

Testing the Equality of Two Poisson Means Using the Rate Ratio

◆ Hon Keung Ng, Southern Methodist University; Man Lai Tang, Harvard Medical School

Southern Methodist University, Dept. of Statistical Science, 3225 Daniel Ave., Dallas, TX 75275-0332

ngh@mail.smu.edu

Key Words: *Poisson rate ratio, level of significance, constrained maximum likelihood estimation, sample size*

We investigate procedures for comparing two independent Poisson variates that are observed over unequal sampling frames (i.e., time intervals, populations, areas or any combination thereof). We consider two statistics (with and without the logarithmic transformation) for testing the equality of two Poisson rates. Two methods for implementing these statistics are reviewed: (1) the sample-based method, and (2) the constrained maximum likelihood estimation (CMLE) method. We conduct an empirical study to evaluate the performance of these two methods with and without logarithmic transformation. In general, all tests perform well for large Poisson rates and equal time/space. In particular we find that the CMLE method works satisfactorily only for the statistic without the logarithmic transformation, while sample-based method performs better for the statistic using logarithmic transformation. The corresponding sample size formulae are provided and valid in the sense that the simulated powers associated with the approximate sample size formulae are generally close to the pre-chosen power level. We illustrate our methodologies with a real example from a breast cancer study.

Two Approaches to the Behrens-Fisher Problem with and without Replication

◆ Paul J. Lupinacci, Villanova University; Terry Hyslop, Thomas Jefferson University

Villanova University, 800 Lancaster Ave., Villanova, PA 19085

paul.lupinacci@villanova.edu

Key Words: *Behrens-Fisher, replication*

We consider the Behrens-Fisher problem of testing means from two independent populations where the variances are not assumed to be equal. We propose an approach based on linear combinations of random variables as in Howe (1974). We also consider a test introduced by Lee and Fineberg (1991) who proposed a fitted test based on the critical values of Lee and Gurland's test (1975). We compare these methods in the standard setting of one observation per individual and then generalize both methods to replicate data. Simulation studies are used to demonstrate the level and power of these tests. A numerical example will be provided to illustrate the methods.

Testing for Latent Unit-by-treatment Interaction in Clustered Randomized Trials with Binary Outcomes

◆ Ed Mascha, Cleveland Clinic Foundation

Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH 44026

emascha@bio.ri.ccf.org

Key Words: *potential outcomes, unit-by-treatment interaction*

Individual treatment effect heterogeneity, also called unit-by-treatment interaction, is a quantification of the proportion of the population that responds differently to one treatment than to the other. It is quite distinct from the usual treatment-by-covariate interaction that is based on differing marginal treatment effects across levels of a covariate. With a given covariate, there may exist both types of interactions, one or the other type, or neither. Unit-by-treatment interaction may well vary over levels of a covariate or covariate patterns, but it is independent of the marginal treatment-by-covariate interaction with any covariate. It may be conceptualized and studied regardless of whether there are covariates related to it. We develop a method to test this type of interaction in parallel group randomized studies where we only observed 1 of the 2 binary outcomes of interest for each individual. The method depends on natural clustering in the data. We use a de-clustering algorithm to estimate the relationship between variabels known to provide information about the interaction of interest.

312 Statistics for Correlated Data ⌘

General Methodology

Wednesday, August 11, 8:30 am–10:20 am

Residual Analysis of Multidimensional Point Process Models for Earthquake Occurrences and Goodness-of-fit Assessment Using a Weighted Analog of Ripley's K-function

◆ Alejandro Veen, University of California, Los Angeles; Frederic P. Schoenberg, University of California, Los Angeles

University of California, Los Angeles, Dept. of Statistics, 8130 Math Sciences Bldg., Box 951554, Los Angeles, CA 90095-1554

veen@stat.ucla.edu

Key Words: *point process, residual analysis, K-function, earthquakes*

This work presents alternative approaches for assessing goodness of fit for multidimensional point process models. One new technique, based on an idea of P. Brémaud, involves augmenting the dataset with an additional dimension to create a residual point process. The procedure is applied to commonly used models for California earthquake occurrence data and the results are compared to other residual analysis methods as rescaling and thinning. Furthermore, a statistic which involves a weighted analog of Ripley's K-function is proposed for evaluating goodness of fit.

A Point Process Distance Metric with Applications to Earthquake Aftershock Activity

◆ Katherine E. Tranbarger, University of California, Los Angeles; Frederic P. Schoenberg, University of California, Los Angeles

University of California, Los Angeles, Dept. of Statistics, 8130 Math Sciences Bldg., Box 951554, Los Angeles, CA 90095-1554

ktranbar@stat.ucla.edu

Key Words: *point process, earthquakes, distance metrics, cluster analysis*

This work considers the problem of how to measure the distance between two point process sequences. Several properties of distance metrics proposed by Victor and Purpura are examined, and algorithms for their calculation are outlined. Using such a distance metric, a prototype sequence can be determined such that the distance between the prototype and all sequences in the dataset is minimized. This measurement approach, and a two-dimensional extension of it, are applied to global earthquake data and used to characterize the typical aftershock activity following a main shock of magnitude 7.5. Clustering techniques are also employed to identify groups of earthquakes with similar aftershock behavior, and to identify those outliers whose aftershock activity differs most from the prototype of the data. Comparisons are also made with the prototype sequence corresponding to the ETAS model, which is commonly used to describe aftershock activity.

Description and Classification of Wildfire Patterns Using Prototype Point Processes

◆ Frederic P. Schoenberg, University of California, Los Angeles; Edward A. Johnson, University of Calgary

University of California, Los Angeles, Dept. of Statistics, 8142 Math-Science Bldg., Los Angeles, CA 90095-1554

frederic@stat.UCLA.edu

Key Words: *point process, wildfires, prototypes, distance metrics*

The behavior and structure of historical wildfire patterns, inferred from tree scars, are summarized using prototype point patterns. These prototypes are used not only as summaries of fire interval patterns but also to identify outliers and to classify regions of forest into groups of trees exhibiting similar behavior.

Parameter Estimation in the Spatial Autologistic Model with Varying Independent Sub-blocks

◆ Johan Lim, Texas A&M University

Texas A&M University, Dept. of Statistics, Room 429 Blocker Hall, 3143 TAMU, College Station, TX 77843-3143

johanlim@stat.tamu.edu

Key Words: *likelihood approximation, spatial AR model*

We propose to approximate the likelihood function with independent sub-blocks; the entire data is subdivided into many sub-blocks which are treated as independent of each other. The approximate MLE denoted by MBILE is shown to have same asymptotic distribution with that of the MLE in the Ising model. The computational load for the MBILE is much lighter than the for the

MLE, and decreases geometrically as the size of a sub-block increases. Also, simulation studies show that, in finite sample, the MBILE is superior to other existing method in terms of mean squared error (MSE) if the interaction between spins are not too high. However, when the interaction is quite strong, the likelihood becomes barely identifiable and all existing methods including the MBILE shows poor performance. This is joint work with Haiyan Liu and Michale Sherman.

Estimating Spatial Patterns across Time: Using Hierarchical Mixed Effects Models for Variograms

◆ Daniel G. Kehler, Dalhousie University; Ransom Myers, Dalhousie University

Dalhousie University, Dept. of Biology, Halifax, NS B3H 4J1 Canada

kehler@mscs.dal.ca

Key Words: *variogram, bias, spatio-temporal, hierarchical mixed effect models*

In the context of spatio-temporal models, spatial information exists at multiple time intervals. Traditional methods for estimating the spatial pattern within a time period involve parametric representations in the form of variograms models. A general dependence in the spatial pattern across time can be imposed by treating variogram parameters as random effects realized from a common distribution. However, since variogram estimates at different spatial lags are often assumed to be independent observations, this can lead to an underestimation of the true parameter uncertainty, and hence an overestimation of the random effect variances. A more accurate assessment of uncertainty is obtained by dividing the data into two or more groups, and estimating multiple variograms in every year. We show, using simulations, how a hierarchical mixed effects model can accurately capture both the variogram uncertainty and correctly estimate variogram parameters in each time interval. The methods are applied to a dataset of fish recruitment in lakes by way of illustration.

A New Class of Joint Models for Longitudinal and Survival Data Accommodating Zero and Nonzero Cure Fractions

◆ Yueh-Yun Chi, University of North Carolina, Chapel Hill; Joseph G. Ibrahim, University of North Carolina, Chapel Hill

yychi@bios.unc.edu

Key Words: *joint models, cure rate models, random effects, breast cancer trial*

Joint models for a longitudinal response process and time-to-event have recently received a great deal of attention in cancer and AIDS clinical trials. Models for repeated measures and event time distribution are assumed to be independent given a common set of latent random effects. We extend the latent process model to multivariate repeated measures, and propose a novel approach to construct the conditional likelihood of the time-to-event, which accommodates both zero and nonzero cure rate structures. The occurrence of a cancer is assumed to be initiated by the mutation of some potential metastasis-competent tumor cells. The rate of generation of metastasis-competent tumor cells over time is modeled as a function of the true longitudinal trajectory as well as baseline covariates, and the promotion time for a potential cell to become a detectable tumor is independently parametrically modeled. A Bayesian paradigm is adopted to facilitate the estimation process and ease the computational complexity. The methodology is applied to a real dataset from an International Breast Cancer Study Group trial.

Estimation of Generalized Simple Measurement Error Models with Instrumental Variables

◆ Jeffrey R. Thompson, North Carolina State University; Randy L. Carter, University of Buffalo

North Carolina State University, 209C Patterson Hall, Campus Box 8203 NCSU, Raleigh, NC 27606

thompson@stat.ncsu.edu

Key Words: *measurement error models, instrumental variables, nonlinear models, generalized linear models, latent class models*

Measurement Error (ME) Models contain at least one independent variable that is imprecisely measured. This leads to an unidentified model and a bias in the naive estimate of the effect of the variable measured with error. One method to correct these problems is through the use of an instrumental variable (IV). We propose a method for parameter estimation, where an IV provides the identifying information, in simple nonlinear ME models. Our estimation method utilizes a “categorization” step and assumes conditional independence given the latent variable, to arrive at maximum likelihood estimates for the parameters of interest, through solving a resulting system of nonlinear equations using estimated generalized nonlinear least squares. New theorems on the identifiability of a large class of ME models are given. We show our estimators have favorable asymptotic properties and provide methods of inference for them. We show how many commonly studied ME models fit into, and can be solved, using the general method we developed.

313 Improving the Quality of Data Collected by Government Organizations ▲ ⌘

Social Statistics Section, Section on Survey Research Methods, Section on Government Statistics

Wednesday, August 11, 8:30 am–10:20 am

The Quality Framework: A Guide for Measuring Quality at the National Center for Health Statistics

◆ Kenneth W. Harris, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

kwh1@cdc.gov

Key Words: *principles, methodology, integrity*

The National Center for Health Statistics (NCHS) is the federal agency responsible for the collection and dissemination of the nation’s vital and health statistics. To carry out its mission, NCHS

conducts a wide range of annual, periodic, and longitudinal sample surveys and administers the national vital statistics registration system. In October 2002, as the result of a mandate from the Office of Management and Budget (OMB) to all federal statistical agencies, NCHS documented its quality guidelines on the NCHS web site. As a follow-up to that activity many agencies, including NCHS, began to systematically identify and document all aspects of their quality related procedures. This paper describes the quality control and quality-enhancing activities conducted at NCHS.

Measuring the Impact of Questionnaire and Envelope Messages on Respondent Behavior

◆ Theresa F. Leslie, U.S. Census Bureau; Deborah H. Griffin, U.S. Census Bureau; Pamela D. McGovern, U.S. Census Bureau; David A. Raglin, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., 7600, Washington, DC 20233-7600

Theresa.F.Leslie@census.gov

Key Words: *survey response, voluntary data collection, government surveys*

Dillman et al. found that reminding respondents that the decennial census was mandatory by providing a “Your response is required by law” message on the envelope led to an increase in mail response of about 10%. A similar test was recently conducted to assess the impact of removing all references to mandatory in a household survey and replacing those messages with motivational information such as “Your response is important to your community.” Two different voluntary messages were tested to assess respondent behavior. The test also evaluated if more user-friendly explanations of a mandatory survey still resulted in high levels of respondent cooperation. This test was conducted using the American Community Survey, a household survey collecting demographic, housing, social, and economic data. We determined that the removal of mandatory messages had a dramatic impact on cooperation rates by mail. We also found that varying messages within voluntary and within mandatory had only minor impacts. Little research exists to measure if varying motivational messages on the outgoing mail envelope leads to different levels of data completeness.

Making a Long Questionnaire Short: Alternatives to Reducing the Items on the Veterans Health Administration Learners’ Perceptions Survey

◆ Stephen J. Dienstfrey, Schulman, Ronca, & Bucuvalas, Inc.; Evert Milander, Dept. of Veterans Affairs; Gianna Dusch, Schulman, Roinca, & Bucuvalas, Inc.

Schulman, Ronca, & Bucuvalas, Inc., 8403 Colesville Rd., Suite 820, Silver Spring, MD 20910

s.dienstfrey@srbi.com

Key Words: *customer satisfaction, health care clinical training, questionnaire construction*

The Office of Academic Affiliations of the Veterans Health Administration began work on the Learners’ Perception Survey (LPS) in 1999. The initial survey was developed after an extensive literature review and 15 focus groups with faculty and trainees in a variety of healthcare professions. The resulting was a questionnaire asked trainee’s satisfaction on 50 specific elements from four domains: Clinical Faculty/Preceptors, Learning Environment, Working Environment and Physical Environment. In the 2003 survey, a fifth domain—Personal Experience—was added, resulting in a questionnaire that had 65 specific elements. There was concern that the growing length of the questionnaire could have a negative impact on the response rate. This paper compares the results of reducing the number of elements by using a domain satisfaction base and an overall satisfaction base for selecting elements to be included in the short form version of the basic LPS.

Use of Respondent Burden Measurements at NASS

◆ Wendy J. Scholetzky, National Agricultural Statistics Service

National Agricultural Statistics Service, 1400 Independence Ave., SW, Room 6344-A South Bldg., Washington, DC 20250-2007

wscholetzky@nass.usda.gov

Key Words: *burden measurements, respondent burden, annual cycle, burden management tools*

Each year, NASS conducts hundreds of surveys covering many facets of U.S. agriculture. For most surveys, the agency uses a list of known farm operators and agribusinesses to define the sampled populations. On an annual basis, NASS updates this list and selects new samples for the upcoming survey year. The sampled populations are fairly small and highly skewed to the right (the mean is much larger than the median) resulting in quite a few records being selected for multiple surveys. The agency is extremely concerned with respondent burden and continuing to achieve high response rates. Because samples are selected for a majority of the surveys during the annual cycle, a measurement of respondent burden for the upcoming survey year can be computed. NASS calculates several burden measurements for an operation since a respondent’s perception of burden can vary. These burden measurements consider the number of surveys, the number of contacts, and the total time required for an operation to complete the various survey questionnaires. Using these measurements, special data collection arrangements can be made for highly burdened operations.

On Inter-ater Reliability Measurement in Human Rights Violation Data Projects

◆ Romesh Silva, Human Rights Data Analysis Group

Human Rights Data Analysis Group, Benetech Initiative, 3/1A Rajakeeya Mawatha, Colombo 07, Sri Lanka

rsilva@martus.org

Key Words: *inter-rater reliability, human rights, categorical data, Kappa coefficient*

This paper reviews methods used in applying inter-rater reliability measures to events-based human rights violations data. It relates the development of coding frames for human rights violations data projects to the reliability of the actual coding processes. Popular measures such as Kohen’s Kappa and the proportion of overall and specific agreement are reviewed in the context of human rights violations data. Methodological Challenges in applying IRR measures to human rights data

projects are explored; for example, reliability measurement of coding processes with multiple coders (n>2) when the number of coders changes during the coding process. Furthermore, empirical results from a human rights data project are presented which present coding reliability for violations which result from acts of commission against ones that result from acts of omission.

Survey Documentation: On the Path to Enlightenment

◆ Shawna L. Waugh, U.S. Dept. of Energy

U.S. Dept. of Energy, Energy Information Administration, EI-70, 1000 Independence Ave., SW, Washington, DC 20585

shawna.waugh@eia.doe.gov

Key Words: *survey documentation, standards, transfer and survey quality, knowledge*

Survey documentation promotes the mission of the Energy Information Administration by assisting the agency to meet current and future challenges with survey planning, data collection, data processing, data dissemination, and evaluation. Survey documentation is critical in knowledge transfer. Survey documentation provides critical information to respondents on data collection activities, employees, and contractors on quality control procedures to minimze errors, and customers concerning reliability of the EIA's data. Survey documentation ensures transparency and reproducibility of the survey process and results. For example, technical notes in EIA publications explain survey methods and statistical techniques used to conduct the survey. This paper will focus on the following questions: Why federal statistical agencies invest in survey documentation? What is survey documentation? What are the best practices? What are strategies for overcoming obstacles to documenting survey processes and results?

Using Results from Census 2000 for Planning the 2010 Census

◆ Florence H. Abramson, U.S. Census Bureau; Agnes S. Kee, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 9200, Washington, DC 20233-9200

florence.h.abramson@census.gov

Key Words: *evaluations, experiments, research and development*

The basis for planning the 2010 Census largely evolved from what we learned in Census 2000. After the completion of Census 2000, the Census Bureau released reports on 87 evaluations, six experiments, and numerous operational assessments. The results and recommendations included in these reports support other information being used to plan and implement national content tests in 2003 and 2005 and operational site tests in 2004 and 2006. There are 11 2010 Research and Development Planning Groups reviewing results and recommendations and using them to develop new and improved methods covering the following areas: American Indian and Alaska Native, Communications, Content, Coverage, Field Activities, Language Program, Overseas Enumeration, Race and Ethnicity, Self-response, Special Place/Group Quarters, and Systems and Software. This paper discusses plans for the 2010 Census tests and how the Census 2000 evaluation and experiment results are used to identify the objectives for 2010 Census research and development.

314 Regular Contributed Posters

Section on Physical and Engineering Sciences, Section on Quality and Productivity, Biometrics Section, Biopharmaceutical Section, Social Statistics Section, Business and Economics Statistics Section, Section on Statistical Computing, Section on Statistics and the Environment, General Methodology, IMS, Section on Statistical Graphics

Wednesday, August 11, 9:00 am–10:50 am

Spectral Analysis with Time-varying Frequency

◆ Liangang Liu, Southern Methodist University; Henry Gray, Southern Methodist University

Southern Methodist University, 3225 Daniels Ave., Room 134 Heroy Bldg., Dallas, TX 75275

lliu@mail.smu.edu

Key Words: *Wigner-Ville Distribution, linear chirp process*

The instantaneous spectrum of M- and G(λ)-stationary process was defined by Gray Woodward and Jiang (2003). The instantaneous autocorrelation function (ACF) was defined by Jiang (2003). The relationship between the instantaneous ACF and the instantaneous spectrum is established for both continuous and discrete process. A new method to estimate lambda and the origin offside of G(λ) stationary processes is proposed, which is based on the cross-entropy. Also, a new class of nonstationary process called Linear Chirp (L.C.) process is developed. The L.C. process has approximately linearly changing frequency. The Wigner-Ville time-frequency distribution is used to estimate the parameters in the L.C. process. Both simulated and real geophysical data are analyzed using the L.C. process.

Temporal-spatial Covariance Structure of Water Use Efficiency of Lentil in Rotation with Wheat

◆ Murari Singh, International Center for Agricultural Research in the Dry Areas; Theib Oweis, International Center for Agricultural Research in the Dry Areas; Mustafa Pala, International Center for Agricultural Research in the Dry Areas; Ashutosh Sarker, International Center for Agricultural Research in the Dry Areas

International Center for Agricultural Research in the Dry Areas, Icarda, PO Box 5466, Aleppo, Syria

m.singh@cgiar.org

Key Words: *spatial, temporal, water, covariance, lentil*

Supplemental irrigation in a crop in dry areas facilitates an efficient use of water and is used as a mechanism for achieving higher and sustainable levels of crop productivity, particularly in crop rotation system. To model water use efficiency of lentil, a grain legume rich in protein, split-plot experiments were conducted at ICARDA station using lentil/wheat rotation system from 1997-2000. The statistical issues addressed are spatial variability in the fields and correlated response over year. Yearwise data were modeled with plot-errors following first-order autocorrelation with/without linear trends in the field, and combined with a constant correlation over time. AIC was used to select the best covariance structure. Spatial correlation with linear trend was detected in several cases. Compared with the standard split-plot analysis, AIC-best models gave stronger significance and higher precision of the main effects and interaction of supplemental irrigation and planting date on lentil water use efficiency. Extraction and use of most appropriate temporal-spatial covariance structures is recommended for estimation of effects of the agronomic factors.

Comparative Analysis of Start-up Demonstration Tests

◆ William S. Griffith, University of Kentucky; Michelle D. Smith, University of Kentucky

University of Kentucky, Dept. of Statistics, Lexington, KY 40506-0027

griffith@ms.uky.edu

Key Words: *demonstration test, run*

Start-up demonstration tests are used to decide whether to accept or reject equipment. Various criteria have been proposed involving consecutive (or total) numbers of failures and successes. For example, a test may terminate when there have been k consecutive successes (accept) or f total failures (reject) whichever comes first. One may also consider clusters of successes or failures rather than consecutive. We shall discuss and compare the analyses of such tests and some associated statistical issues.

Model Reduction of a Three-parameter Model for Survival Distributions to Two-parameter Weibull and Generalized Exponential Models in Estimation of Small Quantiles

◆ Elliott Nebenzahl, California State University, Hayward; Dean H. Fearn, California State University, Hayward

California State University, Hayward, 25800 Carlos Bee Blvd., Hayward, CA 94542-3435

enbenza@csuhayward.edu

Key Words: *survival distributions, parametric models, model reduction*

The exponentiated Weibull (EW) distribution depends on three parameters: a scale parameter and two other parameters that affect its shape. Special cases are the Weibull (W) distribution where one of the shape parameters is equal to 1 and the generalized exponential (GE) which has the other shape parameter equal to 1. We are interested in the performance of the GE and W estimates of the EW quantile and in a comparison of the two, in terms of which one more accurately estimates the true EW quantile. If reasonable estimates of the true quantile can be found, then we can operate with the simpler two-parameter models (Weibull and generalized exponential) rather than the more complicated tree-parameter EW model. We check whether either of the two-parameter models are satisfactory by considering, for a large sample of EW data, the ratio of the GE fit of a quantile to the EW fit of the quantile along with the ratio of the W fit of the quantile to the EW fit. For small values of the quantile we give ranges of the values of the shape parameters of the EW distribution that allow a satisfactory GE and/or W fit.

Using Web Technology to Extend or Customize Available Statistical Software

◆ John C. Nash, University of Ottawa; Sudan Wang, University of Ottawa

University of Ottawa, School of Management, 136 J-J Lussier Private, Ottawa, ON K1N 6N5 Canada

nashjc@uottawa.ca

Key Words: *web technology, user interface, open source*

Applications of statistical software benefit from customization to aid users, especially in the provision of graphical interfaces to simplify the bility of sophisticated software tools. This poster presentation uses web technology, that is, the infrastructure generally employed for working with interactive HTML pages over the internet, as a way to build such interfaces. These may be used remotely, over the internet, but may also be used via a local web server on a single computer. Our example tasks involve the development of interfaces to R and to SciLab to carry out time-series forecasting using trend-season and ARIMA modeling. This experience has been summarized in a set of recommendations as well as a proof-of-concept Java program that "writes" necessary interface program code, in our case in Perl, JavaScript, and HTML. This meta-tool is the focus of ongoing work.

From Nonstationary to Stationary: Continuous-time ARMA and the Kalman Filter

◆ Zhu Wang, Southern Methodist University; Henry Gray, Southern Methodist University; Wayne A. Woodward, Southern Methodist University

Southern Methodist University, SMU PO Box 750332, Dallas, TX 75275

zhuw@mail.smu.edu

Key Words: *nonstationary, time deformation, continuous ARMA, Kalman filter, unevenly spaced*

We consider methodology for analyzing nonstationary time series whose frequency content changes approximately linearly with time. After appropriate time deformation, the evenly spaced nonstationary time series is transformed to unevenly spaced stationary data. To analyze this type of unevenly spaced data, interpolating to obtain evenly space data and re-interpolating to forecast are needed in the previous work. To avoid these interpolations, we use the Kalman filter to fit a continuous ARMA process. For this type of nonstationary data, we discuss techniques of spectral estimation, model selection and diagnosis, filtering, smoothing, forecasting, and simulating realizations. We apply these techniques to simulated and bat echolocation chirp data.

An Overview of Incorporating Data into Deterministic Models of Fluid Transport in the Chesapeake Bay

◆ Gary Fowler, U.S. Naval Academy

U.S. Naval Academy, Mathematics Dept., 572 Holloway Rd., Annapolis, MD 21401

gof@usna.edu

The transport of fluid particles is typically modeled by partial differential equations (e.g., Navier-Stokes). Of course, uncertainty associated with the coefficients and initial conditions produces uncertainty in predicted paths. Complementing the mathematical models are actual observations. Both Eulerian (at fixed positions) and Lagrangian (along fluid paths) observations are available. This paper will outline techniques for incorporating observations into the modeling process.

Monitoring Industrial Generation of Electricity Using Multivariate Statistical Procedures

Robert L. Mason, Southwest Research Institute; Joe H. Sullivan, Mississippi State University; Zachary G. Stoumbos, Rutgers University; ◆ John C. Young, McNeese State University

McNeese State University, Lake Charles, LA 70609

jyoung@incontroltech.com

Key Words: *autocorrelation, Hotelling's T-Square, multivariate statistical process control, regression analysis*

Large industrial operations often use several forms of energy, such as electricity and steam, in their production processes. The most popular unit systems for the internal powerhouse generation of electricity are the steam turbine, the gas turbine, and a cogeneration combine cycle configuration using both steam and gas turbines. We present a case study for the problem of monitoring the overall powerhouse performance, including monitoring the energy conversion efficiency of a turbine unit. We also statistically assign the load among the various units in such a way as to minimize the fuel ges. The statistical procedures used in this case study include multivariate statistical process control methods based on the application of Hotelling's T-Square statistic, multivariable techniques such as regression analysis, and time-series modeling of a nonlinear mean response with autocorrelation.

The Properties of the Standard Deviation Illustrated with Engineering Data

◆ D. W. McCormack, Jr., International Sematech Manufacturing Initiative

International Sematech Manufacturing Initiative, 2706 Montopolis Dr., Austin, TX 78741-6499

don.mccormack@ismi.sematech.org

Key Words: *chi distribution, measurement error, noncentral chi-squared distribution*

The standard deviation is commonly used to develop metrics for engineering data. It is appealing because it expresses the measurand in its original units rather than in squared units. Unfortunately, literature on the statistical properties of the standard deviation is less prevalent than that for the variance. This presentation illustrates the development and application of frequently used properties of the standard deviation, such as its mean and variance, from the context of measuring deviations from an idealized feature on an integrated circuit.

Use of Fraility Models to Analyze Repeated Viral Infection Data

B.P. Madhura, GSK Pharmaceuticals; ◆ Ashwini K. Mathur, GlaxoSmithKline

GlaxoSmithKline, Biomedical Data Sciences India, 9th Floor, 121 Dickenson Rd., Bangalore, 560042 India

ashwini.k.mathur@gsk.com

Key Words: *survival analyses, viral infection, correlated survival times*

For viral infections, the main endpoint that is analyzed is time to first infection and subsequent infections. Drugs and different dosing schemes (continuous vs. intermittent) have been developed to delay these infection times. Comparison of different dosing schemes has been explored in this paper using models to account for correlated infection times within a patient. The results are also compared to basic survival models where the correlated infection times are analyzed independently by first, second, third, and so on occurence of the infection.

Fitting Lognormal Distributions to Data from a Size-sorting Particle Device

◆ Stanley A. Shulman, National Institute for Occupational Safety and Health; Paul A. Baron, National Institute for Occupational Safety and Health; Dawn M. Ramsey, National Institute for Occupational Safety and Health

National Institute for Occupational Safety and Health, 4676 Columbia Pkwy., MS R-3, Cincinnati, OH 45226

sas2@cdc.gov

Key Words: *nonlinear optimization*

The cascade impactor is a device for sampling airborne particles to determine the particle mass-weighted aerodynamic-diameter distribution. The particles enter at the top of the device and deposit by decreasing particle diameter on one of the nine stages. Past experimental data has established the fraction of each particle size expected to deposit on each stage. From the nine masses, estimates of the parameters of the lognormal distribution (either a single distribution or a mixture of two) must be obtained. Procedures based on weighted least squares are evaluated for fitting one or two lognormal distributions to both actual and simulated data. For the simulated data, several different scenarios are considered: geometric means and standard deviations producing distinct peaks or indistinct peaks; distributions with substantial mass in the largest particle size stage. A main conclusion is that the weighted least squares solutions produce imprecise estimates of the lognormal parameters.

An Adequacy Test for Time Series Models Using Estimating Functions

◆ Melody Ghahramani, University of Manitoba; Aerambamoorthy Thavaneswaran, University of Manitoba

University of Manitoba, Dept. of Statistics, Winnipeg, MB R3T 2N2 Canada

umghahra@cc.umanitoba.ca

Key Words: *estimating functions, time series, model identification*

Combined estimating functions method had been used to study filtering for nonlinear time series models in Thavaneswaran and Thompson (2002) and for prediction of nonlinear time series models in Thavaneswaran and Heyde (1999). Typically, model identification or adequacy of linear time series models involves checking normality. However, for many financial time series, nonlinear time series such as ARCH models and GARCH models are more appropriate. We propose a time series identification method using the combined estimating function theorem result. We illustrate the method with a few financial datasets.

Spatial Data Analysis of Lung Cancer Mortalities in Texas

◆ Abhishek Lall, Sam Houston State University; Donald Albert, Sam Houston State University; Ferry Butar Butar, Sam Houston State University

Sam Houston State University, Dept. of Mathematics and Statistics,
PO Box 2206, Huntsville, TX 77341

abhisheklall@hotmail.com

Key Words: *kriging, variogram, prediction, anisotropy, MSE*

"Are neighboring measurements more likely to be similar in value than distant ones?" Crude rate estimate is usually unreliable due to general inaccuracy in number of reported cases in regions of smaller population. However, this estimate can be smoothened by considering the similar observations in the neighboring regions. We analyze mortalities caused by Lung Cancer across 254 counties in Texas during 1990-1997 and try to predict the data at certain locations. We then compare the predicted values against the actual sampled data and examine the spatial dependency of the sampled data. We use the process of kriging to accomplish the objectives. An experimental variogram is constructed, which is run through various testing steps to make it as optimal as possible. This variogram is later used to determine the optimal assigned weights. Using these weights, a minimum error variance linear estimate is constructed at a location where the true value is unknown.

Two Way ("Hour-Month") Time-section Plots as a Tool for Climatological Visualization and Summarization

◆ Charles Fisk, Naval Base Ventura County

Naval Base Ventura County, Newbury Park, CA 91320

cjfisk@worldnet.att.net

Key Words: *climatic/oceanographic data visualization, contour plots, diurnal/seasonal time-section plots*

This study discusses and demonstrates the utility of two-way ("hour-month") time-section plots as a means of visualizing single station climatological or oceanographic data, a mode of presentation certainly not unknown but not believed to be widely used in practice. Data of these kind usually reside in tabular, digitized form, but the recent advent of data analysis/visualization software and desktop computing power makes it feasible to "transform" multiple digital summaries of these kind into single page graphical layouts that give a comprehensive visual feel for the seasonal/diurnal variations of parameter(s) of interest. The graphs also lend themselves to effective presentation on the web. In this mode, calendar month comprises the y-axis, hour of the day the x-axis. Upon the grid, climatic or oceanographic variables are contoured, represented by symbols, or colored. Multiple contour are also possible, and sunrise/sunset demarcation lines can be overlain. Derived variables such as mean vector wind/constancy are also represented effectively in this manner. A number of charts types are presented and explained.

Bayesian Test for Asymmetry and Nonstationarity in MTAR Model with Possibly Incomplete Data

◆ Man-Suk Oh, Ewha Women's University; Soo Jung Park, Ewha Women's University; Dong Wan Shin, Ewha Women's University; Byeong Uk Park, Seoul National University; Woo Chul Kim, Seoul National University

Ewha Women's University, Seo-Dae-Moon Gu, Dae-Hyun Dong 21,
Dept. of Statistics, Seoul, 120-750 Korea

msoh@ewha.ac.kr

Key Words: *nonlinearity, model selection, Markov chain Monte Carlo, multiple test*

We propose an easy and efficient Bayesian test procedure for asymmetry and nonstationarity in MTAR model with possibly incomplete data. Estimation of parameters and missing observations is done by using a Markov chain Monte Carlo (MCMC) method. Testing for asymmetry and nonstationarity is done via test of multiple hypotheses representing various types of symmetry/asymmetry and stationarity/nonstationarity. This allows simultaneous consideration of parameters relevant to asymmetry and nonstationarity of the model, and also enables us to find the sources of asymmetry and nonstationarity when they exist. Posterior probabilities of the hypotheses are easily computed by using MCMC outputs under the full model, with almost no extra cost. We apply the proposed method to a set of Korean unemployment rate data.

Spatial Interpolation of Soil Properties by Incorporating Topographic Attributes

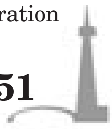
◆ Xuewen Huang, Michigan State University; Alexandra Kravchenko, Michigan State University

Michigan State University, Plant and Soil Science Bldg 175C, Crop and Soil Science Dept., East Lansing, MI, MI 48824

huangxue@msu.edu

Key Words: *multivariate geostatistics, kriging, spatial prediction, soil properties, DEM*

Topography is a major factor of soil formation and soil properties are related to landscape positions. Precision farming requires detailed soil property maps for successful site-specific management. Because measurements of soil properties are expensive and time-consuming, lack of soil observations limits potential benefits of precision farming. In contrast, field elevation data are easy to obtain. This paper will compare multivariate geostatistical methods for incorporating topographic attributes into spatial prediction of soil properties. The techniques will be illustrated using 33 soil sample sites and 5 x 10 m elevation measurements in a 50 hectare field. The analysis procedures will involve integration between statistical methods of data analysis and GIS.



Analysis of Residual Strength of Metals in a Corrosive Environment

◆ Gerald J. Shaughnessy, University of Dayton; Peter W. Hovey, University of Dayton

University of Dayton, 300 College Park, Dayton, OH 45469

shaughne@notes.udayton.edu

An important problem in maintaining aging aircraft is the loss of strength due to environmental degradation. Laboratory tests involve subjecting test specimens to corrosive environments while under load for specified time periods and then testing the specimens for residual strength. Current ASTM standards specify using a uniform distribution to simulate a failure time for specimens that fail during the aging process. This paper investigates the use of a model for residual strength degradation that treats the premature failures as censored data.

Speed as a Risk Factor in Run-off Rd. Accidents: Bayesian Analysis with Differential Measurement Error

◆ Gary A. Davis, University of Minnesota

University of Minnesota, 122 CivE, 500 Pillsbury Dr. SE, Minneapolis, MN 55455

drtrips@umn.edu

Key Words: *Markov chain Monte Carlo, accident reconstruction, traffic safety, Bayesian methods, logistic regression*

Research done in the 1960s and 1970s suggested that Rd. accident risk is a U-shaped function of speed, with risk increasing as one travels both faster and slower than what is average on a Rd.. Debate continues as to the causes of this relationship, and there is reason to suspect that it may be an artifact of measurement error or mixing of different accident types. This paper describes two case-control analyses of run-off Rd. accidents, one using data collected in Adelaide, Australia, and one using data from Minnesota. In both analyses the speeds of the case vehicles were estimated using accident reconstruction, while the speeds of the controls were measured on vehicles travelling the accident site under similar conditions. Bayesian relative risk regression was used to relate speed to accident risk, and uncertainty in the case speeds was accounted for by treating these as additional unknowns with informative priors. Neither dataset supported the existence of a U-shaped relationship, although accident risk clearly tended to increase as speed increased.

Semiparametric Joint Modeling of Longitudinal Measurements and Time-to-event Data

◆ Wen Ye, University of Michigan; Xihong Lin, University of Michigan; Jeremy M.G. Taylor, University of Michigan

University of Michigan, School of Public Health II, 1420 Washington Heights, M4331, Ann Arbor, MI 48109-2029

wye@umich.edu

Key Words: *semiparametric, joint modeling, longitudinal, survival*

Longitudinal studies in medical research often generate both censored time-to-event data and repeated measurements on biomarkers. Recently, joint models using both types of data have been developed. Commonly, the longitudinal covariate is modeled by a linear mixed model. However, in some cases, the biomarker's time trajectory is not linear, such as the prostate specific antigen PSA profile after radio-therapy in prostate cancer study. We propose a two-stage regression calibration approach which models the longitudinal biomarker using a semiparametric mixed model, where covariate effects are modeled parametrically and the individual time trajectories are modeled nonparametrically using a population smoothing spline and subject-specific random stochastic processes. Estimates of the biomarker level and change rate at each time-event are then used as time dependent covariates in the second stage survival model. We also propose an approach to jointly estimate parameters in the two submodels in the two-stage method. The performance of the approaches is illustrated by application to a prostate cancer study.

Analysis of Messy Longitudinal Data from a Study of Pediatric Patients

◆ Shesh N. Rai, St. Jude Children's Research Hospital; Shelly Lensing, St. Jude Children's Research Hospital; Sean Phipps, St. Jude Children's Research Hospital; James Boyett, St. Jude Children's Research Hospital

St. Jude Children's Research Hospital, Dept. of Biostatistics, 332 N. Lauderdale St., Memphis, TN 38105

Shesh.Rai@stjude.org

Key Words: *longitudinal study, quality-of-life study, model-based inference*

"Messy" data are more common than "neat" data in real-world clinical studies. The quality of messy data is always questioned; therefore, inferences drawn from such data need special attention. We describe a messy data problem that confronted us and the way that we dealt with it. The objective of the study that generated the messy data was to relate baseline psychometric status to longitudinal quality-of-life indicators for pediatric patients with cancer. The data were very messy because of the imprecise manner in which they were collected and their incompleteness from a variety of sources. Furthermore, analyses of these data were complicated for several reasons. The longitudinal nature of the data required modeling of the correlation between multiple measurements taken on the same subject. Missing responses necessitated justification of assumptions. Different analytic approaches were used because continuous and ordinal dependent variables were obtained. Simplification and grouping strategies were used to deal with the large number of independent variables. We discuss the model-based inference procedures used to address this problem and describe the application.

Spatial Distribution of Respirable Dust Concentrations in the Vicinity of a Jackhammer Operator

◆ W. Karl Sieber, National Institute for Occupational Safety and Health; Alan Echt, National Institute for Occupational Safety & Health; Daniel Lefkowitz, New Jersey Dept. of Health and Senior Services

National Institute for Occupational Safety and Health, 4676 Columbia Pkwy., Cincinnati, OH 45255

wks1@cdc.gov

Key Words: *spatial distribution, variogram, respirable dust, exposure, occupation*

In many studies, the extent of exposure to a particular agent at locations away from the source of that agent is of interest. In this study the spatial distribution of respirable dust concentrations in the vicinity of a jackhammer operation was investigated. A laser photometer was used to measure respirable dust concentrations at 30 points separated by two foot intervals on a 20 foot by 16 foot grid where the origin of the grid was the location of the jackhammer operator. Mean concentrations varied between 3.23 mg/m3 near the construction worker to 0.04 mg/m3 at the far ends of the grid. A semivariogram was constructed which indicated that the data followed a spherical model with a lag of two and maximum number of lags of four. The plot further indicated that at distances greater than 5.8 feet from the point of highest concentration no spatial correlation with other dust measurements would be found. Such information on the distribution of concentrations in the vicinity of the source of the agent is important in the occupational setting for determining appropriate use of protective personal equipment, establishing safe boundaries, or other control measures for other workers.

315 Late-breaking Session #2 - The Latest Report to NSF on Funding Statistical Research and Education: Challenges and Opportunities

ASA, ENAR, IMS, SSC, WNAR

Wednesday, August 11, 10:30 am–12:20 pm

The Latest Report to NSF on Funding Statistical Research and Education: Challenges and Opportunities

◆ Bruce G. Lindsay, Pennsylvania State University

bgl@psu.edu

The session will be organized to stimulate open discussion about the future of our field. It will be centered on the report but bring in other perspectives through the contributions of the discussants and audience. If successful, this session should help to multiply the number of people who are thinking actively about future directions and helping to bring them to fruition.

316 Using Control Charts to Monitor Process and Product Profiles ▲

Business and Economics Statistics Section, Section on Quality and Productivity, Section on Physical and Engineering Sciences, Business and Economics Statistics Section

Wednesday, August 11, 10:30 am–12:20 pm

Using Control Charts to Monitor Product and Process Quality Profiles

◆ William H. Woodall, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics, Blacksburg, VA 24061-0439

bwoodall@vt.edu

Key Words: *statistical process control, calibration, functional data analysis, splines, linear regression, nonlinear regression*

In most statistical process control (SPC) applications it is assumed that the quality of a process or product can be adequately represented by the distribution of a univariate quality characteristic or by the general multivariate distribution of a vector consisting of several correlated quality characteristics. In many practical situations, however, the quality of a process or product is better characterized and summarized by a relationship between a response variable and one or more explanatory variables. Thus, at each sampling stage, one observes a collection of data points that can be represented by a curve (or profile). We discuss some of the general issues involved in using control charts to monitor such process and product quality profiles and review the SPC literature on the topic. We relate this application to functional data analysis and review applications involving linear profiles, nonlinear profiles, and the use of splines and wavelets. We strongly encourage research in this area and provide some research ideas.

Phase I Analysis of Nonlinear Product and Process Quality Profiles

◆ James D. Williams, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics, 406-A Hutcheson Hall, Blacksburg, VA 24061-0439

jdwilliams@vt.edu

Key Words: *multivariate statistical process control, nonlinear regression, functional data, T^2 control chart*

In many quality control applications, use of a single (or several distinct) quality characteristic(s) is insufficient to characterize the quality of a produced item. In an increasing number of cases, a response curve (profile), is required. Such profiles can frequently be modeled using linear or nonlinear regression models. In recent research others have developed multivariate T^2 control charts and other methods for monitoring the coefficients in a simple linear regression model of a profile. However, little work has been done to address the monitoring of profiles that can be represented by a parametric nonlinear regression model. Here we extend the use of the T^2 control chart to monitor the coefficients resulting from a nonlinear regression model fit to profile data. We give four general approaches to the formulation of the T^2 statistics and determination of the associated upper control limits for Phase I applications. We also consider the use of nonparametric regression methods and the use of metrics to measure deviations from a baseline profile. These approaches are illustrated using the vertical board density profile data presented in Walker and Wright (2002).

A Changepoint Method Based on Linear Profile Data

◆ Mahmoud A. Mahmoud, Virginia Polytechnic Institute and State University; William H. Woodall, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, Dept. of Statistics, Blacksburg, VA 24061-0439

mamahmou@vt.edu

Key Words: *calibration, panel data, likelihood ratio, simple linear regression, statistical process control, two-phase regression*

We propose a changepoint method for testing the constancy of the regression parameters in a linear profile dataset. Each sample collected over time in the historical dataset consists of several bivariate observations for which a simple linear regression model is appropriate. The proposed method is based on the likelihood ratio test for a change in one or more regression parameters. We compare the performance of our proposed method to that of the most effective phase I linear profile control chart approaches using a simulation study. The advantages of our proposed method over the existing methods are greatly improved detection of sustained step changes in the process parameters and improved diagnostic tools to determine the sources of profile variation and the location(s) of the changepoint(s). Also, we give an approximation for appropriate thresholds for the test statistic. The use of our method is demonstrated using a dataset from a calibration application at National Aeronautics and Space Administration (NASA) Langley Research Center.

Profile Monitoring in Semiconductor Manufacturing

◆ Vivek B. Ajmani, General Mills, Inc.

General Mills, Inc., 330 University Ave., SE, Minneapolis, MN 55414

Vivek.Ajmani@genmills.com

Key Words: *linear profile, nonlinear models, regression, EWMA, Hotelling*

Control charts are widely used in semiconductor manufacturing to monitor processes. These control charts are typically used to monitor univariate parameters of interest to detect departures from their intended target. There are instances where it may be of interest to monitor profiles of the variables across time or space. This talk focuses on applications of both linear and nonlinear profiles monitoring in semiconductor manufacturing. Examples using actual data from Intel Manufacturing will be presented.

317 Noether Award Invited Session ▲ ⌘

Noether Award Committee

Wednesday, August 11, 10:30 am–12:20 pm

Nonparametrics and Mixture Models

◆ Tom Hettmansperger, Pennsylvania State University

Pennsylvania State University, 317 Thomas Bldg., University Park, PA 16802

tph@stat.psu.edu

Key Words: *nonparametric statistics, semiparametric models, EM algorithm, nontransitive samples, Efron dice*

In recent years mixture models have provided a very flexible approach to modeling diverse phenomena. EM algorithms have provided a convenient approach to computation so that mixtures can be fitted to data. We will discuss some work on semiparametric mixtures in which the underlying component distributions are not specified by any parametric model. In particular, we do not wish to impose the condition that the components come from the same family. We will discuss situations in which the mixing proportions and the underlying component distribution functions can be estimated along with the number of components in the mixture. When comparing three samples, if the samples arise from certain mixture distributions, nontransitivities can arise. We will discuss the structure of such mixtures and propose nonparametric tests to detect the nontransitivities. All proposed methods will be illustrated on datasets.

Component Selection and Smoothing in Smoothing Spline Analysis of Variance Models

◆ Yi Lin, University of Wisconsin, Madison; Hao H. Zhang, North Carolina State University

University of Wisconsin, Madison, 1210 West Dayton St., Madison, WI 53706-1685

yilin@cs.wisc.edu

Key Words: *model selection, variable selection, reproducing kernel, Hilbert space*

We propose a new method for model selection and model fitting in nonparametric regression models, in the framework of smoothing spline ANOVA. The “COSSO” is a method of regularization with the penalty functional being the sum of component norms, instead of the squared norm employed in the traditional smoothing spline method. The COSSO provides a unified framework for several recent proposals for model selection in linear models and smoothing spline ANOVA models. Theoretical properties, such as the existence and the rate of convergence of the COSSO estimator, are studied. In the special case of a tensor product design with periodic functions, a detailed analysis reveals that the COSSO applies a novel soft thresholding type operation to the function components. We give an equivalent formulation of the COSSO estimator which leads naturally to an iterative algorithm. We compare the COSSO with the MARS, a popular method that builds functional ANOVA models, in simulations and real examples. The COSSO gives very competitive performances in these studies.

318 New Methodologies for QT Interval Prolongation Adjustments in Pharmaceuticals ▲ ⌘

Biopharmaceutical Section

Wednesday, August 11, 10:30 am–12:20 pm

Characterization of the Heart Rate, QT Interval and Appropriate QTc of Conscious Beagle Dogs, Chair-trained Monkeys and Anesthetized Mongrel Dogs in Telemetry Studies

◆ Lori Mixson, Merck & Co., Inc.; Chao-Min L. Hoe, Merck & Co., Inc.; Hugo Vargas, Merck & Co., Inc.; Peter Siegl, Merck & Co., Inc.; Richard Briscoe, Merck & Co., Inc.

Merck & Co., Inc., 770 Sumneytown Pike, PO Box 4 WP37C-305, West Point, PA 19486-0004

lori_mixson@merck.com

Key Words: *telemetry, preclinical, simulation*

Understanding, interpreting, and predicting QT prolongation caused by drug treatment is important for health outcomes, as there is a risk for sudden death. As such, there is also an increased regulatory interest in detecting potential QT prolonging drugs early in development. New statistical methods arise from improved technology to measure cardiac function. Telemetry studies in animals are key to the safety assessment of research compounds. Electrocardiography (ECG) data from telemetry studies conducted on conscious beagles (14 studies), chair-trained monkeys (10 studies), and anesthetized mongrel dogs (14 studies) were analyzed to characterize the heart rate and QT interval. This data consisted of approximately 20 hours of pre-dosing/baseline measurements collected at one-minute intervals. Each study was a 4x4 cross-over design, such that each animal in a study contributed 1,200 minutes per time period to the analysis. The variation over time within animal and variation among animals was used to calculate the variation in the strength of the association between heart rate and QT.

Repeated-measures Models in the Analysis of QT Interval

◆ Brian P. Smith, Eli Lilly and Company

Eli Lilly and Company, Lilly Research Laboratories, Eli Lilly and Company, Drop Code 2233, Indianapolis, IN 46285

smith_brian_p@lilly.com

Key Words: *QT interval, QT correction, repeated-measures models*

Because of the recent regulatory emphasis on issues related to drug-induced cardiac repolarization that can potentially lead to sudden death, the analysis of QT interval has received much attention in the clinical trial literature. The analysis of QT data is complicated by the fact that QT interval is correlated with heart rate and other prognostic factors. Several attempts have been made in the literature to derive an optimal method for correcting QT interval for heart rate. However, obtained QT correction formulas are not universal because of substantial variability observed across different patient populations. It is shown that the widely used fixed QT correction formulas do not provide an adequate fit to QT and RR data and provide bias estimates of treatment effect. The QT correction formulas derived from baseline data in clinical trials are also likely to lead to Type I error rate inflation. We developed a QT interval analysis framework based on repeated-measures models accommodating the correlation between QT interval and heart rate and the correlation among QT measurements collected over time.

A Maximum Likelihood Approach for Estimating the QT Correction Factor Using Mixed-effects Model

◆ Amrik Shah, Schering-Plough Corporation

Schering-Plough Corporation, 2015 Galloping Hill Rd., K-15-2, Kenilworth, NJ 07033

amrik.shah@spcorp.com

Key Words: *Bazett, QT, QTc, prolongation, Friderecia*

Assessment of QT interval prolongation is often used for assessing the cardiac safety of a new drug. However, the correction of the QT interval for varying heart rates has potential bias due to various different correction factors. We propose a maximum-likelihood (ML) approach for calculating the appropriate individual correction factor using the data. The data come from a study with 24 subjects participating in a 10-day multiple dose (NEW RX), placebo-controlled crossover trial with repeat ECG's obtained at baseline and at Day 10. ML techniques were used to fit a random-effects model to observed QT and HR values for estimating the pooled and individual correction factors. QTc values using four correction factors (Bazett, Friderecia, Pooled, and Individual) were also investigated. The relative performance of the various correction factors are given in terms of variability and graphical techniques.

Reliable QT Correction for Clinical Studies: The Best Linear Unbiased Predictor (BLUP) Method

◆ Daniel C. Park, GlaxoSmithKline; Kwan R. Lee, GlaxoSmithKline; Xiwu Lin, GlaxoSmithKline

GlaxoSmithKline, 1250 S. Collegeville Rd, Collegeville, PA 19426 US

daniel_c_park@gsk.com

Key Words: *drug-induced QTc prolongation, QTc interval, best linear unbiased predictor (BLUP)*

The estimation of QT interval prolongation induced by compounds of pharmacological interest has become very important in the development and safety assessment of every drug. However, reliably identifying QT prolongation has been very difficult because of the strong dependence of the QT interval on heart rate. Many heart rate correction formulas have been proposed since Bazett first introduced his correction formula for the QT interval in 1920. These published formulas, however, are problematic: they fail to remove correlation between the QT interval and heart rate, and they often under- or over-correct the QT interval. To improve on these formulas, both pooled and individually optimized regression methods have been proposed. The pooled methods are not always reliable because they do not account for variations among individuals, and the individual methods are not always practical for use in clinical studies. We propose a powerful and practical approach to correcting the QT interval using the random coefficient regression model. Through simulations, we provide comparisons of this method with other correction methods that are currently being used.

Covariate-adjusted Reference Intervals for Diagnostic Data with Application to Electrocardiographic Data

◆ Alexei Dmitrienko, Eli Lilly and Company

Eli Lilly and Company, Lilly Research Laboratories, Drop Code 2233, Indianapolis, IN 46285

Dmitrienko_Alex@lilly.com

Key Words: *QT interval, reference interval, quantile regression*

The analysis of extreme diagnostic measurements—e.g., QTc interval—in clinical trials relies on reference intervals that help drug developers quickly determine whether a particular value is typical or atypical. The distribution of QTc interval is greatly influenced by various covariates such as patient’s age, and it is important to properly account for this influence in the analysis of extreme QTc values. This presentation discusses several approaches to constructing covariate-adjusted reference intervals for QTc interval based on quantile smoothing and stepwise quantile approximations. A detailed review of methods for optimizing the quantile estimation procedures is provided. The described methods for computing covariate-adjusted reference intervals are applied to the derivation of reference intervals for absolute values of and changes in QTc interval and other ECG parameters.

319 History of Data Visualization ⌘

Section on Statistical Graphics, Section on Physical and Engineering Sciences, Section on Quality and Productivity, Section on Statistical Computing

Wednesday, August 11, 10:30 am–12:20 pm

The Milestones Project: A Case Study in Statistical Historiography

◆ Michael Friendly, York University

York University, Psychology Dept., 4700 Keele St., Toronto, ON M3J 1P3 Canada

friendly@yorku.ca

Key Words: *data visualization, thematic cartography, timelines*

The graphic representation of quantitative information has deep roots that reach into the histories of thematic cartography, geodesic measurement, statistics and statistical graphics, medicine, and other fields, and also into enabling developments in technology for reproduction, display, and computation. The Milestones Project attempts to document and illustrate these historical developments leading to modern data visualization and visual thinking. It consists of a comprehensive catalog of important milestones in all fields related to data visualization, with representative images, bibliographical citations, cross-references, web links. In this presentation, I describe the origins and present status of this project, and try to address questions of statistical historigraphy—how to represent, analyze, display, and study themes, antecedants, influences, patterns, trends, and so forth.

Henri’s Line

◆ Antoine de Falguerolles, Université Paul Sabatier

Université Paul Sabatier, Laborartoire de Statistique et Probabilités, 118 route de Narbonne, Toulouse, 31000 France

falguero@cict.fr

Key Words: *QQ plot, normal QQ plot*

A quantile-quantile plot is a standard tool for comparing two distributions. Boosted by the use of popular statistical software programs, the colloquial expression “‘QQ plot” has been recently picked up in the French language. A particular instantiation of the QQ plot is the “normal” QQ plot also known in French in this colloquial form. This language practice, Franglais or Frenglish, masks a surprising oblivion. The idea of an experimental validation of the normal distribution in the form of a QQ plot is ascribed in France to Paul Henri (1848-1907), an artillery captain at that time. His approach was called in textbooks “Henri’s method” and the line describing the fitted distribution as “Henri’s line” (‘droite de Henri’). I will recall the work of Paul Henri and an interesting application of Edmond Lhoste (1880-1948) to the case where the data come from several identified normal distributions with possibly different means and different variances.

A Graphical Legacy of Charles Joseph Minard

◆ Howard Wainer, National Board of Medical Examiners

National Board of Medical Examiners, 3750 Market St., Philadelphia, PA 19104

hwainer@nbme.org

Key Words: *Minard, Napoleon, flow maps, graphical disasters*

Charles Joseph Minard (1781-1870) was a French civil engineer turned economic geographer-cartographer who is known today principally because of a graph he drew of Napoleon’s march on Moscow, which was published in the year before his death. But because Minard’s graphical work was so much broader than that one plot, the mathematician Howard Gray Funkhouser (1898-1984), in his 1937 history of graphical methods, dubbed Minard “the Playfair of France.” Minard’s work had substantial effect on the practice of statistical communication. For 21 years, from 1879 until 1899, the Bureau de la Statistique Graphique of the Ministry of Public Works annually published a large format chart-book. This series, l’Album de Statistique Graphique, rivals in sophistication and polish the very best statistical atlases published today. And most of the displays found within it derive very directly from earlier versions designed by Minard. We will discuss two graphs taken from that series. In addition, we will show how even the most magnificent of graphical forms can be misused.

Georg von Mayr and Emil Eugen Roesle—Two German Pioneers of Statistical Graphics

◆ Rüdiger Ostermann, University of Applied Sciences

University of Applied Sciences, Dept. of Nursing, Leonardo Campus 8, Münster, 48149 Germany

ruediger.ostermann@fh-muenster.de

Key Words: *history, statistics, Germany*

The name of Georg von Mayr (1841-1925) is almost forgotten in the statistical community. But he made important contributions to the development of statistical graphics methodology. Ee present some examples of his early graphics, e.g. a mosaic plot from 1877, and some graphics of the Swiss Graphical and Statistical Atlas of 1897 and 1914 that he influenced. Emil Eugen Roesle (1875-1962) played a major role in the development of medical statistics and epidemiology in Germany. A major part of his work included the graphical display of obtained statistical data. In 1911 he organized the first International Hygiene-Exhibition in Dresden, which more than 5 million visitors attended. They showed their interest in the 259 graphical-statistical figures, such as trellis displays, of 35 national and international exhibitors. Furthermore, Emil Eugen Roesle wrote publications which dealt with the principle structure of graphical-statistical displays.

320 Statistical Consulting in an Academic Environment

Section on Statistical Education, Section on Statistical Consulting

Wednesday, August 11, 10:30 am–12:20 pm

Statistical Consulting within Agriculture

◆ Marcia L. Gumpertz, North Carolina State University

North Carolina State University, Dept. of Statistics, Box 8203, Raleigh, NC 27695-8203

gumpertz@ncsu.edu

Key Words: *consulting, agriculture, collaboration, tenure, training, service*

NCSU has a large College of Agriculture and Life Sciences (CALS). Several statistics faculty positions are partly funded by CALS for the purpose of providing consulting services to the college. The consulting faculty serve as statistical liaisons to CALS departments, providing faculty, staff, and graduate students with unlimited free statistical consulting. This liaison system promotes development of long-term collaborative relationships. Consulting and research responsibilities tend to blend together for statistics faculty in these positions. The liaison model has worked well for CALS; however, there are several challenges. Consulting service to the rest of the university is much more limited, and there is more demand than we can satisfy. Training students in consulting requires commitment from the entire statistics faculty, and everyone is stretched for time. The issue of evaluating consulting faculty for promotion and tenure remains difficult. The NCSU Statistics Department values consulting highly and has written guidelines to assess the scholarly contribution of statistical consulting, but the criteria must be discussed and interpreted anew and reaffirmed frequently.

Statistical Consulting within the University

◆ George P. McCabe, Purdue University

Purdue University, Mathematical Sciences Bldg., 150 N University St., West Lafayette, IN 47907-2067

mccabe@stat.purdue.edu

Key Words: *consulting, collaborative research*

Free access to appropriate advice on the use of statistical methods should be available to all university faculty and researchers. A statistical consulting service impacts the teaching, research, and service missions of the university. Working in this environment provides valuable experience for statistics graduate students. The problems encountered provide statisticians with a source of real-world problems that can lead to interesting and important research. The clients receive valuable assistance with their research. We need to create an environment where researchers have confidence that we can enhance their work and enable them to ask new and interesting questions. An active statistical consulting service should open the doors to collaborative research projects, joint publications, and substantial outside funding for the statistician.

Statistical Consulting within the Medical Sciences

◆ Michael H. Kutner, Emory University; Kirk Easley, Emory University

Emory University, Dept. of Biostatistics, 1518 Clifton Rd., NE, #320, Atlanta, GA 30322

mkutner@sph.emory.edu

Key Words: *statistical consulting, statistical collaboration*

This presentation will focus on statistical consultation that potentially impacts the scientific integrity of the research effort including sound statistical design, study implementation, analysis, and interpretation. However, the consulting statistician needs to foster and promote an environment that establishes the statistician as a true research scientist collaborator. To do this, the statistician must “sell” the researcher client on the “value” of having a statistician as an integral member of the research team. The Biostatistics Consulting Center (BCC) at Emory University provides consultative and collaborative support for Health Sciences Center investigators and will be described in detail. The BCC is a separate service center within the School of Public Health. Issues discussed include: (1) organizational structure and funding, (2) faculty involvement, (3) graduate student training, (4) research assistants, and (5) establishing a clientele.

321 Some Lessons Learned and Analytic Results from the Joint Canada/United States Health Survey

Section on Government Statistics, Section on Survey Research Methods, Section on Health Policy Statistics, Social Statistics Section

Wednesday, August 11, 10:30 am–12:20 pm

Legal and Procedural Issues Encountered in Conducting the 2003 JCUSH

◆ Jennifer Madans, National Center for Health Statistics; Lorna Bailie, Statistics Canada

National Center for Health Statistics, CDC, 3311 Toledo Rd., Hyattsville, MD 20782

jmadans@cdc.gov

Key Words: *legal requirements, ethical requirements, international comparability*

The Joint Canada U.S. Health Survey, a one-time telephone survey conducted in both countries, tests whether two national statistical offices can integrate their survey design and analysis activities to achieve a greater degree of international comparability. To achieve this objective, design and implementation differences at all stages of the survey were controlled and a single data collection agent, Statistics Canada, was used. As a part of the planning process, all differences between U.S. and Canadian legal and ethical requirements were identified and evaluated as to how they might impact the survey. Collection activities were designed in accordance with the requirements found in the respective countries. Although the fundamental principles around privacy, confidentiality and respondent burden were the same in both countries, the processes for implementing these requirements were quite different. These differences had unanticipated consequences for how the survey was carried out. This presentation will review these legal and procedural issues and discuss what was learned about standard survey processes in both countries as these challenges were overcome.

Operational and Methodological Lessons Learned in Conducting the 2003 Joint Canada/United States Survey of Health

Ed Rama, Statistics Canada; ◆ Catherine Simile, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Hyattsville, MD 20782

csimile@cdc.gov

Key Words: *operations, Canada, United States*

The Joint Canada U.S. Survey of Health is a collaborative project undertaken by Statistics Canada and the U.S. Centers for Disease Control and Prevention National Center for Health Statistics. One of the key objectives of the project was to produce highly comparable health data between Canada and the U.S. To ensure comparability, all collection activities were performed by Statistics Canada. This included developing the collection system, administering the same questionnaire in both countries using Canadian interviewers, and monitoring collection activities. Despite the adoption of standardized surveying practices that both organizations adhere to, important procedural and methodological differences were uncovered that impacted the comparability of the data. Chief among them were the different approaches each country has toward the treatment of unresolved phone numbers and the implications this had on the final results. This presentation will outline some of the challenges that these procedural and methodological differences posed and discuss the lessons learned that can be used to facilitate future successful collaborations.

A Comparative Analysis of Four Disability/Functional Limitation Modules in the 2003 Joint Canada/United States Survey of Health

◆ Barbara M. Altman, National Center for Health Statistics; Edward Ng, Statistics Canada

National Center for Health Statistics, 3311 Toledo Rd., Room 6111, Hyattsville, MD 20782

baltman@cdc.gov

Key Words: *activity restrictions, disability, measurement, HUI, international comparison*

International comparison of functional health and disability is often hampered by important differences in key aspects such as survey instrument, data collection methods, etc. The Joint Canada/U.S. Survey of Health (JCUSH) represents the first time the same survey is conducted to measure the health of both Canadian and U.S. citizens, using standardized methodology. This paper compares disability and functional limitation in both countries, based on four different modules, namely, U.S. questions on vocational restriction of activities and on limitation of activities as well as Canadian questions on restriction of activity and those for the construction of the Canadian Health Utilities Index (HUI). These measures are clearly related to one another, but they produce different estimates of the levels and distribution of health status in populations. We will compare the age-adjusted prevalence rates of disability/functional limitation derived by the four modules, and explore whether they produce consistent differentials. These are important first steps in the discussion as to whether and how the questionnaires across countries can be brought closer in disability measurement.

A Comparative Analysis of Healthcare Service Utilization in the United States and Canada: Findings from the Joint Canada/United States Survey of Health

◆ Claudia Sanmartin, Statistics Canada; Debra Blackwell, National Center for Health Statistics; Jane F. Gentleman, National Center for Health Statistics; Michael Martinez, National Center for Health Statistics; Nancy Ross, McGill University; Jean-Marie Berthelot, Statistics Canada

Statistics Canada, R.H. Coats 24th Floor, Ottawa, ON K1A 0T6 Canada

claudia.sanmartin@statcan.ca

Key Words: *health care, utilization, access, Canada, United States*

Canada and the U.S. employ strikingly different approaches to financing medical care for their citizenry. Canada has a single-payer plan in which all residents have government-sponsored health insurance coverage for physician visits and hospitalization. In contrast, in the U.S., most funding for health care services comes from insurance providers. The two countries are similar, however, regarding dental care and prescription drugs for which individuals depend on private insurance coverage. The Joint Canada/U.S. Survey of Health will be used to examine differences in the determinants of health care utilization between the two countries with a particular focus on the effects of health insurance coverage. The analysis will focus on services for which private insurance plays a role in the U.S. but not in Canada and for services for which private insurance plays a role in both countries. Descriptive statistics and multivariate analyses (linear and logistic regression methods) will be used to describe and model health care service utilization.

322 Statistical Challenges in Computational Biology ▲

ENAR, Section on Bayesian Statistical Science

Wednesday, August 11, 10:30 am–12:20 pm

Analysis of Cis-regulatory Signals in the Genome

◆ Wing H. Wong, Harvard University

Harvard University, Statistics Dept., One Oxford St., Cambridge, MA 02138

wwong@stat.harvard.edu

We review advances in the computational discovery of cis-regulatory sequences that control gene transcription.

Lasso Methods for Computational Biology

◆ Rob Tibshirani, Stanford University

Stanford University, HRP Redwood Bldg., Room T101C, Stanford, CA 94305-5405

tibs@stat.stanford.edu

The Lasso is a shrinkage and selection method for linear regression. It minimizes the usual sum of squared errors, with a bound on the sum of the absolute values of the coefficients. It has connections to soft-thresholding of wavelet coefficients, forward stagewise regression, and boosting methods. We illustrate how this set of techniques can be applied to analyse microarray data and, more generally, to solve computational biology problems.

Studies in Statistics and Cancer Biology

◆ Michael A. Newton, University of Wisconsin

University of Wisconsin, Dept. of Statistics, 1210 West Dayton St, Madison, WI 53706

newton@stat.wisc.edu

Key Words: *genomic aberrations, polyclonality, overlapping clusters*

Cancer biology presents numerous statistical problems. I will consider two: the question of tumor polyclonality, and the question of selectively-advantageous genomic aberrations. Advances on the latter involve simple stochastic models, overlapping-cluster analysis, interactions, and context-based feature selection from high-dimensional aberration profiles. Advances on the polyclonality problem, in a collaborative project with W.F. Dove's lab, involve simple image analysis and spatial point processes to describe crypt architecture of the intestinal epithelium.

323 Early Warnings of Bioterrorism: Syndromic Surveillance ▲

Section on Statistics in Defense and National Security, Section on Health Policy Statistics, Section on Statistical Graphics, Social Statistics Section, Cmte on Statisticians in Defense and National Security

Wednesday, August 11, 10:30 am–12:20 pm

New Challenges for Analysis in Public Health Surveillance: BioSense

◆ Henry R. Rolka, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 1600 Clifton Rd., NE, mailstop K74, Atlanta, GA 30333

HRolka@CDC.Gov

Key Words: *bioterrorism, surveillance, syndrome, public health*

The growing abundance and use of administrative data, automated electronic reporting, surveillance integration efforts, and novel methods for data analysis have all engendered broad new requirements for analyzing surveillance data and incorporating statistical evidence into decision-making. This effect is amplified by the threat of bioterrorism and poses formidable new demands for interdisciplinary cooperation among scientists and technologists to develop mechanisms that accomplish useful public health surveillance. As coordination efforts for knowledge management and system interoperability across the broad spectrum of new decision support activities evolve, it is important to focus development and evaluation of analytic components in order to maintain effectiveness for the long term. An overview of CDC's BioSense Initiative will be provided along with a description of its data and analytic components. The objective is to motivate the practical data analysis and statistical challenges, describe some of the ways we're addressing these challenges and to elicit feedback for refining future analytic directions in this important public health activity.

Syndromic Surveillance with Multiple Data Streams

◆ Martin Kulldorff, Harvard Medical School and Harvard Pilgrim Healthcare

Harvard Medical School and Harvard Pilgrim Healthcare, Dept. of Ambulatory Care and Prevention, 133 Brookline Ave. 6th Floor, Boston, MA 02215

martin_kulldorff@hms.harvard.edu

Key Words: *spatial statistics, public health, surveillance, scan statistics, permutation tests*

The diagnosis of disease often requires multiple lab tests that may take anything from a day to a couple of weeks after the first symptoms arise. Whether due to bioterrorism or natural causes, it is thought that syndromic surveillance based on symptoms like vomiting and fever may be more useful for the early detection of disease outbreaks than diagnosed disease events, even though there is more noise in the data. Available syndromic surveillance data sources include hospital emergency visits, ambulance dispatch calls, and over-the-counter pharmacy sales. Most current syndromic surveillance systems evaluate each data stream independently. We will present statistical methods that simultaneously incorporates multiple data streams into a single likelihood function, which is then used with a permutation-based space-time scan statistics for the early detection of disease outbreaks. The method will be illustrated using syndromic surveillance data from New York City.

The Distribution of Interpoint Distances, Cluster Detection, and Syndromic Surveillance

◆ Marcello Pagano, Harvard School of Public Health; Marco Bonetti, Harvard University; Laura Forsberg, Harvard School of Public Health; Al Ozonoff, Harvard School of Public Health

Harvard School of Public Health, Bldg. II, Room 449, 655 Huntington Ave., Boston, MA 02115-6009

pagano@biostat.harvard.edu

Key Words: *detection power, spatial distribution*

Syndromic surveillance with triggers determined by too many individuals is enhanced by simultaneously considering the spatial distribution of the individual cases. The frequency distribution of the distances between individuals can be used to study the normal spatial distribution of individuals; especially when this distribution is used to detect clusters, or outbreaks. We investigate this problem and study the increase in detection power.

The Quantitative Evaluation of Biosurveillance Programs

◆ David Siegrist, Potomac Institute for Policy Studies

Potomac Institute for Policy Studies, 901 North Stuart St., Suite 200, Arlington, VA 22203

DSiegrist@PotomacInstitute.org

Key Words: *DARPA, detection algorithms*

The speaker is the principal investigator for validation of the DARPA Bio ALIRT biosurveillance program. He will discuss the quantitative evaluation of detection algorithms using simulation and the use of deidentified medical records to compare anomaly detection algorithms.

324 IMS Medallion Session 2 ⌘

IMS

Wednesday, August 11, 10:30 am–12:20 pm

Models for Spatial-temporal Processes

◆ Michael L. Stein, University of Chicago

University of Chicago, Dept. of Statistics, 5734 University Ave., Chicago, IL 60637

stein@galton.uchicago.edu

Key Words: *environmental statistics, air pollution, spatial-temporal processes, physical models, covariance functions*

This talk will discuss the use of stochastic and deterministic approaches, separately and in combination, for modeling processes that vary in space and time, with a specific focus on atmospheric processes. I will address a variety of conceptual and computational issues that arise when simultaneously analyzing computer model output and observations at a fixed set of monitoring locations.

325 Identifying and Understanding Product Association

Section on Statistics and Marketing

Wednesday, August 11, 10:30 am–12:20 pm

Investigating the Dynamic Impact of Price Promotions on the Strength of Product Associations

◆ Bernd Vindevogel, Ghent University; Dirk Van den Poel, Ghent University; Geert Wets, Limburg University Centre

Ghent University, Dept. of Marketing, Hoveniersberg 24, Ghent, B-9000 Belgium

bernd.vindevogel@UGent.be

Key Words: *multivariate time-series analysis, product associations, price promotions*

We apply the framework of market basket analysis to measure the strength of 324 product associations. Multivariate time series analysis (VARX) is used to assess the over-time impact of price promotions on the strength of the associations. The resulting 324 VARX models consist of lift, measuring the strength of the association, and the price of the two products as endogenous variables and a feature dummy as exogenous variable. This approach allows us to model both the short run and long run effects of price on lift. In phase two, we measure the dynamic impact of price on lift with impulse response models. In phase three, the results of the dynamic impact are used as a dependent variable in a regression model. In this way, we are able to identify the influence of covariates as promotional intensity, both frequency and depth, the nature of the products, like perishability, and the nature of the association (complementarity, independence or substitution) on the evolution of the strength of the association.

Market Structure Analysis Using Generalized IRT Models and Hierarchical Bayes Estimation Procedures

◆ Lynd D. Bacon, Lynd Bacon & Associates, Ltd.; Douglas Rivers, Stanford University; Jeff Hunter, General Mills, Inc.

Lynd Bacon & Associates, Ltd., Belmont, CA 94002-1512

lbacon@lba.com

Key Words: *market structure analysis, IRT, hierarchical Bayes, MCMC, product strategy, consumer insights*

There is a long tradition of quantitative analysis procedures for doing demand-side market structure analysis. These methods include internal methods in which the dimensions and nature of competition between brands or products is inferred from preference/choice data, and external methods in which the dimensions are identified a priori. Our paper adds to the work in this area by extending a technique described by Clinton, Jackman, and Rivers (2003) for analyzing political role call data, and applying it to the analysis of market structure. The method we describe produces maps that represent consumers and brands/products in a single, low-dimensional space by using a hierarchical Bayes specification of what is effectively a generalized multidimensional Item Response Theory (IRT) model. Our general specification is in the spirit of IRT extensions described by Junker (1997), and by Patz and Junker (1997). The key advantages of our method include that it provides a natural way to integrate perceptual and preference data of multiple measurement types and from multiple sources. We apply our model to consumer brand ge data spanning product categories.

Effect of Simpson's Paradox on Market Basket Analysis

◆ Yuejiao Ma, Pennsylvania State University; Dennis K.J. Lin, Pennsylvania State University

Pennsylvania State University, 326 Thomas Bldg., University Park, PA 16802

hma@stat.psu.edu

Key Words: *Market Basket Analysis, association rule, confidence, improvement, Simpson's Paradox, common improvement*

One of the well-studied problems in data-mining is pruning for association rules in Market Basket Analysis. Association rules discovery is an important database-mining algorithm that finds interesting association or correlation relationships among a set of items. Association rule induction is a powerful method for Market Basket Analysis, which aims at finding regularities in the shopping behavior of supermarket customers, online shops and the like. A useful association rule usually satisfies three measurements: support, confidence, and improvement, with minimum support, minimum confidence, and improvement greater than one. The measurements of Association rules are based on the aggregated dataset. It is very easy for the decision-maker to misinterpret the real relationship reflected in a nonaggregated dataset and miss potentially useful association rules. This well-known phenomenon is due to Simpson's Paradox that was defined by Simpson in 1951. For example, based upon the aggregated dataset, we find that the rule {if a customer buys product A, then this customer will also buy product B} does not satisfy the requirements at all.

A Comparison of Regression Models to Impute Continuous Variables of Partially Complete Observations

◆ Peter G. Wright, Statistics Canada

Statistics Canada, RHC 17-N, 120 Parkdale Ave., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

peter.wright@statcan.ca

Key Words: *imputation, regression, auxiliary data*

In an annual census of 3,800 foreign subsidiaries of Canadian enterprises, Statistics Canada collects values of sales, employment, and book value using a mail-out, mail-back questionnaire, supplemented with administrative data. Every year for over half of the subsidiaries, only book value is available. We compare regression models that were used to obtain sales and employment figures of partial nonrespondent subsidiaries. We address some of the practical challenges associated with the approach, including assessment of the variance and the potential for extrapolation.

326 Dollars, Power, and Genes: Novel Designs for Minimizing Cost and Maximizing Power in Genetic Studies

ENAR

Wednesday, August 11, 10:30 am–12:20 pm

Selective Phenotyping for Increased Efficiency in Genetic Mapping Studies

◆ Chunfang (Amy) Jin, University of Wisconsin, Madison; Hong Lan, University of Wisconsin, Madison; Alan Attie, University of Wisconsin, Madison; Dursun Bulutuglo, University of Wisconsin, Madison; Gary A. Churchill, University of Wisconsin, Madison; Brian S. Yandell, University of Wisconsin, Madison

University of Wisconsin, Madison, Dept. of Statistics, 1210 West Dayton St., Madison, WI 53706

chunfang@stat.wisc.edu

Key Words: *quantitative trait loci, interval-mapping, power, minimum moment aberration, diabetes*

The power of a genetic-mapping study depends on the heritability of the trait, the number of individuals included in the analysis, and the genetic dissimilarity among them. In experiments that involve microarrays or other complex physiological assays, phenotyping can be expensive and time consuming and may impose limits on the sample size. A random selection of individuals may not provide sufficient power to detect linkage until a large sample size is reached. We present an algorithm for selecting a subset of individuals based solely on genotype data that can achieve substantial improvements in sensitivity compared to a random sample of the same size. The selective phenotyping method involves preferentially selecting individuals to maximize their genotypic dissimilarity. Selective phenotyping is most effective when prior knowledge of genetic architecture allows us to focus on specific genetic regions. However, it can also provide modest improvements in efficiency when applied on a whole genome basis. Importantly, selective phenotyping does not reduce the efficiency of mapping as compared to a random sample in regions that are not considered in the selection process.

Picking the Most Likely Candidates for Further Development: Novel Intersection-union Tests for Addressing Multicomponent Hypotheses in Comparative Genomics

◆ David B. Allison, University of Alabama, Birmingham

University of Alabama, Birmingham, Ryals Public Health Bldg., Suite 327, 1665 University Blvd., Birmingham, AL 35294-0022

dallison@uab.edu

Key Words: *microarrays, target identification, intersection union tests, RNAi, Bayesian, frequentist*

In selecting genomic targets and physiological pathways for future study or drug development, investigators select a subset of candidates from among many options. In prioritizing, they may seek potential targets that meet all of several criteria—e.g., genes that are differentially expressed in response to a particular stimulus in each of several species (i.e,. evolutionarily conserved responses), or genes that are differentially expressed in response to a stimulus and also produce a predicted phenotypic change when knocked down in with RNAi. Testing whether several (k) null hypotheses can all be rejected requires intersection-union tests (IUTs). A traditional IUT rejects the union of all k null hypotheses in favor of the intersection of all k alternative hypotheses if a test for every one of the separate k null hypotheses is rejected. This IUT is generally conservative. Moreover, it yields results classifiable as significant or not, but not a quantitative p value. Herein, we examine an approach to frequentist testing to overcome these limitations. We then present a Bayesian approach to the problem that makes more effective use of data when many IUTs are conducted simultaneously.

Two-stage Designs for Gene-disease Association Studies

◆ Varghese George, University of Alabama, Birmingham

University of Alabama, Birmingham, Dept. of Biostatistics, 1665 University Blvd., RPHB 309C, Birmingham, AL 35294-0022

VGeorge@UAB.edu

Key Words: *association, candidate genes, complex diseases, disequilibrium-mapping, linkage, two-stage design*

Linkage and disequilibrium mapping methods to identify causative loci are important tools for understanding the genetic etiology of complex diseases and traits. Refinements of these methods will enhance ongoing efforts to effectively dissect the architecture of complex human diseases. We will discuss a two-stage TDT design in which association and linkage disequilibrium are tested sequentially. A major advantage of this design is that if a significant association observed during the first stage is spurious due to population admixture/stratification, it is likely to disappear in the second stage (unless due to Type I errors). If no significant association is found in the first stage, there is no need to proceed with the second-stage analysis for fine mapping in that region. On the other hand, if we find significance at both stages, that may be concluded as confirmatory finding of a candidate gene. The two-stage approach should be more powerful because fewer tests are performed overall and hence the multiple-testing issue is mitigated. We will compare population-based case-control design, classical TDT, and the two-stage approach in terms of power and efficiency.

Impact of Measurement Errors in Haplotyping in DNA Pooling

◆ Hemant K. Tiwari, University of Alabama, Birmingham

University of Alabama, Birmingham, Ryals Public Health Bldg., Suite 327, 1665 University Blvd., Birmingham, AL 35294-0022

htiwari@uab.edu

Key Words: *association, candidate genes, complex diseases, DNA pooling, haplotyping, measurement errors*

DNA pooling has been proposed as an efficient method to perform large scale association studies for a genome scans or candidate gene analyses with a dense set of markers. The use of DNA pooling has potential for gene-mapping and reduces the genotyping cost considerably. Although, there is a substantial reduction of cost in genotyping, there are several methodological issues that can confound the findings. The estimate of an allele frequency in a pooling experiment is subject to errors due to DNA quantification and formation of pools; amplification of target sequence; frequency estimation with chosen methodology; and sampling variance. It has been shown that haplotype analysis has greater statistical power than single point analysis. However, the extent to which misspecification of allele frequency estimate due to DNA pooling affects haplotype inference has not been rigorously investigated to date. We investigate effect of DNA pooling on haplotype frequencies estimation and then we investigate impact of misspecification of allele frequencies on Type I error and power of the haplotype association test.

327 New Methodology for Estimating Census Coverage Error Using Multiple Systems ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Wednesday, August 11, 10:30 am–12:20 pm

Multiple System Estimation with Erroneous Enumerations

◆ Paul P. Biemer, RTI International and UNC-CH; G. Gordon Brown, RTI International; Dean H. Judson, U.S. Census Bureau

RTI International and UNC-CH, PO Box 12194, Research Triangle Part, NC 27709-2194

ppb@rti.org

Key Words: *latent class analysis, population census, undercount, finite mixture models, correlation bias*

This paper presents a latent class modeling approach for multiple system estimation that accounts for varying levels of incompleteness and undetected erroneous enumerations in the population lists. Our approach assumes that one of the lists is based upon administrative records with errors that are assumed to be locally independent of the other enumeration-based lists. For k >2 lists, the resulting data take the form of an incomplete 2k contingency table which can be represented by a latent class model where the latent variable is an individual's residency true status (i.e., resident or nonresident of the population). Latent class analysis is used to estimate the expected values of the observed cells of this table and then to project these estimate onto the unobserved cells in order to estimate the total number of population members. Using artificial populations, we evaluate the improvement in mean squared error using this approach compared with other log-linear estimation approaches from the census undercount and capture-recapture literature.

Estimating Erroneous Enumerations in the U.S. Decennial Census Using Four Lists

◆ G. Gordon Brown, RTI International; Paul P. Biemer, RTI International and UNC-CH; Dean H. Judson, U.S. Census Bureau

RTI International, 3040 Cornwallis Rd., PO Box 12194, Research Triangle Park, NC 27709-2194

ggbrown@rti.org

Key Words: *census undercount, latent class analysis, capture-recapture, erroneous enumerations, heterogeneity*

We present a method for adjusting the U.S. decennial census day residency counts using a system of four lists: Census, PES, MER, ARL. Our primary objective is to develop a method that can detect erroneous enumerations that are present in the data since failure to account for erroneous enumerations in the lists will result in overestimation of the census day residents. Indicators of residency on the lists are used to obtain estimates of erroneous enumerations in a manner similar to one used in capture-recapture experiments to detect heterogeneity. The erroneous enumerations are assumed to come from a "nonresident" population that is rostered on the four lists with different probabilities than the actual census day "resident" population. Using the four lists, latent class analysis, and the software package LEM, we were able to fit several reasonable models that reflect likely census populations. We discuss the assumptions required for the inference to be valid and indicate the robustness of these models to the violation of these assumptions. The results from a simulation study are used to show when our method is reliable and situations in which it fails.

Individual-level Multiple Systems Estimation: Predicting Census Day Residency

◆ Elizabeth A. Stuart, Harvard University; Alan Zaslavsky, Harvard Medical School

Harvard University, Dept. of Statistics, 1 Oxford St., Cambridge, MA 02143

stuart@stat.harvard.edu

Key Words: *administrative records, multiple systems estimation, census enumeration, SOUP prior, Bayesian hierarchical model*

Administrative records are a promising data source for census enumeration, but a major challenge is determining which administrative records represent Census Day residents, given the interval between the administrative record dates and Census Day. Using a Bayesian hierarchical model of migration and observation in the record systems (including the census), we predict census day residency for each individual and generate estimates of Census Day population size. We extend multiple systems estimation by modeling individual-level behavior and utilizing the full information in the record systems, including covariate and date information. Using the single observation unbiased prior (SOUP) of Meng and Zaslavsky (2002) we also predict the number of unobserved individuals. Results are presented using simulated data, with file types and coverage rates based on StARS 2000 and Census 2000 files; the file systems simulated are census records as well as the StARS 2000 record systems with reliable date information. Coverage rates of Census Day population size and file coverage rates are compared when the unobserved individuals are and are not utilized in the Census Day population estimates.

Multiple Systems Estimation Methods in the Context of a Combination of Census-based and Sample-based Lists

◆ Jana Asher, Carnegie Mellon University; Stephen E. Fienberg, Carnegie Mellon University

Carnegie Mellon University, 26 Underwood Place NE, Washington, DC 20012

asher@stat.cmu.edu

Key Words: *multiple systems estimation, decennial census, administrative records*

This paper describes research into potential multiple systems estimation techniques in which two sources of data (for example, the decennial census and an administrative records list) are considered primary population lists, and a third data source (for example, a post-enumeration survey) is treated as an auxiliary random sample of blocks from a third population list. We discuss the advantages and disadvantages of this approach, and the issues surrounding the use of lists that reference different time frames. We conclude with a discussion of how this research, combined with other multiple systems research presented in this session, might be used in the future in the context of the U.S. decennial census.

328 Models for Decision Support on Colorectal Cancer Screening and Prevention: Why Microsimulation and What Assumptions?

Section on Health Policy Statistics, ENAR

Wednesday, August 11, 10:30 am–12:20 pm

Microsimulation Modeling of Colorectal Cancer for Assessments to Guide Health Policy: Model Developments and Applications

◆ Martin L. Brown, National Cancer Institute; Eric Feuer, National Cancer Institute

National Cancer Institute, 6130 Executive Blvd. EPN 4005, Bethesda, MD 20892-7344

mb53o@nih.gov

Key Words: *modeling, microsimulation, cancer, screening, surveillance*

Colorectal cancer (CRC) is the second leading cause of cancer death in the United States; it is also one of the most preventable cancers. CRC develops when normal colonic epithelium gives rise to precancerous adenomatous polyps, which may grow and ultimately become invasive cancer. Mathematical modeling can describe this adenoma-carcinoma sequence but must be complex enough to describe the underlying progression and regression of disease and how screening and lifestyle factors can affect the disease process. Microsimulation models are characterized by simulation of individual event histories for a hypothetical population and can be used to answer questions about these complex processes including the potential impact of interventions. We present three microsimulation models developed as part of the National Cancer Institute's Cancer Intervention and Surveillance Monitoring Network that take into account lifestyle risk factors and regression of precursor lesions. The models are calibrated using observed data. We use the models to describe and explain CRC incidence and mortality trends in the United States 1975-2000 as a function of changes in screening.

Estimating Adenoma Risk for Microsimulation Modeling

◆ Onchee Yu, Group Health Cooperative Center for Health Studies; Carolyn M. Rutter, Group Health Cooperative Center for Health Studies; Diana Miglioretti, Group Health Cooperative Center for Health Studies; Margaret Mandelson, Group Health Cooperative Center for Health Studies

Group Health Cooperative Center for Health Studies, 1730 Minor Ave., Suite 1600, Seattle, WA 98101-1466

yu.o@ghc.org

Key Words: *meta-analysis, nonhomogenous Poisson, extra-Poisson variability, MCMC*

The risk of developing an adenoma is a key input for microsimulation models of colorectal cancer. Most colorectal cancers arise from adenomas and discovery of adenomas influences future screening. We estimate adenoma risk by combining information from autopsy studies using Bayesian meta-analysis. We use a multinomial distribution for counts with bin probabilities based on a nonhomogeneous Poisson model for individual risk. This allows us to combine information across studies that use different groupings to summarize adenoma counts. Risk is modeled as a function of age and sex and included pseudo-individual-level random effects to incorporate extra-Poisson variability. Models were estimated using WinBugs software. We assessed model fit by comparing posterior predicted values to observed data and validated our model by comparing results from colonoscopy studies to model-based estimates. We estimate that the overall probability of at least one adenoma increases from 0.24 at age 50 to 0.30 at age 60 and 0.36 at age 70. Men were more likely have an adenoma than women (e.g., prevalence at age 60 of 0.36 vs 0.25).

Using Monte Carlo Simulation to Estimate the Effect of Aspirin Use on Progression of Colorectal Cancer

◆ Karen M. Kuntz, Harvard School of Public Health; Jane C. Weeks, Dana-Farber Cancer Institute; Milton C. Weinstein, Harvard School of Public Health; Graham A. Colditz, Brigham and Women's Hospital

kmk@hsph.harvard.edu

Key Words: *colorectal cancer, simulation model*

We used a novel combination of epidemiological analysis and Monte Carlo simulation to infer the effects of aspirin use on unobservable states of colorectal disease. Using a state-transition model of colorectal disease in which aspirin use "slows" disease progression, we simulated a variety of potential aspirin effects and compared their influence on diagnosed colorectal cancer (CRC) with the association between duration of aspirin use and diagnosed CRC observed in a large prospective study. We modified the aspirin effects and repeated the simulations until the simulated relationship between duration of aspirin use and diagnosed cancer was consistent with the observed relationship. The simulation results suggest that aspirin is unlikely to reduce CRC risk solely by decreasing adenoma incidence or solely by slowing adenoma progression to malignancy. Instead, it appears that aspirin acts at multiple points in the disease by exerting a strong influence on adenoma emergence and a relatively weaker influence on adenoma progression. Simulation modeling can play an important role in improving our understanding of disease and informing clinical and preventive guidelines.

Calibration of Microsimulation Models: A Likelihood-based Approach

◆ Amy Knudsen, Harvard School of Public Health; Milton C. Weinstein, Harvard School of Public Health; Karen M. Kuntz, Harvard School of Public Health

Harvard School of Public Health, 718 Huntington Ave., 2nd Floor, Boston, MA 02115

abird@fas.harvard.edu

Key Words: *microsimulation, model calibration, likelihood, colorectal cancer*

Modeling unobserved processes is challenging since limited data are available to directly inform the transitions between unobserved states. Such models can be estimated by constraining model outcomes to calibrate to observable data. We explored a likelihood-based approach to the calibration of a natural history model of colorectal cancer (CRC). We calibrated the model by simulating life histories under a given set of parameters and comparing model outcomes with data on the prevalence, location, and size of adenomas from autopsy and screening studies, and the incidence, stage, and location of CRC from a cancer registry. We assumed each set of data follows a binomial distribution and calculated two likelihoods for each measure: (1) likelihood of generating the data with a given set of parameters (i.e., observed likelihood), and (2) likelihood if the model exactly predicted the data (i.e., maximum likelihood). An overall GOF score was calculated as the sum of the differences in -2 log likelihoods. The parameter space was explored using the Nelder-Mead Simplex algorithm. Likelihood-based methods are valuable tools for model calibration and provide insight into limitations in model structure.

Microsimulation Modeling to Estimate the Impact of Cancer Prevention Strategies for Healthy People 2010

◆ Ann G. Zauber, Memorial Sloan-Kettering Cancer Center; Marjolein van Ballegooijen, Erasmus University; Iris Vogelaar, Erasmus University; Rob Boer, RAND Corporation; Deb Schrag, Memorial Sloan-Kettering Cancer Center; Eric Feuer, National Cancer Institute; Martin L. Brown, National Cancer Institute

Memorial Sloan-Kettering Cancer Center, 307 East 63rd St. Room 357, Dept. of Epidemiology and Biostatistics, New York, NY 10021

zaubera@mskcc.org

Key Words: *microsimulation modeling, colorectal cancer prevention, Healthy People 2010, CISNET*

The Healthy People 2010 (HP 2010) goals serve as the Department of Health and Human Services blueprint for achieving increased years of healthy life and eliminating health disparities by 2010. We use microsimulation modeling with the MISCAN model to project how best to achieve the HP 2010 goal of a 35% reduction in CRC mortality for the U.S. by 2010. We model treatment interventions of new chemotherapy regimes through the relative survival for each stage of disease; screening interventions through the stage of colorectal cancer detected and by detection and removal of adenomatous polyps, and lifestyle modifications through the risk of developing adenomas and the rate of progression of the adenoma to carcinoma. The estimated impact of each type of intervention is derived for the U.S. population to ascertain the anticipated magnitude of effect and whether the stated goal is obtainable. The results of this simulation model were presented to the NCI in June 2004 to assess the potential for reaching the HP 2010 goals for colorectal cancer mortality reduction and to suggest whether increased levels of interventions are required to meet the HP 2010 goal.

329 Software for Time Series Modeling and Seasonal Adjustment

Business and Economics Statistics Section

Wednesday, August 11, 10:30 am–12:20 pm

New Graphics for Seasonal Adjustment

◆ Catherine Hood, U.S. Census Bureau; Roxanne Feldpausch, U.S. Census Bureau

catherine.c.hood@census.gov

Key Words: *time series, X-12-ARIMA, SEATS*

Progress in seasonal adjustment depends on the development of new seasonal adjustment methods, but it also depends on the development of diagnostics for the adjustment. Two widely used seasonal adjustment programs are the U.S. Census Bureau's X-12-ARIMA and Agustin Maravall's SEATS. In previous studies with SEATS and X-12-ARIMA, we have been looking at the properties of a time series that make it a good candidate for adjustment by SEATS or by X-12-ARIMA. As part of our research into various diagnostics in SEATS and X-12-ARIMA, we have developed some new, graphical ways to look at seasonal adjustment diagnostics. We used a version of X-12-ARIMA that has access to the SEATS algorithm. This allowed us to compute similar diagnostics for both programs, including sliding spans and revision diagnostics, so we can compare adjustments between the two programs.

Model Simplification after the Automatic Modeling Procedure of X-12-ARIMA 0.3

◆ Kathleen M. McDonald-Johnson, U.S. Census Bureau; Catherine Hood, U.S. Census Bureau; Roxanne Feldpausch, U.S. Census Bureau

U.S. Census Bureau, ESMPD / TSMS, Rm 3112/4 Mail Stop 6200, Washington, DC 20233-6200

kathleen.m.mcdonald.johnson@census.gov

Key Words: *regARIMA model, seasonal adjustment, time series*

The automatic modeling procedure of X-12-ARIMA version 0.3 is based on the automatic modeling procedure of TRAMO. At the U.S. Census Bureau, we use automatic modeling to help identify regARIMA models for forecast extension and estimating regression effects. The automatic procedure may identify mixed models that are generally too complicated for a concurrent adjustment production setting. We present the results of a study comparing the performance of the automatically identified model to a possibly simplified version of the model and to an airline model. We compared models using goodness-of-fit diagnostics, within-sample and out-of-sample forecasts, and revision history diagnostics.

Further Developments in X-12-SEATS

◆ Brian C. Monsell, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 9100, Washington, DC 20233

brian.c.monsell@census.gov

Key Words: *model-based signal extraction, outliers, intervention effects, seasonal adjustment diagnostics*

In collaboration with the current developers of the SEATS seasonal adjustment program, an experimental version of X-12-ARIMA that produces model-based seasonal adjustments from SEATS has recently been made available to users. This program allows users of seasonally adjusted series to generate X-11 and SEATS sesonal adjustments using the same interface, and compare these seasonal adjustments using a common set of diagnostics. This session will demonstrate how SEATS adjustments are integrated into the X-12-ARIMA procedure and give examples of new modeling options such as seasonal outliers and pulse regressors.

An Implementation of Component Models for Seasonal Adjustment Using SsfPack Software

◆ Siem Jan Koopman, Free University Amsterdam

Free University Amsterdam, FEWEB, De Boelelaan 1105, Amsterdam, NL 1081 HV The Netherlands

koopman@tinbergen.nl

Key Words: *seasonal adjustment, time series, outliers, ARIMA models, canonical decomposition, unobserved components*

An alternative to traditional methods of seasonal adjustment is to use component time series models to perform signal extraction, such as the structural models of Harvey currently implemented in STAMP, or the ARIMA decomposition models of Hillmer and Tiao currently used in SEATS. A flexible implementation allowing easy specification of different models has been developed using the SsfPack software module of the Ox matrix programming language. This allows the incorporation of heavy-tailed distributions into certain components within the model. Examples of robust seasonal adjustments for different model types using this method will be shown.

Structural Time Series Modeling Using SAS 9.1

◆ Rajesh Selukar, SAS Institute Inc.

SAS Institute Inc., SAS Campus Dr., Cary, NC 27513

rajesh.selukar@sas.com

Key Words: *time-series, seasonal adjustment, state space, structural models*

PROC UCM is a new procedure in SAS\ETS 9.1 that provides a very flexible environment for analyzing time-series data using the Structural Time Series models, also called the Unobserved Components Models (UCM). These models represent the observed series as a sum of suitably chosen components such as trend, seasonals, cycles, and regression effects. You can use the UCM procedure to formulate very comprehensive models that bring out the salient features of the series under consideration. Structural models are applicable in the same situations where the Box-Jenkins ARIMA models are applicable; however, the structural models tend to be more informative about the underlying stochastic structure of the series. The UCM procedure can be used for time-series forecasting, model-based seasonal adjustment, and for detection of structural change. It provides extensive model diagnostics and has good plotting capabilities. The UCM procedure will be illustrated using a few examples.

330 Generalized Linear Models with Overdispersion and Zero-inflation

SSC

Wednesday, August 11, 10:30 am–12:20 pm

A Model Selection Criterion for Marginal Zero-inflated Regression for Clustered Data

◆ Daniel B. Hall, University of Georgia; Zhengang Zhang, University of Georgia

University of Georgia, Dept. of Statistics, Athens, GA 30602-1952

dhall@stat.uga.edu

Key Words: *GEE, zero-inflated Poisson, information criterion, longitudinal data, repeated measures, mixture model*

Applying methodology of Rosen, Jiang, and Tanner, Hall and Zhang have recently described methods for fitting marginal versions of zero-inflated regression models to clustered data via the expectation-solution algorithm. This approach involves replacing the estimating equations solved in the M step of the usual EM algorithm for zero-inflated regression under independence with generalized estimating equations (GEEs). These GEEs involve working correlation matrices to account for within-cluster correlation. We propose a model selection criterion which can aid in the selection of appropriate working correlation matrices as well as in the selection of covariates for the linear predictors of the model. The new criterion is akin to Akaike's information criterion (AIC) and also to Pan's adaptation of AIC to the GEE context. The performance of the new criterion is examined via simulation. In addition, the extension of this approach to the more general finite mixture of marginal models context of Rosen et al.'s (2000) paper is considered.

Generalized Linear Model, Zero-inflation, and Overdispersion

◆ Dianliang Deng, University of Regina

University of Regina, 3737 Wascana Pkwy., Dept. of Mathematics and Statistics, Regina, SK S4S 0A2 Canada

deng@math.uregina.ca

Key Words: *binomial models, generalized linear model, overdispersion, Poisson model, score test, zero-inflation*

Discrete data in the form of counts often exhibit extra variation that can not be explained by a simple model, such as the binomial or the Poisson model. Also, these data, sometimes, show more zero counts than what can be predicted by a simple model. Therefore, a discrete generalized linear model (Poisson or binomial) may fail to fit a set of discrete data either because of zero-inflation or because of overdispersion or because there is zero-inflation as well as overdispersion in the data. We deal with the class of zero-inflated overdispersed generalized linear models and propose procedures based on score tests for selecting a model that fits such data. We show that in certain cases and under certain conditions the score tests derived using the general overdispersion model and those developed under specific overdispersion models are identical. Empirical level and power properties of the tests are examined by a limited simulation study. Two illustrative example are given.

Semiparametric Regression Models for Zero-inflated Data

◆ Sujit K. Ghosh, North Carolina State University

North Carolina State University, Raleigh, NC 27695-8203

sghosh@stat.ncsu.edu

Key Words: *overdispersion, regression models, semiparametric, zero-inflated data*

In recent years there has been considerable interest in models for count data that allow for overdispersion due to excess zeros. These models (e.g., ZIP) complement more conventional models for overdispersion that are based on the mean-variance relationship. However, most of the regression models for zero-inflated data involve full parametric specification of the distribution of counts and hence these models may fail to address various features of count data. The proposed semiparametric model is flexible and accounts for overdispersion due to excess zeros. Inference is based on estimating equations derived under a semiparametric framework. Simulation studies are provided to validate the use of proposed model and a real data application is presented for illustration.

Mixture of Zero-inflated Densities

◆ Jean-Francois Angers, Universite de Montreal; Atanu Biswas, Indian Statistical Institute, Kolkata

Universite de Montreal, Dept. de mathematiques et de statistique, C.P. 6128, Succ. centre-ville, Montreal, PQ H3C 3J7 Canada

jean-francois.angers@umontreal.ca

Key Words: *zero-inflated model, mixture of densities, zero-inflated Poisson, EM algorithm, Bayesian analysis*

In several real-life examples one encounters count data where the number of zeros is such that the usual Poisson distribution does not fit the data. Quite often the number of zeros is large, and hence the data is zero-inflated. Furthermore, the histogram is often multimodal indicating that the data come from different sub-populations. In such a situation, a mixture of some discrete probability density functions including some zero-inflated Poisson model(s) can be considered and a Bayesian analysis can be carried out. Using the EM algorithm, Bayesian estimates for the different parameters and also the predictive density are obtained. The techniques are illustrated using a real-life dataset.

Test of Treatment Effect in Pre-drug and Post-drug Count Data with Zero Inflation

◆ Xing Jiang, University of Windsor

University of Windsor, Dept. of Mathematics and Statistics, 401 Sunset Ave., Windsor, ON N9B 3P4 Canada

jiang4@uwindsor.ca

Key Words: *likelihood ratio test, PVC data, score test, treatment effect*

A procedure for testing for treatment effect in data similar to the data on premature ventricular contractions (PVC) (Berry) is presented. We consider a zero-inflated beta-binomial model. Based on this model we develop score tests to test for treatment effect in the data in which observations in the form of counts are recorded before and after applying a therapy. Results of a small simulation experiment, to study small sample behavior of a score test and a likelihood ratio test, are reported and the PVC data are analyzed. Both the score and the likelihood ratio tests show good level properties. Either the score tests or the likelihood ratio tests can be used for testing the presence of treatment effect. The score tests, however, may be preferable because they use estimates of the parameters only under the null hypothesis.

331 Topics in Response Surface Designs II ⌘

Section on Quality and Productivity

Wednesday, August 11, 10:30 am–12:20 pm

Assessing Designs for Mixture Experiments for the Generalized Linear Models Case

◆ Ayca Ozol-Godfrey, Virginia Polytechnic Institute and State University; Christine Anderson-Cook, Virginia Polytechnic Institute and State University; Edward L. Boone, University of North Carolina, Wilmington

Virginia Polytechnic Institute and State University, Dept. of Statistics, Hutcheson Hall, Blacksburg, VA 24061-0439

aozol@vt.edu

Key Words: *response surface, mixture experiments, FDS Plot, non-normal data, robustness to parameter misspecification*

A mixture experiment is a special type of a response surface experiment where the factors are the ingredients or the components of a mixture, and the response depends on the relative proportions of the components. Frequently, while examining these designs, normality for the response is assumed, and the linear model form is used. However, recently different types of non-normal data and the ways to examine them have been studied. We will study mixture experiments using Generalized Linear Models (GLM) for a non-normal response. The Fraction of Design Space (FDS) Plot, a new graphical design evaluation tool by Zahran, Anderson-Cook, and Myers (2003) will be reviewed, and it will be adapted to examine the mixture designs for the GLM case. A variety of types of robustness to initial parameter estimates will be evaluated using these plots. The methods will be illustrated with several examples.

Comparing the Performance of Crossed-array and Single-array Approaches to Robust Design

◆ Daniel D. Frey, Massachusetts Institute of Technology; Xiang Li, Massachusetts Institute of Technology

Massachusetts Institute of Technology, 77 Massachusetts Ave. Room 3-449D, Cambridge, MA 02139

danfrey@mit.edu

Key Words: *robust design, design of experiments*

This paper presents a new way of comparing the effectiveness of robust design methods and applies it to two well-known alternatives—crossed arrays and single arrays. The relative merits of both approaches have been examined in the literature and single arrays are generally considered the preferred alternative. A large number of simulated systems are sampled from a relaxed weak heredity model so that the systems will exibit the properties of effect sparsity, hierarchy, and inheritance. Crossed-array and single-array methods are applied to the systems in an effort to reduce the variance of the response. The percentage of variance reduction achieved on average across the sampled systems is taken as a measure of performance of the method. The simulations indicate that the single array approach performs well if the probability of three-way interactions is zero, but degrades if there is a realistic likelihood of control by control by noise interactions. By contrast, crossed-array approaches performed well across a broader range of system types.

A Bayesian Approach to Dual Response Surface Modeling

◆ Younan Chen, Virginia Polytechnic Institute and State University; Keying Ye, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, 406 A Hutcheson Hall, Dept. of Statistics, Blacksburg, VA 24061

yochen@vt.edu

Key Words: *dual response surface, Bayesian method*

In modern quality engineering, dual response surface methodology is a powerful tool to monitor an industrial process by using both the mean and standard deviation of the measurements as the responses. Least squares method in regression is often used to estimate the coefficients in the mean and standard deviation models, and various decision criteria are proposed by researchers to find the optimum conditions. We discuss a Bayesian approach to investigate dual response surface modeling. The comparisons of the results derived from both Bayesian and non-Bayesian regressions are made by using simulation data and a real dataset.

Response Model vs. Dual Response Model Methods: Relative Efficiencies of Estimators

◆ Mary McShane-Vaughn, Southern Polytechnic State University; Russell Heikes, Georgia Institute of Technology

Southern Polytechnic State University, 1100 S Marietta Pkwy, IET Dept., Marietta, GA 30060

mvaughn@spsu.edu

Key Words: *robust design, design of experiments, response model, dual response model*

We contrast the performance of the response model to that of another popular robust design method, dual analysis. The relationship between the relative efficiencies of the methods' estimators and the underlying model parameters will be determined in general, and then presented for a specific design choice. Using relative efficiency as the criterion, recommendations concerning the appropriate use of the response model method will be made.

The Practical Study on Response Surface Methodology

◆ Chosei Kaseda, Yamatake Corporation

Yamatake Corporation, 1-12-2 Kawana, Fujisawa-shi, 251-8522 Japan

kaseda-chosei@jp.yamatake.com

Key Words: *response surface methodology, black-box modeling, radial basis function, thin plate spline, artificial neural network*

Response Surface Methodology is one of effective statistical approaches for optimum design using experimental data. But conventional RSM has some practical problems on generating response surface. Generally, polynomial model is applied to experimental data based on Design-of-Experiment. But it sometimes becomes difficult to easily generate the surface when data is not ideal DOE data due to experimental limitations. Also it becomes more difficult in the case of complex nonlinear targets. Neural network is also sometimes used. But it requires model-tuning by trial-and-error. So the conventional RSM does not always improve the efficiency of product design works because of requirement of much labor and time for surface-generation. To solve this problem, we apply multivariate spline based on radial basis function to surface-generation. This RSM can contribute to making more efficient design on practical use. Practical problems of conventional RSM shall be discussed first. Then, as a means of solving those problems, our RSM and developed software shall be introduced. Lastly, we'll show results of applications to pharmaceutical design and electrical device design, etc.

332 Statistics in Veterinary Medicine

Section on Bayesian Statistical Science

Wednesday, August 11, 10:30 am–12:20 pm

Bayesian Inference for Single Case Designs

◆ Richard Evans, Iowa State University

Iowa State University, VDPAM, College of Veterinary Medicine, Ames, IA 50011

revans@iastate.edu

Key Words: *Bayesian, changepoint*

Single case statistical inference has been investigated and implemented in the behavioral sciences for many years. However, there has been little use of this method in other sciences and statistical development has been limited. Single case designs can be useful for research on rare or expensive animals or diseases, where sufficient sample sizes required for traditional statistical methods may be difficult to obtain. Typical single case designs apply and withdraw interventions (e.g., ABAB or ABC) over time. Our approach is to use Bayesian changepoint methods to calculate the posterior probability that the intervention affected the subject and changed the outcome during the course of treatment relative to other treatments or baselines. Note that observations within a treatment time may not be exchangeable and the analysis may require different numbers of cutpoints (e.g., A|BC and A|B|C). We apply this method to data obtained from an emu that was trained to walk on a pressure platform.

Modeling Genetic Change in the PRRS Virus in Pigs

◆ Philip Dixon, Iowa State University

Iowa State University, Dept. of Statistics, 120 Snedecor Hall, Ames, IA 50011-1210

pdixon@iastate.edu

Key Words: *mutation, hierarchical model, latent abundance*

The PRSS virus causes catastrophic losses in production pig operations. The virus and disease were unknown prior to the late 1980s. The epidemiology of the disease suggests that the virus genome mutates rapidly. Mutation rates were estimated by in-vivo passage of the virus through seven generations of pigs. Each generation, the viral pool was sampled by sequencing 15 viral clones. This approach underestimates genetic change because rare clones are unlikely to be detected. A model for the genetic change can be constructed as a Bayesian hierarchical model. This combines a model for the mutation process, a model for the abundance of each sequence over time, and a model for the sampling process. Such a model can account for unobserved mutations. Inference is by MCMC. The model will be validated by comparing data collected from generations 8-15 to model projections.

Exploratory Relative Risk-mapping Applied to Regional Data on Echinococcus Multilocularis Infections in Red Foxes

◆ Olaf Berke, University of Guelph

University of Guelph, Dept. of Population Medicine, Ontario Veterinary College, Guelph, ON N1G 2W1 Canada

oberke@uoguelph.ca

Key Words: *veterinary, epidemiology, Bayes, kriging, cluster, scan test*

Spatial epidemiological investigations are often exploratory with limited knowledge about the putative risk factors. Indeed, often the primary motivation for the spatial epidemiological analysis is to identify unknown geographically varying risk factors. An exploratory approach to mapping the spatial relative risk is to scale the risk map by the background risk of the unexposed population. Exposure to unknown spatial risk factors is defined via specific cluster analysis using the spatial scan test. Identification of spatial disease clusters separates the population spatially into inside and outside high risk areas, i.e., the exposed and unexposed populations. To account for unequal regional populations at risk the data were spatially smoothed by empirical Bayesian estimation. Then the latent risk surface map is generated from regional data by universal kriging. This exploratory approach to relative risk-mapping gives the investigator an impression about the importance and geographical distribution of the unknown spatial risk factors. The approach is applied to data on infections of Red foxes with Echinococcus multilocularis from 43 administrative regions in Lower Saxony, Germany.

Bayesian Nonparametric ROC Curve Estimation

◆ Adam Branscum, University of California, Davis

University of California, Davis, One Shields Ave., Davis, CA 95616

ajbranscum@ucdavis.edu

Key Words: *Polya trees*

Nonparametric and semiparametric models are used to provide flexible inferences for example relaxing unimodality and other consequences of strong distributional assumptions. A standard epidemiologic problem involves the estimation of receiver operating characteristic curves based on serologic data that have been collected on diseased and nondiseased individuals respectively. There are many frequentist parametric and nonparametric solutions to this problem in the literature. We illustrate the Bayesian nonparametric approach to ROC curve estimation, in the absence of a gold standard test based on a single sample of serologic outcomes with unknown disease status. In addition, we are able to estimate the mixing proportion, or prevalence of disease in the sampled population, using this approach.

A Bayesian Model for Livestock Disease Infection Prevalence Rates

◆ Eric A. Suess, California State University, Hayward; Ian A. Gardner, University of California, Davis; Wesley O. Johnson, University of California, Davis

California State University, Hayward, Dept. of Statistics, 25800 Carlos Bee Blvd., Hayward, CA 94542

esuess@csuhayward.edu

Key Words: *disease surveillance, prevalence estimation, Bayesian, MCMC*

For livestock disease infection surveillance, we develop a Bayesian model for estimating infection rates. By incorporating expert prior opinion related to the animal disease along with test results from animals sampled within herds we produce posterior estimates of infection rates. The model is used to estimate the within-herd prevalence and the average within-herd prevalence, the prevalence of infected herds, and the regional prevalences. Correlation between the infection status of animals within herds part of the model. We use data from a national survey of Johne's Disease. The model is implemented using WinBugs and the MCMC output is analyzed in R.

333 Bayesian Design and Analysis for Biomedical Data ▲

Section on Bayesian Statistical Science

Wednesday, August 11, 10:30 am–12:20 pm

A Statistical Model for Evaluating Synergy

◆ Gary L. Rosner, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030

glr@odin.mdacc.tmc.edu

Key Words: *Bayesian methods, isobolograms, synergy, dose-response, drug development*

Quite often, drugs are combined with the hope that the joint effect is better for the patient than the effect that can be safely achieved by each drug alone. An important question in drug development concerns maximizing the effect of drug therapy and learning if two or more agents are synergistic with respect to their biological effect. One may also be concerned with determining if drugs are antagonistic, i.e., reducing biologic effect when combined. We discuss estimation of synergy and antagonism from a Bayesian perspective.

Measures of Explained Variance and Pooling in Multilevel Models

◆ Iain Pardoe, University of Oregon; Andrew Gelman, Columbia University

University of Oregon, Lundquist College of Business, Eugene, OR 97403-1208

ipardoe@lcbmail.uoregon.edu

Key Words: *shrinkage, adjusted R-squared, Bayesian inference, hierarchical model, multilevel regression, partial pooling*

Explained variance (R-squared) has been generalized in various ways to multilevel models for hierarchical data structures in which individuals are grouped into units, and there are variables measured on individuals and each grouping unit. The models are based on regression relationships at different levels, with the first level corresponding to the individual data, and subsequent levels corresponding to between-group regressions of individual predictor effects on grouping unit variables. We present an approach to defining R-squared at each level of the multilevel model, rather than attempting to create a single summary measure of fit, by comparing variances within the model. In simple regression, our measure generalizes the classical adjusted R-squared. We also discuss a related variance comparison to summarize the degree to which estimates at each level of the model are pooled together based on the level-specific regression relationship, rather than estimated separately. This pooling factor is related to the concept of shrinkage in simple hierarchical models. We illustrate the methods on a dataset of radon in houses within counties using a series of multilevel models.

Bayesian Screening Tool for Clinical Safety Data

◆ Jayanti Gupta, Merck & Co., Inc.; Saurabh Mukhopadhyay, Merck & Co., Inc.

Merck & Co., Inc., Rahway, NJ 07065

jayanti_gupta@merck.com

Key Words: *Bayesian, clinical trials, hierarchical models*

Developing a unified screening tool to compare adverse experiences among treatment groups that occur in clinical trials is of considerable interest. Since Aes occur within body systems, it is also of interest to see if particular body systems are of more concern than others with respect to the treatments being tested. Bayesian hierarchical mixed models in the manner outlined by Berry (2002) holds promise, as it accommodates simultaneous estimation of several Aes occurring in the same body system. However, different choices of model development exist and their relative long-term performance needs to be examined. Different choices of hierarchical models will be studied through simulations, and methods will be proposed to provide consistent and long-term support to analyze safety data.

Posterior Simulation with Polya Tree Priors

◆ Ying Yang, University of Texas M.D. Anderson Cancer Center; Peter Mueller, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Dept. of Biostatistics, Houston, TX 77030

yingyang@odin.mdacc.tmc.edu

Key Words: *regression, Polya tree, Markov chain Monte Carlo*

The error distribution in the standard regression model is modeled as a Polya tree prior constrained to have median 0. A noninformative prior is placed on the regression parameters. We discuss several alternative methods to implement posterior simulation: finite Polya tree, Polya tree predictive density, or Polya tree marginal density. Markov chain Monte Carlo algorithms are described for each approach.

Analysis of Ordinal Longitudinal Data

◆ Xian Zhou, University of Texas M.D. Anderson Cancer Center; Benjamin N. Bekele, University of Texas M.D. Anderson Cancer Center; Peter Mueller, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77030

xianz@mdanderson.org

Key Words: *mixture model, probit model, moving average model, Dirichlet prior, reversible jump Markov chain Monte Carlo*

Albert & Chib proposed a Bayesian ordinal probit regression model using the Gibbs sampler. Their method defines a relationship between latent variables and ordinal outcomes using cut-point parameters. However, the convergence of this Gibbs sampler is slow when the sample size is large because the cutpoint parameters are not efficiently sampled. Cowles proposed a Gibbs/Metropolis-Hastings (MH) sampler that would update cutpoint parameters more efficiently. In the context of longitudinal ordinal data, these algorithms might require the computation of a multivariate normal cumulative probability function to calculate the acceptance probability of MH sampler. We introduce a mixture of probit model with latent variables following a moving average model. This mixture model can successfully model the ordinality of the data while holding constant the cutpoint parameters. Gibbs samplings under a Dirichlet prior and reversible jump Markov chain Monte Carlo are carried out to estimate a known or unknown number of components of a mixture model, respectively.

334 Quality Assessment of Administrative Records Data ▲ ⊞

Section on Government Statistics, Section on Quality and Productivity, Social Statistics Section, Section on Survey Research Methods

Wednesday, August 11, 10:30 am–12:20 pm

Assessing Industry Codes on the IRS Business Master File

◆ Paul B. McMahon, Internal Revenue Service

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

paul.b.mcmahon@irs.gov

Key Words: *administrative records, NAICS*

The Internal Revenue Service uses a variation on the North American Industry Classification System (NAICS) Codes that generally collapses the six-digit classifications to reduce respondent burden. The data are self-reported and subject to data abstraction errors, but those that are among the list of codes IRS uses are retained on the Business Master File. We will compare these validated codes to the values assigned during the Statistics of Income Programs data abstraction process. In addition, some of the "invalid" codes may, in fact, be NAICS Codes from the full Census listing, and we investigate whether this is confirmed by the SOI data edits.

Data Interpretation across Sources: A Study of Form 990-PF Information Collected from Multiple Databases

◆ Melissa R. Ludlum, Internal Revenue Service

Internal Revenue Service, PO Box 2608, Statistics of Income, Washington, DC 20013

melissa.r.ludlum@irs.gov

Key Words: *data quality, consistency, discrepancies, data collection methods, correction procedures*

Many private foundation data come from a single source, the information return, Form 990-PF. Forms 990-PF are available to the public for analysis and inspection via the internet. Additional statistics are created by the National Center for Charitable Statistics at the Urban Institute and are based on annual extracts from the IRS Returns Transaction File (RTF), which are provided to NCCS by IRS staff. A third resource is the Statistics of Income study of private foundations, which is based on a weighted annual sample of Forms 990-PF that are key-entered by IRS staff and reviewed for mathematical and reporting errors. However, the statistics derived from Forms 990-PF may vary widely, depending on the manner in which the data are collected. The paper will compare Form 990-PF data on the SOI sample file to those from the annual RTF extracts and make efforts to explain discrepancies. In addition, data from returns filed by a sample of large private foundations will be obtained from the internet and further compared to the other sources to evaluate the effects of IRS data entry and correction procedures on the various statistics.

The Effect of Content Errors on Bias and Nonsampling Error in Estimates Derived from Samples of Administrative Records

◆ Barry W. Johnson, Internal Revenue Service; Darien B. Jacobson, Internal Revenue Service

Internal Revenue Service, Statistics of Income, PO Box 2608, Washington, DC 20013

barry.w.johnson@irs.gov

Key Words: *bias, nonsampling error, administrative records*

The Statistics of Income Division of the IRS uses a number of methods for ensuring the quality and integrity of the data it produces for tax administration research. As a first line of quality assurance, codes and mathematically related fields are extensively tested as they are entered into computer systems. In addition, for a subsample of returns selected and processed in most studies, SOI assigns a second employee to enter and edit the data. Values from the original and second versions are then computer matched. A supervisor resolves discrepancies discovered during the match. The original value, second value, and correct values are all collected as a part of the quality review system, along with a set of codes that describe the cause of the error. This paper will use quality review data from federal estate tax returns selected into the calendar year 2002 Statistics of Income (SOI) estate tax sample to study editor errors in the SOI data. We will examine the effects of these nonsampling errors on the final estimates, with particular focus on identifying patterns that can be addressed during editor training.

Editor Judgment Effect: Modeling a Key Component of Nonsampling Error in Administrative Data

◆ Kimberly A. Henry, Internal Revenue Service; Yahia Z. Ahmed, Internal Revenue Service; Ellen Legel, Internal Revenue Service

Internal Revenue Service, Statistics of Income, PO Box 2608, Washington, DC 20013

Kimberly.A.Henry@irs.gov

Key Words: *data quality, reconciliation sample, bias, variance component estimation*

The presence of sampling and nonsampling errors can significantly affect an agency's national estimates. While sampling errors in Statistics of Income's (SOI) Corporation data are reported and published in the form of Coefficients of Variation (CVs), nonsampling errors are discussed, but not quantified. Instead, they are classified into two types: random errors whose effects may cancel out, and systematic errors whose effects tend to remain relatively fixed, resulting in bias. The abstraction of statistical information from administrative records places SOI's editors in a position to make judgments while recording, cleaning, and processing data. Errors created from these judgments are the largest source of nonsampling error in the corporate data. This paper will attempt to quantify the effect of editor judgment on the data using a quality review reconciliation sample.

The Evolution of IRS Telephone Quality Measures

◆ Laura R. Rosage, Internal Revenue Service

Internal Revenue Service, Statistics of Income, PO Box 2608, Washington, DC 20013

laura.r.rosage@irs.gov

Key Words: *defects per opportunity, accuracy, embedded quality*

The Interal Revenue Service has been measuring the quality of the services it provides to taxpayers since the late 1980s. The way in which telephone quality is measured has evolved over the years from test calls, to live monitoring of telephone calls, and soon, to whole contact recording. This paper details the Service's latest move from a pass/fail method of measurement to a defects-per-opportunity methodology. It discusses the strengths and limitations of each method, the overall impact on quality rates, and future plans for improvement to the measurement process.

335 Thar's Gold in Them Thar Hills: Benefits, Cautions, and Pitfalls for Training Novices to Use Large Archival Databases

Section on Statistical Education

Wednesday, August 11, 10:30 am–12:20 pm

Thar's Gold in Them Thar Hills: Benefits, Cautions, and Pitfalls for Training Novices to Use Large Archival Databases

◆ Susan C. Losh, Florida State University; ◆ Christopher M. Tavani, Florida State University; ◆ Carolyn Funk, Virginia Commonwealth University; ◆ Matthew DeBell, American Institutes for Research; ◆ Valerie Martin Conley, The Ohio University

Th Ohio University, 388 McCracken Hall, Athens, OH 45701

conleyv@ohio.edu

Key Words: *data archives, large databases, statistics training, data analysis*

Rather than collecting new data, many scholars now employ original analyses of archival data. The increasing proliferation of (particularly online) databases brings new riches to data analysis. However, analysts often neglect to explore—or may not know—the limitations, errors, and other pitfalls in secondary data analysis before they begin. We aim our experience in different sources of error and cautions to heed toward statistics and research methods educators training advanced undergraduates, graduate students, and practitioners in the social and behavioral sciences (including educational research). Our presenters are archive creators and analysts, and primary data collectors and designers. We discuss survey archives, national educational assessments (e.g., National Assessment of Educational Progress), and post-secondary institutional characteristics (e.g., SEStat or WebCASPAR). We hope more educators will use these resources, alerting their students to the benefits and training them to ask the fruitful questions to avoid the pitfalls in their use.

336 Consulting with CAM Investigators: Challenges, Opportunities, or Both? ▲

Section on Teaching Statistics in the Health Sciences

Wednesday, August 11, 10:30 am–12:20 pm

Consulting with CAM Investigators: Challenges, Opportunities, or Both?

◆ Cynthia R. Long, Palmer Center for Chiropractic Research; ◆ Christina M. Gullion, Kaiser Permanente Center for Health Research; ◆ Kim McFann, National Center for Complementary and Alternative Medicine

National Center for Complementary and Alternative Medicine, 6707 Democracy Blvd., Suite 401, MSC 5475, Bethesda, MD 20892

mcfannk@mail.nih.gov

Key Words: *biostatistics, complementary and alternative medicine, consulting, challenges*

This panel of PhD biostatisticians will discuss the challenges and opportunities in supporting research efforts in the area of complementary and alternative medicine (CAM) drawing from their collective experiences in working with CAM investigators. The panel includes (1) a faculty member of a research university biostatistics department who has a long history of consulting with CAM scientists, (2) an investigator in a federally funded CAM center within a research center of an HMO-environment, (3) a faculty member in a research center of a CAM institution, and (4) a biostatistician at the National Center for Complementary and Alternative Medicine at NIH. A major challenge of working with CAM investigators is that while they are typically credentialed providers in the therapy, they often have little or no training in research methods. An interesting aspect of CAM therapies is that many involve a multimodal, individualized, holistic approach to health care which is difficult to study in the standard randomized clinical trial. Therefore, challenges include the limited basic science and early phase clinical research available and issues related to blinding, placebo controls, the need for measurement development and analyzing data when there are many outcomes of interest in relatively small sample sizes.

337 Design of Experiments I

Section on Physical and Engineering Sciences

Wednesday, August 11, 10:30 am–12:20 pm

Optimal Design for Spherical Regions

◆ Robert Mee, University of Tennessee; Bradley Jones, SAS Institute Inc.

University of Tennessee, Knoxville, TN 37996-0532

rmee@utk.edu

Key Words: *central composite design, D-optimal, I-optimal, polynomial model, response surface model*

Most optimal design algorithms for quantitative factors constrain the treatment combinations to a hypercube region. In many applications, however, a spherical region seems more readily justified. We explore the impact of optimality criteria on the choice of an exact optimal design in a spherical region for experiments run in a single block, as well as for applications where a second-order design is augmented with another block to make possible fitting a cubic polynomial model. Our conclusions are based on a combination of empirical results from specific examples and theoretical results from the continuous design literature.

Optimal Main Effect Plus Plans for 2^m Factorial Experiments

◆ Ying Tian, University of California, Riverside; Subir Ghosh, University of California, Riverside

University of California, Riverside, 3473 Avocado, Riverside, CA 92507

xbbtian@yahoo.com

Key Words: *fractional factorial plans, estimation capacity, optimal designs, search designs*

In main effect plans for 2^m factorial experiments, we assume that the factors do not interact with each other. Such an assumption may or may not be true in practice. We present optimal plans that can not only estimate the general mean and main effects but also estimate all sets of k two-factor interactions for some practical values of k, m, and the number of runs (n). For even values of k, our plans are search designs in searching for (k/2) nonzero two-factor interactions from the set of all two-factor interactions and then estimating them with the general mean and the main effects.

SELC: Sequential Elimination of Level Combinations by Means of Modified Genetic Algorithms

◆ Abhyuday Mandal, Georgia Institute of Technology; Jeff C.F. Wu, Georgia Institute of Technology

Georgia Institute of Technology, 443 Lynch Ave. NW, Atlanta, GA 30318

amandal@isye.gatech.edu

Key Words: *orthogonal arrays, fractional factorial designs, response surface methodology, Bayesian variable selection*

To search for an optimal design in a large search space, Wu, Mao, and Ma (1990) suggested the SEL-method to find an optimal setting. Genetic algorithms (GA) can be used to improve upon this method. To make the search procedure more efficient, new ideas of forbidden array and weighted mutation are introduced. Relaxing the condition of orthogonality, GA is able to accommodate a variety of design points which allows more flexibility and enhances the chance of getting the best setting in fewer runs, particularly in the presence of interactions. The search procedure is enriched by a Bayesian method for identifying the important main effects and two-factor interactions. Illustration is given with the optimization of three functions, one of which is from Shekel's family.

Comparison of Designs Using Change of Variance Functions for Measuring Possible Correlation Effects

◆ Yun Shen, University of California, Riverside; Subir Ghosh, University of California, Riverside

University of California, Riverside, Apt. 309, 3275 Sepulveda Blvd., Los Angeles, CA 90034

sheny01@student.ucr.edu

Key Words: *correlation, design, linear model, optimality, ordinary least squares estimator, precision*

Zhou (2001) introduced a change of variance function for a linear function of parameters in the standard linear model when there may be some possible correlation present in observations. In this

paper we propose four additional changes of variance functions for the same purpose. The change of variance functions are used for measuring the possible correlation effected in the variance-covariance matrix of the estimated parameters. We also use them for the comparison of designs. We find a design by minimizing one of our proposed changes of variance functions in a simple response surface setup. We then compare its performance with all variance design, all bias design, and the design making the average variance equal to the average squared bias.

Average Correlations in Projections Designs

◆ Cheryl Dingus, The Ohio State University; Angela Dean, The Ohio State University

The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210-1247

dingus.7@osu.edu

Key Words: *orthogonal arrays, projection properties, generalized wordlength pattern, isomorphism, main effect plan*

Orthogonal arrays are used as screening designs to identify active main effects, after which the properties of the subdesign for estimating these effects and possibly their interactions becomes important. The subdesign consisting of the columns of the orthogonal array corresponding to the active factors is known as a projection design. Two projection designs are said to be isomorphic or combinatorially equivalent if one can be obtained from the other by a sequence of row and column permutations and permutations of symbols in each column. Nonisomorphic projection designs are often evaluated by the generalized wordlength pattern. This can distinguish many, but not all, nonisomorphic projections. An alternative criterion is presented. This criterion is based on the average squared correlations of complete sets of orthonormal contrasts and it will be shown that the criterion is independent of the contrasts selected. Examples will demonstrate that the criterion is better able to rank order and distinguish projections from three-level orthogonal arrays than the generalized wordlength pattern. (Joint work with H. Evangeleros and C. Koukouvinos.)

Experiment Design in the Presence of Restricted Factor Combinations and Resource Constraints

◆ Joanne R. Wendelberger, Los Alamos National Laboratory

Los Alamos National Laboratory, Statistical Sciences Group, MS F600, Los Alamos, NM 87545

joanne@lanl.gov

Key Words: *experimental design, restrictions, factorial aliasing, augmented design*

In the face of limited resources, statistical design of experiments provides techniques for selecting experimental runs for efficient collection of information. When the number of factors is large, standard design tools such as fractionation and the use of optimal design criteria may be useful. However, practical considerations such as restrictions on certain factor combinations and resource constraints may limit the applicability of standard designs, requiring some ingenuity in the way statistical tools are used to select runs from a set of possible candidates. A materials compatibility study will be used to illustrate the process of identifying a suitable set of runs that makes efficient use of runs in a restricted design space. Careful aliasing of factors in a nonstandard manner leads to a fractional design that allows estimation of the desired effects, while avoiding undesirable factor combinations. The use of optimal design methodology to augment this set of runs with additional candidate runs is also investigated.

Optimal Design for Multilevel Experiments

◆ Christopher I. Vahl, Kansas State University; George A. Milliken, Kansas State University

Kansas State University, Dept. of Statistics, 101 Dickens Hall, Manhattan, KS 66506

vahl@stat.ksu.edu

Key Words: *split-split-plot, optimal design, multilevel design, exchange algorithms, split-plot*

A multilevel experiment contains different sizes of experimental units. Examples include split-plot and strip-plot experiments, which are commonly used in the agricultural and engineering sciences. Optimal designs for experiments with independent errors can be constructed using exchange algorithms. A modification of these algorithms is used to account for the covariance matrix in split-plot and split-split-plot experiments and some results are discussed.

338 Sample Size Determination and Power

Biopharmaceutical Section

Wednesday, August 11, 10:30 am–12:20 pm

Toward a New Way of Thinking about Power and Sample Size

◆ John Johnson, Rho, Inc

Rho, Inc, 100 Eastowne Dr., Chapel Hill, NC 27514

jjohnson@rhoworld.com

Key Words: *clinical trials, power and sample size, group sequential*

Recent advances in the use of the group sequential clinical trials and the difficult questions they raise are going to require us to rethink power and sample size. No longer can we categorize errors as Type I or II, set a probability for each, and solve for a number or even a set of boundaries. For example, NonStop methodology, which is an application of group sequential to the conduct of clinical trials, assumes that there will be no Type I error. For otherwise, we would stop, declare efficacy, and file an NDA. Therefore, the role of Type I error probability is muddled. We refine the roles of alpha and beta in creating group sequential designs and clarify their interpretation. Then we explore a richer set of possibilities of how a group sequential trial can go wrong. Finally, we discuss the implications on the design of such a trial.

Sample Size Re-estimation—Time to Reconsider

◆ Carl-Fredrik Burman, AstraZeneca R&D; Christian Sonesson, Göteborg University

AstraZeneca R&D, AstraZeneca R&D Mölndal, Statistical & Mathematical Science, Mölndal, SE-431 83 Sweden

carl-fredrik.burman@astrazeneca.com

Key Words: *adaptive design, clinical trial, interim analysis, sufficiency principle*

Adaptive designs in clinical trials allow great flexibility while preserving the Type I error. Features such as the sample size may be changed during the course of the trial in response to interim analyses or external information. Research within this area has increased rapidly the last couple of years. However, sample size re-estimation (SSRE) can be criticized for several reasons: (1) It is not clear how to apply SSRE optimally; (2) Group sequential methods based on likelihood ratios are in some sense more efficient than SSRE; and (3) SSRE implies violating fundamental inferential principles. The violation of inferential principles is explored with focus on sufficiency, ancillarity and likelihood principles and we highlight by examples the difficulties inherited in SSRE. We discuss different approaches taken to justify SSRE in some situations.

Uncertainty in Planning Sample Size of Phase III Trial Based on Phase II Data

◆ Sue-Jane Wang, U.S. Food and Drug Administration; H.M. James Hung, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857

wangs@cder.fda.gov

Key Words: *uncertainty, sample size, phase II, phase III*

A recent informal survey suggested that the success rate of phase III clinical trials is low. Many research efforts are taking place to explore possible causes. Traditionally, phase II trials are relatively small and can be expected to result in a large degree of uncertainty in the estimates based on which phase III trials are planned. We investigate the impact of various treatment effect estimates or strategies on the sample size planning for phase III trials.

Statistical Power and Sample Size Determination for Clinical Trials with Multiple Primary Endpoints

◆ Chengjie Xiong, Washington University, St. Louis; J. Philip Miller, Washington University, St. Louis

Washington University, St. Louis, Division of Biostatistics, Campus Box 8067, 706 S. Euclid, St. Louis, MO 63110

chengjie@wubios.wustl.edu

Key Words: *Alzheimer's disease, Central Limit Theorem, intersection-union test*

This article concerns the statistical power and sample size determination for testing the treatment efficacy from clinical trials involving multiple primary endpoints. We consider a two-arm clinical trial in which the definition of treatment efficacy requires the simultaneous statistical evidence favoring the active treatment from all primary endpoints. We study the significance level and the statistical power for the test of treatment efficacy using the intersection-union test for several different situations based on the type of primary efficacy variables. We also present the appropriate method of sample size determination for such clinical trials and compare it to the intuitive approach of selecting the maximum sample size over those obtained from individual tests of multiple primary variables. We point out that the maximum sample size from multiple individual tests is often inadequate for the test of treatment efficacy. A real Alzheimer's clinical trial is used to demonstrate the sample size method.

Can Larger Sample Size Provide More Power?

◆ Mingxiu Hu, Pfizer Global Research & Development; Gordon Lan, Aventis Pharmaceuticals; Rebecca J. Benner, Pfizer Inc.

Pfizer Global Research & Development, 50 Pequot Ave., MS A5146, New London, CT 06371

mingxiu_hu@groton.pfizer.com

Key Words: *clinical trial design, amount of information, sample size, power, logrank statistics, treatment effect*

In clinical trials, we recruit more patients to achieve higher power under the assumption that all recruited patients have similar treatment effects. However, in practice, we may experience deteriorating treatment effects with the patient population, i.e., patients recruited earlier (probably with severer conditions) may have more improvement than patients recruited later (probably with milder conditions). In this case, recruiting more patients will not necessarily lead to more power. This presentation examines how power changes with sample size under some deteriorating patterns and shows when the power reaches its maximum. We will extend this concept to survival analysis to investigate how logrank statistics changes over time assuming the new therapy truly delay the occurrence of death. Under a survival setting, we follow patients longer to obtain more events to increase power. However, the logrank statistics may not increase in a monotone fashion as events accumulate, which often leads to misinterpretation of long-term treatment effect. Examples will be provided to demonstrate and explain this phenomenon.

Setting Tolerance Limits in Clinical Trials with Dichotomous Outcomes

◆ Boris G. Zaslavsky, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 1401 Rockville Pike HFM-219, CBER, Rockville, MD 20852

zaslavsky@cber.fda.gov

Key Words: *clinical trial, confidence, tolerance, StatXact, binomial distribution, negative binomial distribution*

Statistical evaluation and planning of clinical trials frequently utilizes the notion of confidence intervals. The definition of confidence intervals is well known and is broadly accepted by biomedical researchers. In some clinical trials the notion of tolerance interval is utilized. This notion is applicable to the clinical trials that evaluate products based on their performance in a part of the population. The tolerance intervals are rarely mentioned in clinical protocols, even though they are used often in some clinical studies. When the outcomes of a trial are continuous, the tolerance intervals are used in validation and process control of medical products. Tolerance intervals for normally distributed variables can be calculated using available statistical tables. For dichotomous outcomes, the definition of a tolerance interval is complicated, and no tables are available to calculate it. This research provides an algorithm for calculating tolerance intervals using StatXact.

Simulation Studies of the Effects of Sample Size, and Stratification Factors on Performance of the Fisher's Exact and Cochran-Mantel-Haenszel Tests

◆ David Carlin, MedImmune, Inc.; Harry Yang, MedImmune, Inc.

MedImmune, Inc., 35 W. Watkins Mill Rd., Gaithersburg, MD 20878

carlind@medimmune.com

Key Words: *simulation, Fisher's exact, Cochran-Mantel-Haenszel, sample size, power*

Sample size determination, selections of stratification factors and statistical test methods are important steps in planning a clinical trial. Since the Cochran-Mantel-Haenszel test is based on an asymptotic distribution requiring large sample size, the Fisher's exact test normally serves as an alternative method when the sample size is small. However, it does not allow for adjustment of mean and variance estimation based on stratification factors. Therefore, it may decrease the chance for detecting treatment effect. Presented in this paper are results of a simulation study conducted to evaluate the impact of sample size and stratification factors on the performance of the Fisher's exact and Cochran-Mantel-Haenszel (CMH) tests. The study demonstrates that even for a moderate sample size (n≥60), CMH test is more powerful in detecting differences. Also included are several applications of simulation methods to determine sample size not readily calculated by commercially available software packages, and develop stochastic process control criteria.

339 Clinical Trials

Biopharmaceutical Section

Wednesday, August 11, 10:30 am–12:20 pm

Proportion of Similar Response to Evaluate Similarity of Immune Response in Vaccines Clinical Trials

◆ Katherine E.D. Giacoletti, Merck & Co., Inc.; Joseph F. Heyse, Merck & Co., Inc.

Merck & Co., Inc., UNA-102, 785 Jolly Rd., Blue Bell, PA 19422

katherine_giacoletti@merck.com

Key Words: *similarity, equivalence*

Studies to demonstrate the similarity of immune responses are important parts of vaccine clinical development programs. In such studies there is particular interest in demonstrating that the distributions of antibody responses are equivalent for multiple vaccine lots drawn from the manufacturing process and for different formulations of a vaccine. Traditionally, a comparison on the basis of Geometric Mean Titer is used for this analysis of equivalence. Proportion of Similar Response (PSR) has been proposed as an alternative method. We will investigate the properties of PSR analysis in the settings of (a) three or more distributions, (b) unequal variances, and (c) baseline covariate adjustments. The properties will be investigated using simulation methods. An example from a vaccine development program will be provided to illustrate the methods.

A Comparison of Methods for the Joint Evaluation of Incidence and Severity

◆ Xiaoming Li, Merck & Co., Inc.; Devan V. Mehrotra, Merck & Co., Inc.

Merck & Co., Inc., UN-A102, 785 Jolly Rd., Blue Bell, PA 19422

xiaoming_li2@merck.com

Key Words: *joint evaluation, incidence, severity*

In randomized comparative clinical trials, joint evaluation of incidence and severity is often of interest. For example, two vaccines could be compared with respect to their relative ability to prevent the disease in question (e.g., chicken pox), along with the relative severity of the disease among those infected. We will compare few statistical methods for the joint evaluation of incidence and severity. The methods include Lachenbruch's "two-part" model, O'Brien's method, Fisher's combined p value method, and Simes's method. An illustrative example and simulation results will be provided to support some recommendations.

Phase I Trial Statistical Assessment of the Tetrameric Single-chain Anti-CD20

◆ Kuang-Ho Chen, University of Alabama, Birmingham

University of Alabama, Birmingham, Dept. of Medicine, 1078 Ste. A, Liberty National Bldg., 2001 3rd Ave. S, Birmingham, AL 35294

kuangho.chen@ccc.uab.edu

Key Words: *statistical efficacy assessment, fusion estimate of pre-targeted tumor cells, kinetics inference, hypothesis testing on coefficient matrix, linear system of ordinary differential equations*

A Pretarget ® Radioimmunotherapy of colorectal cancer in phase I trial under lower doses of B9E9FP and DOTA-Biotin with the short period before clearance has statistically confirmed not inferior than double dose of the tumor cell specificity fusing protein and/or the related antibody in the same or longer period. The kinetics of the targeted tumor cells fused by the specificity B9E9FP bound later by radiolabeled DOTA-Biotin has been statistically modeled for estimation and employed as an endpoint. The trial designed the four cohorts using combinations of loading concentrations of 160 and 320 mg/m2 of B9E9FP either mixed with 186Re-B9E9FP, of 1.3 and 0.65 mg/m2 of DOTA-Biotin, and 48 to 72 hours clearance durations for synthetic clearing agent. The confirmation is an application of the kinetic inference derived from the statistic compartment modeling with the Gaussian-type distribution. The distribution was empirically modeled from the noises induced from the treatment under the trial and the trial design. The modeling of the study is applicable to structuring a kinetic, not only for the current comparison.

On the Adaptiveness of a Simple Bayesian Model for Treatment Assignments in Clinical Trials

◆ Abdul J. Sankoh, Aventis Pharmaceuticals

Aventis Pharmaceuticals, PO Box 6890, 200 Crossing Blvd, BX2-300E, Bridgewater, NJ 08807

abdul.j.sankoh@aventis.com

Key Words: *adaptiveness, Bayesian, clinical trials, treatment assignment*

The need to design cost-effective and ethically sensitive clinical trials has re-energized the drive for more flexible clinical trial designs and statistical methods. There is a renewed research focus on the development of statistically and ethically efficient adaptive clinical trial methodologies by scientists both at the industry and regulatory agencies. This presentation discusses the properties of a simple Bayesian adaptive treatment assignment tool in clinical trails to minimize the number of subjects' exposed to the inferior treatment arm. The close resemblance of this simple Bayesian adaptive model to some popular frequentist models is examined.

Comparison of One-way and Two-way Analysis of Variance for Center-stratified Sparse Data

◆ Roger Qu, Forest Laboratories, Inc.; Bin Cheng, University of Wisconsin, Madison; Zhen R. Luo, Forest Laboratories, Inc.

Forest Laboratories, Inc., Harborside Financial Center, Plaza V, Jersey City, NJ 07311

peng.qu@frx.com

Sparse data occur often in multicenter clinical studies. When the randomization of a study is stratified by center, for continuous response the two-way analysis of variance (ANOVA) model is usually used for statistical analysis. However, when data are sparse with many centers having small number of patients relative to the number of treatments, a common practice is to ignore the center effect, and to use a one-way ANOVA model for statistical inference. Much of the debate has been in place as to whether or not the one-way analysis is appropriate for data arising from center-stratified randomization. This presentation addresses the issue by comparing the sizes and powers of the both methods. We show that for testing the hypothesis corresponding to the two-way additive ANOVA model (Type II analysis), the size of the one-way test is generally uncontrolled unless the sample sizes are proportional within centers. We also indicate that even under the most favorable situation to the one-way ANOVA (i.e., no center effect), the power gain in using a one-way instead of a two-way test is practically negligible.

Analysis of Clinical Data with Varying Time Intervals

◆ Cong Chen, Merck & Co., Inc.

Merck & Co., Inc., BLX 27, PO Box 4, West Point, PA 19486

chencong_04@yahoo.com

Key Words: *GEE, longitudinal data, mixed effect*

In clinical studies, repeated measures of a response variable collected at pre-specified clinical visits are routinely analyzed as visit-driven longitudinal data. However, such an analytic strategy always has a difficulty in dealing with data taken at a rescheduled or unscheduled visit. In practice, based on a time-window for each visit, these data are often either excluded from the analysis or carried (forward or backward) to a pre-specified visit in the same neighborhood. Examples in textbooks on longitudinal analysis are often based on such "cleaned" data. Both approaches have problems. The exclusion of otherwise admissible data leads to the loss of power for the detection of treatment effect; valuable time information on data measurements is lost when data is carried to a different time point. More sophisticated approach may consider it a complicated missing data problem, and deal with it accordingly. We instead propose a simple analytic strategy that is supported by standard procedures in SAS, and consistent with common practice of clinical data reporting. It is easily acceptable to the medical community, and readily accessible to statistical practioners.

The Importance of Balance on Covariates in Clinical Trials: When Covariate-adjusted Analyses are Not Enough

◆ Brenda L. Gaydos, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center DC2233, Indianapolis, IN 46285

blg@lilly.com

Key Words: *ANCOVA, stratification, adaptive randomization*

Stratification or adaptive randomization to insure balance on important covariates can add to the complexity and costs of a clinical trial. In addition, it has been argued that gains to inference above those achieved through covariate-adjusted analyses are minimal or nonexistent in many cases. Therefore, researchers often opt for simple randomization. It is crucial, however, not to underestimate the risk of imbalance on inference. In some cases, covariate-adjusted analyses can mislead investigators. The resulting p value will not be predictive of the risk in concluding drug effectiveness. Development of a promising new treatment may be stopped or development of an ineffective one continued inappropriately. The limitations of covariate-adjusted analyses in the presence of baseline imbalance for the normal model will be discussed. Examples of clinical trials will be provided where baseline imbalance was present but not statistically significant, yet the data was uninterpretable. Guidelines will be provided to help understand the risks of simple randomization when designing a trial, and the impact of imbalance on analyses when reporting results.

340 From Mail-order Brides to Collegiate Athletes: Analysis of Unique Populations Around the World ▲

Social Statistics Section

Wednesday, August 11, 10:30 am–12:20 pm

Ecological Inference and Radicalism

◆ David E. Booth, Kent State University; Stephane E. Booth, Kent State University

Kent State University, M&IS Dept., College of Business Administration, Kent, OH 44242

dbooth@bsa3.kent.edu

Key Words: *radicalism, coal miners, mine war, ecological inference, robust regression*

It would be interesting to know how radical coal miners that participated in the Southern Illinois mine war actually were. Ecological inference under some conditions might help obtain that information. We review ecological inference and use it, along with robust regression, and voting records to determine the level of radicalism of southern Illinois miners.

What is the Ideal Size for an Iowa Town?

◆ Russell V. Lenth, University of Iowa; Tom W. Rice, University of Iowa

University of Iowa, Dept. of Statistics, 241 Schaeffer Hall, Iowa City, IA 52242

russell-lenth@uiowa.edu

Key Words: *community size, splines, bootstrap, constrained estimation, smoothing*

We will present an analysis of information collected from 99 randomly selected small Iowa towns (populations range from 500 to 10,000). The information consists of community-level U.S. census data and responses to surveys of the towns' residents. Some of the survey variables measure civic attributes about the communities, such as the extent to which respondents agreed that their whole community comes together when something needs to be done and the extent to which people look out for each other in their community. Analysis shows a curvilinear relationship between these civic attributes and town size, with towns of approximately 800 tending to be more civic than small or larger communities. To investigate this relationship, we fitted (suitably constrained) splines to the satisfaction data, with certain demographic variables as covariates, and used bootstrap methods to derive confidence bounds on the optimum population. We will discuss the models used, how the constraints are implemented, the results, and plans for further investigations when more up-to-date data become available.

Comparison of Vital Statistics from Ohio Counties

◆ David Banaszak, Tri County Right to Life Educational Foundation

Tri County Right to Life Educational Foundation, 168 Stratmore, New Carlisle, OH 45344

davebanaszak@juno.com

Key Words: *statistics, Ohio, rates, correlation, outliers, scatterplots*

This paper compares recent vital statistic variables from 88 Ohio counties. The variables are (1) Live Birth rate per 1,000 population, (2) Low Birth Weight rate per 1,000 live births, (3) Births to Unwed Mothers per 1,000 live births, (4) Teen Birth rate per 1,000 female population aged 10-19, (5) Total Death rate per 1,000 population, (6) Infant Death rates per 1,000 live births, (7) Neonatal Death rate per 1,000 Live Births, (8) Marriage rate per 1,000 population, (9) Divorce rate per 1,000 population, (10) Income per Capita, (11) Population Density, and (12) Abortion rate per 1,000 live births. Univariate analysis will look for outliers and check for the marginal normality of the variables. Mahalanobis distance may also identify outliers. A correlation matrix and matrix scatter plots discern relationships between the variables. Correlation of abortion rates with other variables is of interest. Principal component analysis determines dimensionality and identifies outliers. Cluster analysis investigates similarities between the counties and the appearance of outliers. Identifying outliers provides county, state and federal governments with insights to county needs.

Assessing Outcomes for Intercollegiate Athletes at Liberal Arts Colleges

◆ John D. Emerson, Middlebury College

Middlebury College, Dept. of Mathematics, Middlebury, VT 05753

jemerson@middlebury.edu

Key Words: *intercollegiate athletics, educational outcomes, liberal arts colleges, empirical research*

Empirical research published in two books by William Bowen and his collaborators, The Game of Life and Reclaiming the Game, calls attention to the substantial impact of intercollegiate athletics at selective colleges. Longitudinal studies of thousands of college students indicate that the academic credentials of entering athletes increasingly differ from those of other students at these colleges. When coaches recommend athletes for admission, the acceptance rate is substantially increased, often by a factor of two. Intercollegiate athletes, especially male athletes and recruited athletes, perform less well academically than their test scores and other characteristics predict—they underperform. These findings have led to policy changes at some colleges, and especially at the conference and national levels. Empirical investigation is now needed to assess outcomes and to monitor progress made from efforts to reduce the gaps in academic achievement between athletes and other students. One stated goal is to ensure that athletes are fully representative of their student bodies.

Successful Self-representation in Cyberspace: Evidence on Russian Brides from an Internet Marriage Agency

◆ Padma Rao Sahib, University of Groningen; Ruud Koning, University of Groningen; Arjen van Witteloostuijn, University of Groningen

p.rao.sahib@eco.rug.nl

Key Words: *mail-order brides, probit models, personal advertisements*

We investigate the factors associated with success in a specialized marriage market. This marriage market is generated by a marriage agency that arranges meetings between women from the former Soviet Union and men from the United States, Canada, and Western Europe. It is the modern version of the mail-order bride industry that has successfully migrated to the internet. Russian women seek partners by providing profiles of themselves on a web site. Success in this market is defined as being able to find a suitable match culminating in engagement or marriage. We analyze data on profiles of two groups of Russian women using probit models. One group consists of single women who seek partners while and the other is a sample of women who have successfully found a partner via this method. By comparing such profiles, we are able to identify the factors associated with success in this marriage market. We find that stated personal qualities make little difference. However, women reporting lower weight are more likely to find a partner than women who report higher weight.

Forecasting Population Changes in Croatia until 2050

◆ Ante Rozga, Ekonomski Fakultet Split

Ekonomski Fakultet Split, Faculty of Economics, Matice Hrvatske 31, Split, 21000 Croatia

rozga@efst.hr

Key Words: *population, mortality, life expectancy, immigration*

Croatia has undergone dramatical demographic changes in recent years. Population felt from 4,540,641 in 1991 to 4,422,248 in 2003. This is partially the consequence of the war, but there has been a downward trend for decades. Number of live births fell from 76,156 in 1960 to just under 40,000 in 2003, a drop of 47.5%, which is among the biggest drops in the Europe. The number of deaths increased from 41,361 in 1960 to 50,569 in 2003. Using an optimistic scenario, we could expect 4,388,085 people living in Croatia in the year 2050. Infant mortality rates would fall from 11.46 in 1991 to 3.65 in 2050. The life expectancy would rise from 69.51 to 78.47 years. Deaths per 1,000 population would remain steady, from 12.82 in 1991 to 13.01 in 2050. This is the projection if some measures would be taken to stimulate population growth. If these measures were taken but unsuccessful, the worst scenario would be 3,350,000 people living in Croatia in the year 2050. The number of people over 65 years of age would outnumber those under 15, posing a big problem for Croatian economy. In this scenario there would be a lack of work force, and the immigration policy should be employed.

Initial Results from the National Jewish Population Survey 2000-01

◆ David A. Marker, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

davidmarker@westat.com

Key Words: *random digit dialing, religion, intermarriage*

Initial results are now available from the NJPS 2000-01. This large-scale telephone survey of the U.S. population was conducted using random digit dialing. Over 175,000 households were screened for religious identity. There has been a lot of debate in the press about the findings of this study. Among the topics covered will be: how many Jews are really in the U.S.; religious composition of the U.S.; demographic outlook for the next 10 years; what are the effects of intermarriage on the number of Jews; and selected data by denomination.

341 Nonparametric Approaches to Regression and Correlation

Section on Nonparametric Statistics, Biometrics Section, WNAR

Wednesday, August 11, 10:30 am–12:20 pm

Interval Estimation for Rank Correlation Coefficients Based on the Probit Transformation with Extension to Measurement Error Correction of Correlated Ranked Data

◆ Bernard A. Rosner, Harvard Medical School; Robert J. Glynn, Brigham & Women's Hospital

Harvard Medical School, Channing Laboratory, 181 Longwood Ave., Boston, MA 02115

bernard.rosner@channing.harvard.edu

Key Words: *ANOVA, nonparametric statistics, nutritional epidemiology, regression dilution bias*

The Spearman rank correlation coefficient ρ_S is routinely used as a measure of association between two non-normally distributed variables. However, confidence limits for ρ_s are only available for ρ_s\neq 0 under the assumption of bivariate normality. We introduce an indirect approach based on the use of the probit transformation for obtaining confidence limits for rho_s for an arbitrary bivariate distribution for (X,Y). This will also allow us to test the hypothesis H_0: ρ_s = ρ_0 vs. H_1: ρ_s \neq ρ_0 for arbitrary ρ_0. In some nutritional applications, the rank correlation between nutrient intake as assessed by a gold standard instrument and a surrogate instrument is used as a measure of validity of the surrogate instrument. However, if only a single replicate (or a few replicates) are available for the gold standard instrument, then the estimated rank correlation will be downwardly biased due to measurement error. We use the probit transformation as a tool for specifying a ANOVA-type model for replicate ranked data.

A Nonparametric Alternative to Analysis of Covariance

◆ Arne C. Bathke, University of Kentucky

University of Kentucky, 875 Patterson Office Tower, Dept. of Statistics, Lexington, KY 40506-0027

arne@ms.uky.edu

Key Words: *nonparametric model, rank test, analysis of covariance, ordered categorical data, ordinal data, dependent observations*

The Analysis of Covariance (ANCOVA) is designed for the many practical situations in which factor effects are obscured by concomitant variables, or the main purpose of the investigation lies in assessing the effect of the concomitant variables. We consider a nonparametric model with covariates. The information contained in the covariates is used to minimize the variance of certain nonparametric estimators for the response variable. This model combines the power gain through introduction of covariates into a factorial design with the robustness of nonparametric procedures. We discuss asymptotic inference for factor effects as well as the effect of covariates. Application of the suggested methods to real data is demonstrated using a SAS-IML macro. The tests can be used for data with ties, and even for purely ordinal data, including ordinal covariates. The number of covariates that can be included into the model is not restricted. Simulations show extremely good small-sample performance. In many situations, the proposed tests only require sample sizes around 10.

Testing for Covariate Effects in the Fully Nonparametric Analysis of Covariance Model

◆ Lan Wang, University of Minnesota; Michael G. Akritas, Pennsylvania State University

University of Minnesota, 385 Ford Hall, School of Statistics, 224 Church St. SE, Minneapolis, MN 55455

lan@stat.umn.edu

Key Words: *covairate, ANCOVA, nonparametric, ANOVA*

In the context of the fully nonparametric analysis of covariance model of Akritas et al. (2000), we propose methods to test for the presence of covariate main effects and interaction effects between the covariate and factors. The idea underlying the proposed procedures is that covariates can be thought of as factors with many levels. The test statistics are closely related to some recent developments in the asymptotic theory for analysis of variance when the number of factor levels is large. They are very easy to compute and have asymptotic normal laws under the null hypotheses. The methods are investigated on simulated and real data.

Analysis of Heteroscedastic Multifactor Designs with Large Number of Factor Levels

◆ Haiyan Wang, Pennsylvania State University; Michael G. Akritas, Pennsylvania State University

Pennsylvania State University, 200 N. Danielle Dr., Apt. 328, Pleasant Gap, PA 16823

hwang@stat.psu.edu

Key Words: *rank tests, unbalanced designs, nonparametric hypotheses, asymptotic theory of quadratic forms, projection method*

Testing methods for ANOVA designs where some of the factors have a large number of levels have received a lot of attention recently. Most attention, however, has been restricted to procedures using the original observations, either in the balanced homoscedastic case or with no more than two factors. We consider possibly unbalanced and heteroscedastic multifactor designs. Such procedures require strong moment assumptions and are sensitive to outliers. Thus we also develop (mid-)rank procedures for the same general setting. The main asymptotic tools are the asymptotic rank transform and Hajek's projection method. Simulation results show that the present rank statistics outperform those based on the original observations, in terms of both Type I and Type II error rates. A real dataset from a microarray experiment is analyzed.

Bayesian Nonparametric Inferences Based on Ranks

◆ Xiaojiang Zhan, Pennsylvania State University

xzhan@stat.psu.edu

Key Words: *Bayesian nonparametric, ranks, location models, regression models*

We would like to explore how to implement the nonparametric procedures based on ranks in the Bayesian context. When prior information exists, it would be desirable to incorporate it in the data analysis, even when we are using rank-based nonparametric methods. We will begin with one sample location model and two sample location model where several common scores statistics are used, such as the sign statistic, the Wilcoxon signed rank statistic, the normal scores statistic in the one sample case and the Mann-Whitney-Wilcoxon statistic, the Mood's median statistic in the two sample case. Regression models (including the DOE models) are of primary interest in this paper. The estimation and model selection problems will be addressed.

A Unified Nonparametric Approach for Unbalanced Factorial Designs

◆ Xin Gao, York University; Mayer Alvo, University of Ottawa

York University, Dept. of Maths & Stats, 4700 Keele St., North York, ON M3J 1P3 Canada

xingao@mathstat.yorku.ca

Key Words: *linear rank statistics, main effect, interaction effect, nested effect, Pitman alternative, quantitative trait*

Motivated by questions arising from the field of statistical genetics, we consider the problem of testing main, nested, and interaction effects in unbalanced factorial designs. Based on the concept of composite linear rank statistics, a new notion of weighted rank is proposed. Asymptotic normality of weighted linear rank statistics is established under mild conditions and consistent estimators are developed for the corresponding limiting covariance structure. A unified framework to employ weighted rank to construct test statistics for main, nested, and interaction effects in unbalanced factorial designs is established. The proposed test statistics are applicable to unbalanced designs with arbitrary cell replicates that are greater than one per cell. The limiting distributions under both the null hypotheses and Pitman alternatives are derived. Monte Carlo simulations are conducted to confirm the validity and power of the proposed tests. Genetic datasets from a simulated backcross study are analyzed to demonstrate the application of the proposed tests in quantitative trait loci mapping.

Hierarchical Generalized Additive Models

◆ Changchun Xie, McMaster University

McMaster University, Apt. 202, 65 Charlton Ave. West, Hamilton, ON L8P 2C2 Canada

xie@ccc.mcmaster.ca

Key Words: *hierarchical model, hierarchical quasi-likelihood, additive model*

We suggest hierarchical generalized additive models which allow an extra error component in the predictors of generalized additive models. Since the component is not directly observed, it is hard to know its distribution. Here, we make no assumptions on the functional form of the distribution other than putting some constraint on the mean and variance function of the extra error component.

342 Nonresponse Adjustments ◬

Section on Survey Research Methods

Wednesday, August 11, 10:30 am–12:20 pm

Does Weighting for Nonresponse Increase the Variance of Survey Means?

◆ Sonya L. Vartivarian, University of Michigan; Roderick J. Little, University of Michigan

slvartiv@umich.edu

Key Words: *missing data, nonresponse adjustment, sampling weights, survey nonresponse*

Nonresponse weighting is a common method for handling unit nonresponse in surveys. A widespread view is that the weighting method is aimed at reducing nonresponse bias, at the expense of an increase in variance. Hence, the efficacy of weighting adjustments becomes a bias-variance trade-off. This note suggests that this view is an oversimplification—nonresponse weighting can in fact lead to a reduction in variance as well as bias. A covariate for a weighting adjustment must have two characteristics to reduce nonresponse bias—it needs to be related to the probability of response, and it needs to be related to the survey outcome. If the latter is true, then weighting can reduce, not increase, sampling variance. A detailed analysis and simulations of bias and variance are provided in the setting of weighting for an estimate of a survey mean based on adjustment cells. The analysis and simulations suggest that the most important feature of variables for inclusion in weighting adjustments is that they are predictive of survey outcomes; prediction of the propensity to respond is a secondary, though useful, goal.

Efficient Nonresponse Weighting Adjustment Method Using Response Probability

◆ Jae-Kwang Kim, Yonsei University

Yonsei University, Seoul, Korea

kim00020@yahoo.com

Key Words: *missing data, survey sampling, jackknife, propensity score*

A nonresponse weighting adjustment method using response probability is proposed. The proposed weighting adjustment method also uses weighting adjustment cells, but does not require that the response probabilities are the same within each cell. By using the response probability, the proposed method adjusts for the nonresponse biases, but is less influenced by the sampling error of the estimated response probability. A suitable choice of weighting adjustment cell makes the point estimator efficient as well. Variance estimaton is also discussed.

Model-based Estimation in Network Sampling Using Samples Obtained by Subsampling

Hong Xu, Pennsylvania State University; ◆ Mike Kwanisai, Pennsylvania State University

Pennsylvania State University, 326 Thomas Bldg., University Park, PA 16802

kwanisai@stat.psu.edu

Key Words: *graph, snowball sampling, MCMC, data augumentation, Bayesian estimates*

In snowball sampling links or relations are followed to include new subjects into the study. In practice it is sometimes not feasible to follow all the relations but only a fraction of the available relations. We present a model-based approach for estimating population proportions using sample data obtained from subsampling. As examples, Bayesian estimates from a simulated and a real dataset are given.

Statistical Profile of the JOLTS Program

◆ Charlotte D. Mueller, Bureau of Labor Statistics; Mark Crankshaw, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 4840, Washington, DC 20212

mueller.charlotte@bls.gov

Key Words: *sample survey, labor statistics, response bias, data collection, nonresponse adjustment*

The Job Openings and Labor Turnover Survey (JOLTS) at the Bureau of Labor Statistics produces monthly estimates for job openings, hires, and separations. The goal of the JOLTS data is to reflect the current status of the labor market as well as show trends over time. JOLTS sampling procedures such as careful frame construction and frame stratification guarantee a sample that is statistically representative of the U.S. economy. Further, the JOLTS procedures for unit and item nonresponse adjustment were selected to translate the statistical validity of the sample into valid estimates. But methodology is not the last step in guaranteeing statistically valid estimates. Unit and item response rates indicate differences in how likely different types of establishments are to report, as well as differences in which data elements the establishments can or will provide. Predictably, examination of the reporting patterns reveals differences by industry and by establishment size class. As a sample survey, we strive to collect as much data from as many respondents as possible, but we also must aim for balance among reporters at the establishment and item level in order to reduce response bias. Maintaining high-quality estimates depends on monitoring reporting patterns, taking measures to reduce imbalances in unit and item response rates, and applying methodology in the estimation process to counteract large response bias if it occurs despite our efforts during the data collection process.

Evaluation of Response Rates over Time in a Mixed-mode Survey

◆ Holly B. Shulman, Centers for Disease Control and Prevention; Brenda Colley Gilbert, Centers for Disease Control and Prevention; Amy Lansky, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 205 College Ave., Swarthmore, PA 19081-1408

hbs1@cdc.gov

Key Words: *response rates, survey, mixed-mode*

Survey researchers are finding it increasingly difficult to maintain high response rates on surveys, as the public is becoming more resistant to their efforts. Many telephone surveys have seen response rates decline in recent years. Mail surveys, if properly designed, have been more successful at maintaining adequate response rates. We examine response data over time from the Pregnancy Risk Assessment Monitoring System (PRAMS), a state, population-based surveillance system of women who have recently delivered a live-born infant. PRAMS identifies eligible women from birth certificates and contacts them two to six months after delivery. The PRAMS methodology, which is consistent across participating states, consists of a mail survey with telephone follow-up for nonresponders. Response rate data from 1996 and 2001 are examined to determine response rates, contact rates, cooperation rates, and refusal rates. Logistic regression is used to examine the relationship between maternal and infant characteristics available from the birth certificate and the likelihood of response for the 23 states participating during 2001.

Applications of Propensity Scoring for the Medical Expenditure Panel Survey—Insurance Component

◆ Steven Riesz, U.S. Census Bureau; John Sommers, Agency for Healthcare Research and Quality; David Kashihara, Agency for Healthcare Research and Quality

steven.riesz@census.gov

Key Words: *nonresponse, post-stratification*

The MEPS-IC is an establishment survey, sponsored by the Agency for Healthcare Research and Quality, that collects data about employer-provided health insurance. Currently, weight adjustments for nonresponse and post-stratification are done using a cell-based system, where cells are defined using employment size and industry classification. This system was developed in the first years the survey was conducted when no survey data were available. This system and any new system must consider a three-stage interview process that adjusts for multiple levels of response and nonresponse. This paper discusses the evaluation of a proposed new method to correct for nonresponse and post-stratify the weights. The method will use response-propensity-scoring and raking to produce weights that consider a larger set of variables to correct for nonresponse, and will produce weights whose values have less variation. This method will likely reduce both the bias and the variance of our estimates.

Modeling of Response Propensities in the Joint Canada/United States Survey of Health

◆ Meena Khare, National Center for Health Statistics; Karen Davis, National Center for Health Statistics

National Center for Health Statistics, CDC, 10189 Maxine St.,
Ellicott City, MD 21042-6351

mxk1@cdc.gov

Key Words: *response propensity, weighting, nonresponse*

The Joint Canada/United States Survey of Health (JCUSH), a one-time national random-digit-dialing telephone survey, was conducted jointly by the National Center for Health Statistics (NCHS) and Statistics Canada from November 2002 to July 2003. The JCUSH is designed to compare selected health characteristics of both the U.S. and the Canadian adult population 18 years of age and older. Persons aged 65 years and over were sampled at a higher rate than other adults. Weighting class adjustments are commonly used in government-sponsored sample surveys to adjust for unit nonresponse. An alternative approach is to use logistic regression to model response propensities and use them to form adjustment cells. A response indicator variable is regressed upon potential covariates that are available for both respondents and nonrespondents to obtain predicted probabilities. This paper describes an evaluation of the weighting methodology used to adjust for noncoverage and unit nonresponse in the U.S. sample.

343 Survival Analysis: Informative Revisioning, Correlated Outcomes, and Model Testing

Section on Statistics in Epidemiology, Biometrics Section

Wednesday, August 11, 10:30 am–12:20 pm

A Kaplan-Meier Generalization in the Presence of Correlated Left Truncation Times

◆ Todd MacKenzie, Dartmouth Medical School

Dartmouth Medical School, Section of Clinical Research, HB7505,
One Medical Center Dr., Lebanon, NH 03756

Todd.MacKenzie@Dartmouth.edu

Key Words: *time-to-event, cumulative hazard, prevalent cohort, right-censoring, estimating equations, copula*

The Kaplan-Meier estimator can be used to estimate survival for left-truncated right-censored survival times, under the assumption of independence of the survival times and left truncation times. This paper presents semiparametric estimating equations for estimating survival when the left truncation time is correlated with survival. The approach uses a copula to describe the dependence structure between the survival time and the left truncation time. These estimating equations yield the Kaplan-Meier estimator if the independence copula is used.

Estimation of Probabilities of Adverse Events under Informative Dropouts

◆ Masako Nishikawa, National Institute of Public Health; Toshiro Tango, National Institute of Public Health, Japan; Makiko Ogawa, Aventis Pharma Ltd

National Institute of Public Health, 3-6 Minami 2 chome, Wako,
351-0197 Japan

mnishikawa@niph.go.jp

Key Words: *competing risk, time-to-event data, cumulative incidence function, recurrent event, nonparametric estimation*

In long-term treatment or in treatment with frequent severe adverse events (AE), such as those for oncology, it is important to know occurrence rate of AE across time and its severity.

Conventional method is to calculate proportion of patients with AE to all patients enrolled neglecting dropout or to estimate occurrence rate of AE by Kaplan-Meier estimator assuming that dropout is noninformative. However, this assumption cannot be always true and not be validated by data. Furthermore assumed population is only hypothetical. We propose a method applying competing risk analysis by defining events of dropout prior AE and AE prior dropout. We focus on an AE one by one. We distinguish obvious non-informative censoring from other censorings that may not be noninformative. Therefore, our approach does not need independent assumption for dropout. The cumulative incidence function (CIF) of the AE by severity can be obtained by viewing categorized severities as competing risk. We estimate CIF of sequential occurrence of the same AE by forming a restricted risk sets. Some useful examples of graphical presentation are shown.

Modeling the Relationship between Two Correlated Survival Outcomes

◆ Ying Guo, Emory University; Amita K. Manatunga, Emory University

Emory University, Dept. of Biostatistics, 1518 Clifton Rd. NE, Atlanta, GA 30322

yguo2@sph.emory.edu

Key Words: *correlated survival times, agreement, cancer epidemiology*

Often, disease occurrence time of the same subject can be measured by using two different methods. It is of interest to compare the disease free rates between the two methods and to determine whether both methods measure disease freedom in the same manner for each subject. To achieve these goals, we model the relationship between event times and its association to covariates. In addition, we propose several indices to quantify the agreement between the event times and discuss their statistical properties. We derive the relationship of these indices to commonly used odds ratio and discuss their interpretations. Finally, we illustrate our results with a prostate cancer epidemiology study where patients with prostate cancer are treated with simultaneous radiotherapy and their prostate specific antigen is measured by two definitions every six months after the treatment.

Structural Models for Quasi-dependent Failure and Truncation Times

◆ Emily C. Martin, Harvard School of Public Health; Rebecca A. Betensky, Harvard School of Public Health

Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115

emartin@hsph.harvard.edu

Key Words: *quasi-independence test, truncated data, structural model*

Randomly truncated survival data arise when the failure time is observed only if it falls within a subject-specific truncating interval. Most estimators of the survival function and regression models based on such data rely on the key assumption of quasi-independence, i.e., factorization of the joint density of failure and truncation times into a product proportional to the individual densities in the observable region, as well as the usual assumption of independent censoring. We propose semiparametric structural models for latent failure or truncation times applicable under quasi-dependence of observed failure and truncation times and independent censoring. Estimation is based on either observed failure time conditional on latent truncation time, or latent failure time conditional on observed truncation time, as determined by the chosen structural model. A quasi-independence test statistic that conditions on covariates of interest provides an estimating function for model parameters. Modifiers of the effect of truncation on failure appear in the structural model for the latent random variable. The approach is illustrated using real datasets and performance is simulated.

A Study of Deaths Due to Cardiovascular Diseases through Markovian Multistate Models

◆ Somesh Chattopadhyay, Florida State University; Daniel McGee, Florida State University

Florida State University, Dept. of Statistics, Tallahassee, FL 32306-4330

somesh@stat.fsu.edu

Key Words: *multistate models, Markov models, multivariate survival analysis, cardiovascular disease, competing risks, recurrent events*

In the survival analysis of cardiovascular diseases the reason for death may be one or more reasons. Also there are different states of a patient from which the patient can move only to a few other states. Some of the conditions of the patient can lead to other conditions. Some of the conditions of the patient may recur as well. In this situation, times to different events are correlated. It is important to investigate the relationship of different causes for death and different events and their joint impact on the patient. In this set-up, a multistate model may be the most appropriate to study the relationship between different cardiovascular events and death. We propose a Markovian multistate model which also retains some past history of cardiovascular disease events of the patient. We investigate the relationship of different causes of death through this kind of model using the data from Framingham study.

Entropy-based Testing of the Proportional Hazards Assumption

◆ Ramin B. Arani, Johnson & Johnson Pharmaceutical R&D, LLC

Johnson & Johnson Pharmaceutical R&D, LLC, 1125 Trenton-Harbourton Rd., Titusville, NJ 08560

rarani@prdus.jnj.com

Key Words: *entropy, relative entropy, residual lifetime distribution*

Let X,Y=0 denote lifetimes and F and G their respective distribution funtion. It can readily be shown that the relative entropy, I(X,Y) (i.e., Kulback-Leibler distance) corresponding to residual lifetimes of X and Y is constant over time if and only if F and G belong to proportional hazards family of distributions. Utilizing this property, one can formulate a formal test for the proportionality of the hazard functions based on a regression estimator of the slope that joins the emperical estimate of relative entropy at each failure time. The sensitivity of this approach is investigated through simulation under various scenarios.

344 Spatial Epidemiology

Section on Statistics in Epidemiology, Biometrics Section, Section on Health Policy Statistics

Wednesday, August 11, 10:30 am–12:20 pm

A Multiscale Method for Disease Mapping in Spatial Epidemiology

◆ Mary Louie, Channing Laboratory; Eric Kolaczyk, Boston University

Channing Laboratory, 181 Longwood Ave., Boston, MA 02115

mary.louie@channing.harvard.edu

Key Words: *disease mapping, multiscale spatial process, standardized mortality ratio, statistical inference, confidence map*

Maps by their very nature display spatial information that varies with resolution (or “scale”). The effects of scale in disease mapping are well recognized. Here we provide a statistical inference framework for describing the distribution of relative risk simultaneously across a hierarchy of multiple scales. This is accomplished via an extension of the Poisson multiscale spatial process model of Kolaczyk and Huang (2001) to measures of relative risk for one of two canonical data structures in disease mapping, that of tract count data. We develop the basic framework for the standardized mortality ratio (SMR), derive (Bayesian) posterior-based estimation strategies, and provide methods for characterizing uncertainty. We illustrate the method through an application to data on gastric cancer in Tuscany. The overall framework provides a tool from which informative disease and confidence maps can be produced across a hierarchy of multiple scales, without the need to first try to identify a single appropriate scale of analysis.

Simultaneous Bootstrap Confidence Intervals for Lorenz Curves Reflecting the Spatial Distribution of Disease Risk

◆ Owen J. Devine, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, MS E-87, 1600 Clifton Rd., Atlanta, GA 30333

ojd1@cdc.gov

Key Words: *disease rates, public health surveillance, spatial analysis, mapping*

Lorenz curves are often used to illustrate the distribution of economic measures within a given population. In more recent applications, Lorenz curves have also been used to evaluate spatial heterogeneity in the distribution of disease risk across geographic areas. In this application, an increase in curvature is indicative of an uneven spatial distribution of risk. An observed Lorenz curve, however, is subject to random variability in that, for example, some curvature is expected even if the underlying distribution of the true risk is constant. We propose a parametric bootstrap approach for constructing simultaneous confidence intervals for the coordinates of observed Lorenz curves. In addition, we use Monte Carlo methods to define a null region corresponding to the variability expected in an observed Lorenz curve given spatial homogeneity of true risk. The performance of the approach is evaluated through a simulation study and the method is applied in an examination of the spatial distribution of sexually transmitted disease risk across counties in North Carolina.

Space-time Interaction Models for Mortality Data

◆ James Leeper, University of Alabama; Jing Yu, University of Alabama

University of Alabama, PO Box 870326, Tuscaloosa, AL 35487-0326

jleeper@cchs.ua.edu

Key Words: *spatial-temporal, space-time interaction, general linear mixed model, generalized additive mixed model, mortality, doubly repeated measures*

When mortality rates are collected over time and space, modeling is a challenge because it is likely that observations from the same time unit are correlated, and it is also likely that observations from the same geographic area are correlated. Data with this structure are known as doubly repeated measures. Besides the fact that observations are correlated in two dimensions, space-time interaction issues are also of interest. Space-time interaction can be interpreted as different time trends in different geographic areas. This study is aimed at exploring spatial-temporal statistical methods for modeling mortality data with a concentration on the space-time interaction. This study extends a general linear mixed model (GLMM) for mortality data with a focus on space-time interaction. Data transformation is used to deal with the non-normality of mortality rates. The new model is compared to a generalized additive mixed model (GAMM), which does not assume normality, by simulation studies to compare power. The models are applied to county-specific infant mortality rates over a 30-year period in Alabama. These models are able to incorporate time-invariant and time varying covariates.

A Bayesian Analysis of the Mortality Curve for Small Area

◆ Rong Wei, National Center for Health Statistics; Balgomin Nandram, National Center for Health Statistics; Wenxing Zha, National Center for Health Statistics

National Center for Health Statistics, 14003 Marleigh Lane, CDC, Clifton, VA 20124

rrw5@cdc.gov

Key Words: *mortality curve, Heligman-Pollard law, nonlinear model, Griddy Gibbs sampler, State Life Tables*

A pilot study showed that U.S. mortality data have a pattern of age-specific death rates which can be parameterized with the Heligman-Pollard (HP) law. In constructing decennial State Life Tables, difficulties rise in smoothing mortality curves for some states with incomplete and sparse data (i.e., the HP law cannot be fitted to each state separately). To overcome this difficulty, a Bayesian method for small-area estimation is adopted. National mortality data with state code for the period 1999-2001, individuals aged 0 to 80 years, are modeled for demographic subpopulations. Death rates by age and state with independence are assumed to share a common beta distribution with one set of HP parameters for the entire U.S. Some HP parameters are transformed so that all parameters lie in the unit interval smaller than 1. Thus, independent uniform priors are placed on these parameters providing proper posterior inference. The Griddy Gibbs sampler is used to fit the overall model for the nation, and then an output analysis is used to obtain the HP law for each state.

Spatial Summaries of Small-area Temporal Effects from an Age-period-cohort Model

◆ Theodore R. Holford, Yale University School of Medicine; Linda W. Pickle, National Cancer Institute

Yale University School of Medicine, 60 College St., PO Box 208034, New Haven, CT 06511

theodore.holford@yale.edu

Key Words: *spatial models, age-period-cohort analysis, Bayes method, MCMC methods, rates, lung cancer*

The age-period-cohort (APC) model provides useful insights into the analysis of time trends for the disease rates when one considers estimable functions of the parameters. However, when one wishes to consider whether these temporal trends are spatially consistent, the precision of the estimated effects suffer due to small numbers of cases, especially in regions with small populations. A temporal-spatial model is described in which a conditional autoregressive prior distribution is applied to the region-specific temporal effects, thereby providing a means of obtaining Bayesian estimates of estimable functions of the model parameters using Markov chain Monte Carlo techniques. Methods for extending this spatial APC model to the analysis of trends in regions with relatively small populations are described. The estimable functions that can be used to describe the temporal trends include not only the temporal-spatial smoothed rates, but estimates of drift which describe the net temporal change that is occurring in a population. This model will also be extended to allow for the inclusion of covariate information that seeks to identify factors related to the temporal trend.

Semiparametric Spatial Modeling of Binary Outcomes with Application to Aberrant Crypt Foci in Colon Carcinogenesis Experiments

◆ Tatiyana V. Apanasovich, Texas A&M University; Raymond J. Carroll, Texas A&M University

Texas A&M University, Dept. of Statistics, 3143 TAMU, College Station, TX 77840

tanya@stat.tamu.edu

Our work is directed towards the analysis of aberrant crypt foci (ACF) in colon carcinogenesis. ACF are morphologically changed colonic crypts that are known to be precursors of colon cancer development. The colon is laid out as a gridded rectangle and the occurrence of an ACF within the grid is noted. The biological question of interest is whether these binary responses occur at random; if not, this suggests that the effect of environmental exposures is localized regionally. To understand the extent of the correlation, we cast the problem as a spatial binary regression with underlying Gaussian latent process. Marginal probabilities of ACF indicators are modeled semiparametrically, using fixed-knot penalized regression splines and single-index models. We modeled the underlying latent process in a nonstationary manner as the convolution of latent stationary processes. The dependency of the correlation function on location is also modeled semiparametrically. We fit the models using pairwise pseudo-likelihood methods. Assuming that the underlying latent process is strongly mixing, we proved asymptotic normality and derived the optimal rate of convergence for penalty parameters.

Modeling Small-area Variations in Mental Health Service Use with Geographically Weighted Regression

◆ Lisa Lix, University of Manitoba; Geoffrey DeVerteuil, University of Manitoba

University of Manitoba, Manitoba Centre for Health Policy, 408-727 McDermot Ave., Winnipeg, MB R3E 3P5 Canada

lisa_lix@cpe.umanitoba.ca

Key Words: *small-area variations, geography, mental health services, spatial correlation, local regression, neighborhoods*

A spatial model for the relationship of health service use to socioeconomic indicators for small-area data was investigated. Geographically weighted regression (GWR) was used, in which estimates of location-specific (i.e., local) regression parameters are obtained by kernel smoothing, with a bandwidth that adapts to spatial locations. GWR was applied to models of mental health service use obtained from acute care in-patient hospitalizations and physician billing claims for 75 neighborhoods in Winnipeg, Canada, a city of approximately 650,000 residents. The independent variables were indicators of population distribution, economic risk, social isolation, social disorganization, and health care infrastructure. Based on model fit statistics, the GWR Poisson regression model captured spatial associations better than a conventional Poisson model for the in-patient data. However, there was a high degree of spatial smoothing across neighborhoods. Little improvement in model fit was observed for the physician data. Accounting for spatial correlation in models of health service use may be beneficial for identifying predictors of the need for mental health services.

345 Regular Contributed Posters

Section on Statistical Education, Section on Statistical Consulting, Section on Survey Research Methods, Section on Government Statistics

Wednesday, August 11, 12:00 pm–1:50 pm

Classroom Simulation: Are Variance-stabilizing Transformations Really Useful?

◆ Bruce E. Trumbo, California State University, Hayward; Eric A. Suess, California State University, Hayward; Rebecca E. Brafman, California State University, Hayward

California State University, Hayward, Statistics Dept., 25800 Carlos Bee Blvd., Hayward, CA 94542

btrumbo@csuhayward.edu

Key Words: *variance-stabilizing transformation, non-normal data, simulation, Minitab/S-Plus/R, teaching, log/arcsine/square root*

When population variances of observations in an ANOVA are a known function of their population means, many textbooks recommend variance-stabilizing transformations. Examples are square root transformation for Poisson data, arcsine of square root for binomial proportions, and log for exponential data. To investigate the usefulness of transformations in one-factor ANOVAs with nonnormal data, we use simulations to approximate the true significance level and power of F-tests—with and without various variance-stabilizing transformations. In our examples, logarithmic transformations of exponential data prove especially valuable. Simulation code for S-Plus and R is provided (Minitab and SAS for some). Classroom use of these simulations in a second statistics course reinforces concepts of significance level and power, illustrates appropriate use and interpretation of transformations, encourages exploration, and teaches computer skills important in the job market.

Exploiting the Equivalence between Confidence Intervals and Hypothesis Tests in Introductory Statistics Classes

◆ Kenneth Constantine, Taylor University

Taylor University, 236 W. Reade Ave., Upland, IN 46989-1001

knconstantine@tayloru.edu

Key Words: *pedagogy, confidence intervals, tests of hypotheses*

We propose a focus for the inference portion of an elementary statistics course: the equivalence between confidence intervals and tests of hypotheses. This equivalence receives scant attention in introductory texts but seems helpful to beginning students. We advocate that, in order to make this equivalence fully applicable for students, an introductory statistics course ought to include coverage of one-sided confidence intervals. Student reference sheets which employ this equivalence in common inferential settings will be made available. We also acknowledge that the confidence interval-test equivalence is not accessible to beginning students in some inferential settings, e.g., tables for frequency data. Some pedagogical issues relating to classroom implementation will be indicated.

An Efficient Way to Display Pairwise Comparison Results in ANOVA

◆ J. Burdeane Orris, Butler University; Bruce Bowerman, Miami University of Ohio

Butler University, 4600 Sunset Blvd., Indianapolis, IN 46208

orris@butler.edu

Key Words: *ANOVA, pairwise, Tukey, teaching*

While several statistical analysis packages display confidence intervals for post-hoc analysis, they do not make it easy for students to see patterns in the data. This paper discusses an efficient way to display pairwise comparison results in analysis of variance. Both individual pairwise comparisons and Tukey simultaneous comparisons are discussed for one-factor ANOVA, randomized blocks ANOVA, and two-factor ANOVA. This method is implemented using the Excel add-in, MegaStat, which takes advantage of Excel's ability to sort data and format cells to help students see how the group means cluster to form patterns in the data. The data are also displayed graphically to help students see the concept of within-group and between-group variation.

Engaging Students in Statistical Discovery

◆ Amy G. Froelich, Iowa State University; W. Robert Stephenson, Iowa State University; William M. Duckworth, Iowa State University

Iowa State University, 324 Snedecor Hall, Ames, IA 50011-1210

amyf@iastate.edu

Key Words: *introductory statistics, conceptual knowledge, activities*

We are developing new course materials for the introductory statistics course designed to engage students in discovery. More than simply present the ideas and methods of statistics, the materials actively involve students in the design and implementation of data collection and the analysis and interpretation of the resulting data. Our goal in developing the course materials is to have students adopt a statistician's point-of-view and begin to think like statisticians. This poster will present several of the new activities. These materials have been incorporated into a small section of an introductory statistics course for two semesters. Preliminary results on student attitudes towards the materials will be presented.

An Introduction to R: Simulating Birthday Matches in the Nonuniform Case

◆ Clayton W. Schupp, California State University, Hayward; Bruce E. Trumbo, California State University, Hayward; Eric A. Suess, California State University, Hayward

California State University, Hayward, Statistics Dept., 25800 Carlos Bee Boulevard, Hayward, CA 94542

cschupp@horizon.csuhayward.edu

Key Words: *birthday matching problem, nonuniform birthday distribution, probability modeling, teaching R in the classroom*

Appropriate simulations can enliven a beginning probability course by focusing on model building, exploring generalizations that would lead to analytically intractable results, and teaching computer skills that are valuable in the job market. Here we use R to simulate the probability of birthday matches in a room of n people. The first simulation model assumes a uniform distribution of birthdays throughout the year, giving results that agree with a standard elementary formula. A very similar second simulation model drops the uniformity assumption, giving results that are analytically beyond the level of undergraduate probability. Data for the actual distribution of U.S. birthdays in 1997-99 provide a basis for the second model. An entry-level exposition of the required functions in R is provided.

Introductory Statistics as an Across-the-curriculum Course

◆ Debra Hydorn, Mary Washington College; Jeffrey Edmunds, Mary Washington College

Mary Washington College, 1301 College Ave., Fredericksburg, VA 22401

dhydorn@mwc.edu

Key Words: *statistics education, race and gender awareness, environmental awareness*

As part of its General Education program, Mary Washington College requires courses from five Across-the-Curriculum (ATC) areas: Writing Intensive, Speaking Intensive, Global Awareness, Environmental Awareness, and Race and Gender Awareness. ATC courses can be at any level from any discipline. This poster presentation will provide information for teaching an Introductory Statistics course under each of the Race and Gender and Environmental Awareness designations. Example data sets for in-class discussion and class projects will be provided along with assessments of the impact on student attitudes and learning. A comparison will also be made between these ATC sections and sections taught using examples and projects from a wider variety of applications.

A Computer Program for Friendly Introductory Statistics Help

◆ Gordon P. Brooks, The Ohio University; Holly Raffle, The Ohio University; Marsha Lewis, The Ohio University; Valerie Blom, The Ohio University

The Ohio University, 305A McCracken Hall, Dept. of Educational Studies, Athens, OH 45701

brooksg@ohio.edu

Key Words: *teaching statistics, computer software, descriptive statistics, correlation*

It is important for students in applied statistics courses to understand the statistics they use, both conceptually and mathematically. As computers have become more prevalent, fewer students are calculating statistics by hand, perhaps losing the mathematical insight provided by performing such calculations. The FISH: Friendly Introductory Statistics Help program has been written to help students gain these mathematical insights without actually performing the calculations themselves. The FISH program helps students to learn the mathematics behind introductory descriptive statistics such as standard deviations, z scores, and correlations. The program performs calculations in a step-by-step fashion so that students can visualize calculations as they are made. The program accepts data entered by the user or will generate data. FISH creates output such as frequency tables and histograms for single variables and predicted values and scatterplots for two variables. The program also can generate multiple samples, providing an example of a sampling distribution of means. The purpose of this presentation is to demonstrate the free FISH program.

The Institutional Review Board: Friend or Foe in a Graduate Student's Career

◆ Terry Tomazic, Saint Louis University; Jennifer Cipfl, St. Louis University; Will Haynes, Saint Louis University; Kevin McIntyre, Saint Louis University; Mark Morgan, Saint Louis University; Christine Ross, Saint Louis University; Melanie Shaafsma, Saint Louis University; Katrina Taylor, Saint Louis University; Richard Wosman, Saint Louis University

Saint Louis University, 3750 Lindell Blvd., McGannon Hall, Suite 230, St. Louis, MO 63108

tomazictj@slu.edu

Key Words: *Institutional Review Board, internet surveys, attitudes*

Graduate students face many challenges as they begin their career in research. They must master the skills and theories in their respective discipline, learn various methodologies and statistics for answering research questions in their discipline, and then actually carry out a research project of their own. Along the way, those involved in research with human subjects encounter the Institutional Review Board. How do those graduate students perceive that IRB? An internet survey of graduate students at a midwestern university provides a glimpse of the knowledge, attitudes, and perceptions regarding this part of a graduate student's career. Overall, attitudes toward the IRB are generally positive. However, most still felt confused or frustrated by the process. This confusion or frustration was not alleviated by level of knowledge. Such findings can be valuable to IRB's and other programs that must train graduate students in the details of research dealing with human subjects.

Incorporating Critical Reading and Writing into an Introductory Statistics Course

◆ Elizabeth Walters, Loyola College in Maryland

Loyola College in Maryland, Dept. of Mathematical Sciences, 4501 N. Charles St., Baltimore, MD 21210

ewalters@loyola.edu

Key Words: *introductory statistics, teaching methods*

Over the past few years, added emphasis in undergraduate education has been placed on teaching across the disciplines. A main concern has been the ability of students to be careful and critical readers and writers. This poster session will present techniques for strengthening the reading and writing skills of undergraduate students by incorporating these skills into an introductory statistics course. Student comments on these techniques, both positive and negative, will be shared. A freshman program at Loyola College in Maryland, called The Alpha Program, intended to develop the habits of careful reading, critical thinking, and scholastic conversation in freshman students across all disciplines, will also be discussed.

Course Materials for a Hybrid Introductory Statistics Course

◆ Patti B. Collings, Brigham Young University; Paul J. Fields, Brigham Young University

Brigham Young University, 231 TMCB, Provo, UT 84602

collingp@byu.edu

Key Words: *teaching, hybrid, course materials, online*

Faculty in the Department of Statistics at Brigham Young University have developed course materials for use in their hybrid introductory statistics course. The materials are designed so that an instructor can decide which lessons to discuss in class and which to have the student do independently online. The materials were developed using Macromedia Flash and QuickTime. Concept demonstrations, movie clips, applets, and self-assessment quizzes are included in the online materials. Corresponding materials were created for use in the classroom. These materials will be demonstrated in the poster session.

A Student-run Consulting Program Providing Community Service

◆ Jing Cheng, Purdue University; Nilupa S. Gunaratna, Purdue University; John R. Stevens, Purdue University

Purdue University, Dept. of Statistics, 150 N. University St., West Lafayette, IN 47907-2067

cheng14@stat.purdue.edu

Key Words: *statistical consulting service, graduate student training, volunteer, service*

In 2003, graduate students in Purdue University's Department of Statistics assisted a mayor-appointed task force with the design and analysis of a citizen survey to help develop an enhancement plan for a section of the City of West Lafayette, Indiana. This consulting was provided free of charge by Statistics in the Community, or STATCOM, a student-run statistical consulting program to provide service to local nonprofit and government groups. Other clients have included school districts and community centers. Involvement in this program allows students to enhance their education through a unique combination of consulting and community service. We present a possible model to develop such a program at other universities, including the organizational structure, methods to attract community clients, and the recruitment and training of student consultants.

Trashball: A Logistic Regression Classroom Activity

◆ Christopher H. Morrell, Loyola College in Maryland; Richard E. Auer, Loyola College in Maryland

Loyola College in Maryland, Mathematical Sciences Dept., 4501 North Charles St., Baltimore, MD 21210

chm@loyola.edu

A classroom activity is described that may be used to motivate and illustrate the use of logistic regression for a binary response variable. The activity involves students attempting to toss a ball into a trash can from various distances. The outcome is whether or not the student is successful in tossing the ball into the trash can. The resulting data allows the instructor to discuss the type of response variable, the need to a logistic regression model, fitting and interpreting the model, multiple logistic regression, and variable selection in this context.

Finitizing Power Series Distributions

◆ Martin S. Levy, University of Cincinnati; Saeed Golnabi, Mellon Financial Corporation

University of Cincinnati, Dept. of Quantitative Analysis, PO Box 21030, Cincinnati, OH 45221

martin.levy@uc.edu

Key Words: *discrete distribution, infinite support, finite support, Taylor series expansion, moments, matching distribution*

We introduce a methodology called finitization that converts an infinitely supported discrete power series distribution into another distribution having support of specified finite size, and one having finite support into a density with an even smaller sized support. We illustrate the methodology with four well-known power series distributions: Poisson, negative binomial, binomial, and logarithmic. An interesting moment-preserving property is presented, namely the first n moments of a finitized power series distribution coincide with those of its parent distribution. Some finitized distributions have interesting and unexpected applications. We offer suggestions on how to incorporate finitization into intermediate probability courses.

Contracting for Research and Development Services

◆ E. Ann Dimler, U.S. Census Bureau

U.S. Census Bureau, 4700 Silverhill Rd., Rm. 3134-4, Stop 9100, Suitland, MD 20746

Elizabeth.Ann.Dimler@census.gov

Key Words: *contracting, research, survey, methodology*

The Census Bureau has successfully contracted for 11 years to obtain the services of statisticians, methodologists, social scientists, data analysts, cost and management analysts, marketing specialists, and computer specialists to complete some of its most challenging and difficult research and development activities. The research cuts across all Census Bureau programs and covers every process associated with survey design and implementation. The services are procured using the Research and Development Contracts which are indefinite delivery, indefinite quantity multiple task order contracts. Services are concentrated in six highly technical areas: Assessment, Planning and Analysis, Data Analysis and Dissemination, Statistical Analysis, Methodological Research, Sub-Population Research, and Survey Engineering. The contractors provide a pool of specialists/experts to tackle some of the Census Bureau's most difficult research and include educational institutions, university-supported firms, and privately owned firms that concentrate in sample survey research, methodology, and applications.

346 Health Policy Statistics Speaker Luncheon (Fee Event)

Section on Health Policy Statistics

Wednesday, August 11, 12:30 pm–1:50 pm

Data Needs for Monitoring Population and Public Health in Canada

◆ John Frank, University of Toronto

University of Toronto, CIHR - Institute of Population and Public Health, 100 College St., Banting Bldg., Suite 207-L, Toronto, ON M5G 1L5 Canada

john.frank@utoronto.ca

Key Words: *health status, public health system, health policy*

This presentation will examine indicators for measuring Canadian population health, including average levels and inequalities in distribution of health status, with a focus on life cycle. Key elements of a successful modern national public health system will be identified, and programming and policy implications discussed.

347 Biopharmaceutical Section Roundtable Luncheons (Fee Event)

Biopharmaceutical Section

Wednesday, August 11, 12:30 pm–1:50 pm

Utility of Flexible Designs in Drug Development?

◆ H.M. James Hung, U.S. Food and Drug Administration

U.S. Food and Drug Administration, DB1, CDER, Room 5062, 1451 Rockville Pike, Rockville, MD 20852

hung@cder.fda.gov

A number of flexible designs have been proposed in statistical literature. Most of them embed multiple stage planning. The concept of multiple stage planning seems natural. How should such stepwise planning fit in the current framework of the three-phase clinical trials? What are anticipated obstacles needing resolution? What are the issues with statistical inference for the data resulted from such designs?

The Composition, Role, and Operating Features of Data Safety Monitoring Committees

◆ Kenneth J. Koury, Schering-Plough Research Institute

Schering-Plough Research Institute, 2015 Galloping Hill Rd., K-15-2,2435, Kenilworth, NJ 07033

kenneth.koury@spcorp.com

"DSMCs" has been a topic of great interest in the recent past. Varying views of their compostion, role, and operating features have been expressed by experts in the field. We recently conducted a survey of the implementation of DSMCs by several companies. We would like to share the results of the survey and discuss certain aspects of DSMCs in detail.

Practical Issues in Sample Size Re-estimation in Clinical Trials

◆ Jeff Maca, Novartis Pharmaceuticals

Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07869

jeff.maca@pharma.novartis.com

Key Words: *sample size re-estimation, clinical trials*

Planning and designing clinical trials depends in many ways upon various types of information. When many trials are undertaken, much of this information will not be available at the desired level of precision, or possibly there is no directly relevant estimate available. There will almost certainly be some differences between the parameter in the current trial and the source of the estimate. Indeed, it may frequently be the case that the final data from a clinical trial provides the first opportunity to confirm or contradict the sample size assumptions. One may see at that point that these incorrect assumptions led to a trial which had considerably different power than was anticipated. The potential implications are that a successful trial outcome may be jeopardized because power was insufficient, or substantial resources may have been wasted because the trial was needlessly large. The possibility of updating the sample size as the trial progresses if the observed data seems to disagree with the pre-trial assumptions would be useful in this case. The practical issues pertaining to designing and running such a trial will be discussed.

Use of Composite Endpoints in Time-to-event Clinical Trials

◆ Nacer E. Abrouk, Amgen, Inc.

Amgen, Inc., One Amgen Center Dr., 24-2C, Thousand Oaks, CA 91320-1799

nabrouk@amgen.com

Key Words: *composite endpoints, cardiovascular mortality and morbidity, time-to-event trials*

Various aspects of composite endpoints in time-to-event clinical trials will be discussed. Specific issues such as power, precision, analysis methods, clinical interpretation, and regulatory viewpoint will be addressed.

Quantitative Pharmacovigilance for Signal Detection

◆ A. Lawrence Gould, Merck & Co., Inc.

Merck & Co., Inc., BL3-2, West Point, PA 19486

goulda@merck.com

Key Words: *Bayesian*

Quantitative pharmacovigilance often uses proportional reporting ratios (PRRs) to identify drug-event associations reflecting possible toxicity in spontaneous reporting system (SRS) databases. Reports from SRS databases generally cannot be used to establish causality, but may identify potential toxicity issues to be confirmed clinically and epidemiologically. PRRs will be imprecise if there are few reports. Recently described Bayesian methods provide a statistically well-founded way to incorporate this uncertainty into deciding whether a particular finding represents a potential signal or just noise. Quantitative pharmacovigilance methods may be useful for the public health purpose of identifying previously unsuspected potential associations between drugs and combinations of adverse events as well as for ascertaining the adverse-event reporting profile corresponding to a single drug or a class of drugs. Even a small number of event reports may signal a potential association between a drug and the event if the event is rarely reported with other drugs. The methods also allow investigation of complex relationships between drugs, events, and concomitant variables.

348 Section on Bayesian Statistical Science Roundtable Luncheons (Fee Event)

Section on Bayesian Statistical Science

Wednesday, August 11, 12:30 pm–1:50 pm

Bayesian Methods in Bioinformatics Research

◆ Jun S. Liu, Harvard University

Harvard University, Dept. of Statistics, 1 Oxford St., Cambridge, MA 02138

jliu@stat.harvard.edu

Key Words: *gene network, machine learning, microarray, MCMC, sequence analysis, transcription regulation*

Discussions will be focused on pressing issues in bioinformatics research, especially those that are at the interface of statistics, genetics, and genomics. We show that the Bayesian modeling approach is a suitable framework for combining and synthesizing information and for handling such complex problems as understanding gene regulation, inferring relationships among proteins, and analyzing functionalities of conserved structural cores of proteins. More detailed mathematical formulations will be given in the short course "Tutorial on the Bayesian methods in bioinformatics" given by JSL.

Missing Data in Regression Models

◆ Joseph G. Ibrahim, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Dept. of Biostatistics, McGavran-Greenberg Hall, Chapel Hill, NC 27599

ibrahim@bios.unc.edu

Key Words: *EM algorithm, Gibbs sampling, maximum likelihood, missing at random, nonignorably missing, sensitivity analyses*

Missing data is a major issue in many applied problems, especially in the biomedical sciences. We will discuss four common approaches for inference in missing data problems. These are (i) Maximum Likelihood (ML), (ii) Multiple Imputation (MI), (iii) Weighted Estimating Equations (WEE), and Fully Bayesian (FB) methods. There is considerable interest as to how these four methodologies are related, the properties of each approach, the advantages and disadvantages of each methodology, computational implementation, software, and practical considerations of the methodologies. We will examine data that is missing at random (MAR) and nonignorably missing. Issues regarding model identifiability, model assessment, sensitivity analyses, and robustness will also be discussed.

349 Section on Government Statistics Roundtable Luncheon (Fee Event)

Section on Government Statistics

Wednesday, August 11, 12:30 pm–1:50 pm

Is It Good Enough? Comparing Item Quality Across Surveys

◆ Pat Doyle, U.S. Census Bureau

U.S. Census Bureau, Demographic Surveys Division, Washington, DC 20233-8400

patricia.j.doyle@census.gov

How do you assess the quality of an item for which there is no good independent benchmark? How do you compare item quality among multiple surveys that use different procedures and have varying levels of precision and response rates? Pat Doyle, U.S. Census Bureau, will lead a discussion based on the work of the Item Quality subgroup of the Interagency Household Survey Nonresponse Group, a subcommittee of the Federal Committee on Survey Methodology.

350 Section on Physical and Engineering Sciences Roundtable Luncheons (Fee Event)

Section on Physical and Engineering Sciences

Wednesday, August 11, 12:30 pm–1:50 pm

Issues in Design and Analysis of Discrete Choice Experiments

◆ Stephen P. Jones, Boeing Company

Boeing Company, 6709 Elderberry Ave. SE, Snoqualmie, WA 98065

stephen.p.jones@boeing.com

Key Words: *discrete choice experiments, design of experiments*

This roundtable will give an overview of the popular market research tool of discrete choice experiments, illustrating the methods with some examples from product development at the Boeing Company. We will discuss some of the statistical issues that distinguish discrete choice experiments from the typical application of design of experiments methodology. This roundtable is intended both for practitioners and researchers. The objective is to encourage new people to enter this area of research as well as to allow practitioners and researchers to share some insights about issues that they believe are of current interest from both a practical and a research perspective.

Using Data from Customers to Improve Engineering Decisions

◆ Jeffrey A. Robinson, General Motors R&D Center

General Motors R&D Center, Mail Code 480-106-256, Vehicle Development Research Lab, Warren, MI 48090-9055

jeffrey.a.robinson@gm.com

Key Words: *customer, engineering, decisions*

Companies often collect a lot of data about their prospective customers and product feedback from their current customers. But using the data effectively is a challenge. This roundtable discussion will summarize some of the issues related to using customer data for effective engineering decision making in the auto industry. It is intended for those wanting to learn a little more as well as for experts from other industries willing to share their experience. We are likely to raise more questions than answers, but the goal is to encourage statisticians to think more seriously about ways to help engineers make better decisions. Some of the data sources could include warranty data, customer satisfaction surveys, and product preference data from customers. Some of the decisions engineers have to make are how to run the plant, how to allocate resources to solve product problems, and how to incorporate customer feedback into future designs.

351 Section on Quality and Productivity Roundtable Luncheons (Fee Event)

Section on Quality and Productivity

Wednesday, August 11, 12:30 pm–1:50 pm

Design of Experiments Trials and Tribulations

◆ Shari Kraber, Stat-Ease, Inc.

Stat-Ease, Inc., 2021 E. Hennepin Ave., Suite #480, Minneapolis, MN 55413

shari@statease.com

Key Words: *design of experiments*

Participate in a lively discussion of the trials and tribulations of planning and running designed experiments. Plan to share your experiences and learn from the experiences of others. Discuss the most common pitfalls that experimenters encounter and learn how to avoid them. Explore problems with fractional factorials, mixtures, missing data, pass/fail data, etc.

A Mixed Salad: Tips, Issues and Learnings on Applying Six Sigma to Transactional Projects

◆ Michael Cook, DuPont Company

DuPont Company, PO Box 80840, Wilmington, DE 19880-0840

michael.cook@usa.dupont.com

Key Words: *Six Sigma, transactional projects, business processes*

Applying Six Sigma to improve the performance of manufacturing processes and product quality is well known. To broaden use and understanding of Six Sigma in other markets, this discussion will focus on applying Six Sigma to business and transactional and service processes. We'll cover items related to teaching and coaching people with transactional projects as well as project execution and useful tools. We'll discuss this in the context of projects for different transactional arenas, such as engineering, construction, purchasing, human resources, facility services, consulting, finance, and the hospitality industry. The goal is to make participants more aware of opportunities on "how" and "where" to apply Six Sigma outside the mainstream of manufacturing.

352 Section on Statistical Graphics Roundtable Luncheon (Fee Event)

Section on Statistical Graphics

Wednesday, August 11, 12:30 pm–1:50 pm

Interactive Linked Micromap Plots

◆ Juergen Symanzik, Utah State University

Utah State University, Dept. of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322-3900

symanzik@math.usu.edu

Key Words: *federal data, graphics, visualization, web*

Linked micromap (LM) plots provide a new statistical paradigm for the viewing of geographically referenced statistical summaries in the corresponding spatial context. The main idea behind LM plots is to focus the viewer's attention on the statistical information presented in a graphical display. Multiple small maps are used to provide the appropriate geographic reference for the statistical data. Linked micromap plots were first presented to the statistical community at the 1996 Joint Statistical Meetings in Chicago. Since then, micromaps have experienced many upgrades and are now available for use on interactive web sites. We will discuss basic ideas of micromaps, how to produce them in S-Plus or nViZn, and which updates have to be made to the NCI Java code to be able to install a local interactive micromap web server. Source code can be made available upon request.

353 Section on Statistics and the Environment Roundtable Luncheon (Fee Event)

Section on Statistics and the Environment

Wednesday, August 11, 12:30 pm–1:50 pm

Cluster Detection in Environmental Health

◆ Andrew B. Lawson, University of South Carolina

University of South Carolina, 800 Sumter St., Columbia, SC 29208

alawson@gwm.sc.edu

Key Words: *cluster, modeling, space-time, object recognition, parameterization*

Cluster detection is very important in the analysis of environmental health data. In both the spatial and temporal domain the isolation of "unusual" aggregations of cases of disease is of importance. Public Health departments are often asked to analyse what is perceived to be a "cluster" of disease post hoc. There are many approaches to this problem varying in their degree of parameterization, and paradigm. Various issues related to how clusters are to be detected will be considered, including the definition of clusters, use of random effects, object recognition, space-time clusters, and model parameterization. The role of testing vs. modeling will also be covered.

354 Section on Statistics in Epidemiology Roundtable Luncheon (Fee Event)

Section on Statistics in Epidemiology

Wednesday, August 11, 12:30 pm–1:50 pm

Future Directions for the Design and Analysis of Nutritional and Environmental Studies with Exposure Measurement Error

◆ Donna Spiegelman, Harvard School of Public Health

Harvard School of Public Health, Dept. of Biostatistics and Epidemiology, 677 Huntington Ave., Boston, MA 02115

stdls@channing.harvard.edu

What are the limitations to current methods for bias correction of generalized regression slopes with covariate measurement error? Why are these methods used less frequently than we might hope by practitioners? What are the most fruitful directions for future research? The roundtable leader will share her experiences developing and applying these methods in the Nurses' Health Study at Harvard University, and hopes others will come prepared to do the same.

355 Section on Survey Research Methods Roundtable Luncheons (Fee Event)

Section on Survey Research Methods

Wednesday, August 11, 12:30 pm–1:50 pm

Integrated Survey and Administrative Data

◆ Julia I. Lane, Urban Institute

Urban Institute, 2100 M St. NW, Washington, DC 20037

jlane@ui.urban.org

Key Words: *integrated data, employer-employee data, confidentiality*

The focus of this discussion will be on discussing the various challenges and rewards derived from integrating survey and administrative data, with lessons learned from a variety of experiences in different countries. Topics will include (but are not limited to): addressing policy challenges; developing a ble database infrastructure; developing products and a user constituency; identifying demand; interpreting data to create ble information; using geography as an integration element; addressing confidentiality challenges. Although the main discussion will be focused on the lessons learned from the development of integrated employer-employee datasets, there will also be a discussion of the integration of other data—notably health services research and geographic research—and the similarities and differences in the challenges that are faced.

Respondent Retention in Longitudinal Surveys

◆ Rupa Datta, NORC, University of Chicago

NORC, University of Chicago, 712 Walker Ave., Oakland, CA 94610

datta@norcmail.uchicago.edu

Key Words: *attrition, response rates, longitudinal survey, panel survey*

In March, 2004, representatives of many major American longitudinal surveys and longitudinal survey sponsors will convene for a one-day Conference on Respondent Retention in Longitudinal Surveys sponsored by the Technical Review Committee of the National Longitudinal Surveys of the Bureau of Labor Statistics, U.S. Department of Labor, and the National Institute of Child Health and Human Development. The Conference is structured as a dialogue that helps to assess the current survey environment and to exchange accumulated knowledge about respondent retention in longitudinal surveys. One set of discussions will address determinants of attrition, including the personal characteristics of attriters, how prior round participation contributes to attrition, and the role of data collection mode in exacerbating or stemming attrition. A second set of discussions will focus on how surveys respond to attrition, including payment of incentives, structuring the interview to avoid attrition, and adopting fielding strategies that minimize attrition. The roundtable will be an opportunity to review key ideas and evidence presented during the March conference, and to extend those ideas.

Web Survey Design and Implementation

◆ Mick P. Couper, University of Michigan

University of Michigan, Survey Research Center, PO Box 1248, Ann Arbor, MI 48109

mcouper@umich.edu

Key Words: *web survey, internet survey*

Web surveys are becoming increasingly common as an alternative or supplement to telephone and mail surveys. Web surveys introduce their own set of issues for survey methodologists and statisticians. This session will be a forum for sharing recent experiences and insights about this increasingly important arena.

356 Social Statistics Section Roundtable Luncheon (Fee Event)

Social Statistics Section

Wednesday, August 11, 12:30 pm–1:50 pm

New Challenges Conducting Surveys of Physicians

◆ Marc Berk, NORC, University of Chicago; Sara Thran, American Medical Association

NORC, University of Chicago, 7500 Old Georgetown Rd., Bethesda, MD 20878

berk-marc@norc.net

Key Words: *physicians, response rates*

Health policy researchers often rely on surveys of physicians and other medical providers. Over the last 20 years it has become increasingly difficult to survey this population and these difficulties may be accelerating. Physicians are more mobile and difficult to locate. Changes in the organization of physician practices may decrease the autonomy physicians have in making participation decisions. New privacy regulations may also contribute to physician reluctance to participate. We will discuss recent experience in conducting physician surveys. Emphasis will be placed on identifying design features that facilitate the implementation of high-quality surveys of medical providers. Items to be covered include mode of interview, use of knowledgeable proxies, and payment of monetary incentives.

357 Section on Teaching of Statistics in the Health Sciences Roundtable Luncheon (Fee Event)

Section on Teaching Statistics in the Health Sciences

Wednesday, August 11, 12:30 pm–1:50 pm

The Role of Statisticians in Teaching Evidence-based Practice to Future Clinicians

◆ Renee Stolove, New York Medical College

New York Medical College, School of Public Health, Program in Physical Therapy, Valhalla, NY 10595

Renee_Stolove@nymc.edu

Key Words: *teaching, evidence-based practice, health sciences, curriculum*

The aim of this discussion is to explore the role of all stakeholders in teaching statistics to future clinicians. Emphasis on Evidence-Based Practice (EBP) has increased the importance of statistical literacy for clinicians and the need to critically evaluate the literature. The manner in which statistical skills are being taught is changing. Rather than being presented in isolated courses or modules, the material is frequently integrated into EBP and clinical courses. Clinical faculty has been assuming increasing responsibility for this. We will discuss the importance of collaboration between statistical and clinical faculty with regard to curricular planning and teaching.

358 Section on Risk Analysis Roundtable Luncheon (Fee Event)

Section on Risk Analysis

Wednesday, August 11, 12:30 pm–1:50 pm

Risk within Transportation

◆ Cliff Spiegelman, Texas A&M University/The Texas Transportation Institute

Texas A&M University/The Texas Transportation Institute, Laboratory for Statistical Science in Chemistry, Dept. of Statistics-3143, College Station, TX 77843-3143

cliff@stat.tamu.edu

Key Words: *highway safety, risk, cell phone, airline passenger screening, transportation*

This roundtable will discuss issues ranging from risks associated with highway safety from cell phone use to building barriers to minimize damage from truck bombs to the screening of airline passengers.

359 Section on Statistical Education Roundtable Luncheons (Fee Event)

Section on Statistical Education

Wednesday, August 11, 12:30 pm–1:50 pm

Learning Styles in the Introductory Statistics Classroom

◆ Carmen O. Acuna, Bucknell University

Bucknell University, Dept. of Mathematics, Lewisburg, PA 17837

cacuna@bucknell.edu

Key Words: *Bloom's taxonomy*

The statistics education reform of the last 15 years has substantially changed the way in which the introductory statistics course is taught. The focus has shifted from numerical and symbolic manipulations to an emphasis on the main elements of statistical thinking and the big ideas of statistics. This change has added to the degree of abstraction of the class. Students who must take this course as part of a major requirement still struggle with the introductory statistics course; now for different reasons. We'll discuss how to capitalize on the students' different learning styles to elicit from them the higher levels of thinking that this shift in emphasis is demanding from them.

Fallacies and Myths in Elementary Statistics

◆ Bernard Harris, University of Nebraska

University of Nebraska, 917 Oldfather Hall, Lincoln, NE 68588-0323

bharris4@unl.edu

Key Words: *teaching statistics, fallacies, myths*

SOLD OUT

The purpose of this roundtable discussion is to highlight several myths and fallacies encountered in elementary statistics texts. These include the role of the number 30 in justifying the use of normal approximations, the treatment of outliers, and the use of squared error loss in estimation. Many textbooks provide incorrect methods for the calculation of quantiles—in particular, the calculation of quartiles. The role of the mean and median as descriptive measures is often incorrectly stated. Also, the use of pooled variance estimates in two sample t-tests and the pooling of the variance estimate in regression are sometimes given in a contradictory manner. Some exercises and textbook examples on probability contradict the role of probabilistic methods in statistical inference. This discussion is intended to be controversial and will provide opportunity for feedback and debate.

Service Learning Opportunities in Statistics Courses

◆ Thomas H. Short, Indiana University of Pennsylvania

Indiana University of Pennsylvania, Mathematics Dept., 210 S. 10th St., Indiana, PA 15705

tshort@iup.edu

Key Words: *nonprofit organization, community service, real data, consulting*

Service learning involves students completing relevant work that fills a real unmet need in a community. Participants at this roundtable will explore possibilities for service learning in statistics courses and discuss the implementation of service learning opportunities for statistics students. Specific topics will include identifying statistical needs in the community, tailoring projects to a statistics course syllabus, incorporating a service component into a statistics course, assigning statistical tasks to students, and ideas for assessment in service learning statistics.

360 Complex Data Structures

Section on Statistical Graphics, SSC

Wednesday, August 11, 2:00 pm–3:50 pm

Some New Developments in the Design and Analysis of Computer Experiments

◆ Derek R. Bingham, Simon Fraser University; Crystal Linkletter, Simon Fraser University; David Higdon, Los Alamos National Laboratory; Nicolas Hengartner, Los Alamos National Laboratory

Simon Fraser University, Dept. of Statistics and Actuarial Science, Burnaby, BC V5A 1S6 Canada

dbingham@cs.sfu.ca

Key Words: *Bayesian variable selection, model calibration, screening experiments, spatial process*

The design and analysis of experiments continues to make important and far reaching contributions to scientific investigation. The rapid growth in computing power has made the computational simulation of complex systems feasible and helped avoid more physical experimentation. Consequently, the design and analysis of computer experiments has become an integral part in the exploration of scientific and industrial processes. This talk has two main features: new approaches to identifying important factors in computer experiments are introduced and their impact on design selection is discussed, and the research goals of the NPCDS initiative on computer experiments will be described.

Finding Differentially Expressed Genes in Microarray Experiments Augmented by Gene Annotation Data

◆ Rafal Kustra, University of Toronto

University of Toronto, Dept. of Public Health Sciences, Faculty of Medicine, McMurrich Bldg., #307, Toronto, ON M5S-1A8 Canada

r.kustra@utoronto.ca

Key Words: *multivariate analysis, regularization, microarray, differential expression, high-dimensional data*

High-throughput genomics (HPG), including microarray techniques, is one of the biggest stories in biology and life sciences, and will likely map one of the key future directions for biostatistics and applied statistics. The HPG data are uniquely challenging to statisticians, and classical methods of our field are rarely applicable directly. One of the defining elements of HPG datasets is a huge number of correlated measurements done on small number of independent samples. Typically in microarray studies, tens of thousands of genes have their expression levels probed, but the number of samples are between a dozen and few hundred. A number of ways was proposed to deal with this huge-dimensionality problem, either as a pre-analytic step (clustering genes, PCA) or incorporated in the analysis. We will describe our approach that incorporates auxiliary sources of information, such as geneontology classification, to build effective regularizers into our models. We concentrate on the problem of finding differentially expressed genes and use crude measures of similarity based on hierarchical, GO annotation to augment the covariance matrix in multivariate models.

Shift Function Plots for Regression Fitting

◆ David Bellhouse, University of Western Ontario; Zilin Wang, University of Western Ontario

University of Western Ontario, Dept. of Statistics and Actuarial Science, London, ON N6A 5B7 Canada

bellhouse@stats.uwo.ca

Key Words: *parametric and nonparametric regression, bootstrapping*

Although a nonparametric regression model allows us to obtain a graphical display of the functional relationship between the response and independent variables, the exact form of the regression function is not evident. Such a functional relationship is explicit in a parametric model. Our objective is to introduce a new graphical approach, called the shift function plot, with which a hypothesis is constructed to evaluate the goodness of fit of a parametric regression model. We parameterize the model under the null hypothesis. Then we use the nonparametric representation as the unrestricted regression model through which we examine how well the parametric model fits the data. Inferential issues relating to the shift functions are examined and a chi-square test statistic is developed with the aid of bootstrapping.

361 Chemometrics ▲

Section on Physical and Engineering Sciences, Section on Statistical Graphics, Section on Quality and Productivity

Wednesday, August 11, 2:00 pm–3:50 pm

Interpretability of Trees From High-throughput Screening Data

◆ Yan Yuan, University of Waterloo; William J. Welch, University of British Columbia; Hugh Chipman, University of Waterloo

University of Waterloo, Dept. of Statistics and Actuarial Science, Waterloo, ON N2L 3G1 Canada

y4yuan@math.uwaterloo.ca

Key Words: *high-throughput screening (HTS), sequential screening, structure-activity relationship, classification tree, tree harvesting*

Millions of compounds are available as potential drug candidates today, and high-throughput screening (HTS) is widely used to assay compounds for activity against a particular biological target. Often, classification trees are built to relate activity to variables characterizing chemical structure. These models are used to predict the activity of unscreened compounds, and hence select promising compounds to be screened in a sequential strategy. With a large dataset and a complex relationship between chemical structure and activity, a tree with hundreds of nodes may result. We describe a "tree harvesting" algorithm that simplifies a tree, making it easier to interpret. This facilitates the identification of interesting groups of compounds, to serve as diverse leads for drug optimization. Several datasets will be used to illustrate the methodology.

Variable Selection for Prediction Where There are an Excessive Number of Variables

◆ Jack Liu, NISS; S. Stanley Young, NISS

NISS, PO Box 14009, RTP, NC 27709

Jack_Liu@niss.org

Key Words: *variable selection, n<p, over-determined systems, prediction*

Assay systems are now available where a very large number of descriptors are available for each sample. For microarrays and proteomics there can be tens of thousands of descriptors. New methods of assay for metabolites can give quantitative values for thousands of descriptors. Often there are only tens or a few hundred samples where we want to predict some characteristic of the sample, tumor versus normal or diseased versus normal. Predictions can be made using linear combinations of large numbers of predictors or by focusing on small sets of predictors. There is a need to have interpretable decision rules so we will focus on small sets of predictors in trying to understand these complex datasets. We use a microarray and a metabolic dataset to demonstrate our methods.

Pooling for Detection of Synergistic Compounds

◆ Jacqueline M. Hughes-Oliver, North Carolina State University

North Carolina State University, Dept. of Statistics, Box 8203, Raleigh, NC 27695-8203

hughesol@stat.ncsu.edu

Key Words: *antagonism, blocking, drug discovery, group testing, high-throughput screening, synergism*

Drug discovery is a multiphase process that requires identification of chemically distinct classes of lead compounds. Given a particular target or biological assay, lead compounds must demonstrate high potencies. Screening to identify potential leads or active compounds can be both time-consuming and expensive because of the large numbers of compounds available in either real or virtual libraries. Motivated by historical observations that only a small fraction, say 5%, of tested compounds are active to a particular assay, some pharmaceutical companies now use pooling experiments for screening. Pooling experiments, where multiple compounds are tested in a single well, are cost-effective and information-dense. They are also ideal for detecting synergism, where two or more compounds are required to achieve activity. Unfortunately, their analysis can be fraught with difficulties. One such difficulty arises from blocking mechanisms that result in an inactive pool containing active compounds. I will present some methods for analyzing pooling experiments, discuss their successes and failures, and illustrate application to a real dataset.

362 Practical Issues in Resampling Environmental Data ▲ ⌘

Section on Statistics in Epidemiology, Section on Statistics and the Environment, Section on Survey Research Methods

Wednesday, August 11, 2:00 pm–3:50 pm

Bootstrapping-dependent Data: Difficulties and Potential Solutions

◆ Mark L. Taper, Montana State University

Montana State University, Dept. of Ecology, Bozeman, MT 59717

taper@rapid.msu.montana.edu

Key Words: *bootstrapping, dependent data, parametric bootstrap, estimating equations*

Bootstraping is an increasingly important statistical tool, allowing statistical inferences to be made in situations where analytic solutions are difficult. However, bootstrapping is not a panacea. A fundamental assumption of bootstraping is that the resampled entities are independent. In many ecological problems, such as time-series, spatial data, and genetic data independent entities are dificult to find. I discuss strategies for constructing useful bootstrap inference in such cases using parametric bootstrapping, conditioning to achieve independence, bootstraping of evidence functions, and bootstraping of random effects.

Practical Issues in Resampling

◆ Tim C. Hesterberg, Insightful Corporation

Insightful Corporation, 1700 Westlake Ave. N, Suite 500, Seattle, WA 98109-3044

timh@insightful.com

Key Words: *bootstrap, standard errors, central limit theorem, finite population*

I'll focus on some practical issues in resampling. The first is that resampling is really only practical using software that makes it easy. I'll demonstrate S-PLUS software, suitable for use by introductory statistics students on up. The graphical diagnostics included give intro stat students a better concept of sampling variability and distributions, and demonstrate to more advanced statisticians that traditional inferences based on normal approximations are often woefully inaccurate. A second practical issue is the number of bootstrap samples used; in particular you need a lot for Bca confidence intervals, many fewer for bootstrap tilting intervals. A third practical issue is that resampling should mimic how the data were collected. This is not always straightforward. For example, the naïve way of doing finite-population bootstrapping can result in greater variance than sampling without replacement. The final issue is that bootstrapping underestimates standard errors, substantially in the case of stratified sampling with small strata. There are four simple remedies.

Essential Modifications of Bootstrap for Some Common Problems

◆ David V. Hinkley, University of California, Santa Barbara

University of California, Santa Barbara, Dept. of Statistics and Applied Probability, Santa Barbara, CA 93106

hinkley@pstat.ucsb.edu

Key Words: *bootstrap*

The simplicity of the basic bootstrap idea and its application makes it very attractive for calculating standard errors, confidence limits, and so forth. However, for some fairly common but complex modeling problems, modifications are necessary to incorporate the correct pattern of variability in the simulated datasets. The main focus of this talk will be on regression methods, including generalized linear model fitting and variable selection methods (including principal component regression, wavelet smoothing, etc).

Graphical Diagnostics for the Bootstrap

◆ Angelo J. Canty, McMaster University

McMaster University, 218 Hamilton Hall, 1280 Main St. W, Hamilton, ON L8S 4K1 Canada

cantya@mcmaster.ca

Key Words: *bootstrap, diagnostics, graphics, discreteness, inconsistency, pivot*

Over the past 25 years the bootstrap has become one of the most widely used tools for many applied statisticians. Like any statistical method, the bootstrap is based on various assumptions and when these assumptions are violated the results of a bootstrap can be very misleading. In many applications of the bootstrap, however, the underlying assumptions are never tested. One of the reasons for this has been the lack of any useful diagnostic tools to assist the applied statistician in checking the assumptions. I will describe some simple graphical diagnostics that can be used to look for potential problems with the bootstrap and suggest remedies where possible. Among the problems that I shall consider are: discreteness of the bootstrap distribution, the effect of outliers, and inconsistency of the bootstrap method. Another topic that I shall examine is the choice of scale that makes a bootstrap statistic approximately pivotal and so stabilizes confidence interval calculations.

363 Statistics in Genomics ▲ ⌘

WNAR, Biometrics Section

Wednesday, August 11, 2:00 pm–3:50 pm

Statistical Methods for Detecting RNA Splice Sites

◆ Mark Segal, University of California, San Francisco

mark@biostat.ucsf.edu

Key Words: *mRNA splicing, random forests, classification, penalization*

Pre-messenger RNA splicing, whereby introns are precisely removed from primary transcripts, is an essential step in the expression of most eukaryotic genes, especially in humans. While a great deal of knowledge has been accumulated on splicing factors, much less is known about the nature of pre-mRNA sequences that are recognized by the nuclear splicing machinery for selection of splice sites. We pursue such characterization by utilizing a large dataset of real and decoy 5′ and 3′ splice sites, partitioned into training (189422) and test (94727) sequences. We apply classification procedures, including random forests and penalized techniques, and introduce some context-specific refinements. Comparisons with previously used maximum entropy modeling are drawn. Methods are contrasted with respect to predictive performance and interpretative yield.

Parametric and Nonparametric Empirical Bayes Adjustments to Multiple P Values for the Detection of Differentially Expressed Genes in Microarray Experiments

◆ Susmita Datta, Georgia State University

Georgia State University, Dept. of Mathematics and Statistics, 30 Pryor St., Atlanta, GA 30303-3083

sdatta@mathstat.gsu.edu

Key Words: *p values, microarray, empirical Bayes, differentially expressed genes*

An empirical Bayes adjustment to multiple t-tests has been shown to improve the sensitivity of the overall procedure. Here we propose an empirical Bayes adjustment to the p values rather than the test statistics. As a result, it is applicable to other types of multiple tests (e.g., F, chi-squared) as well. Both parametric and nonparametric versions of the empirical Bayes adjustment are considered. Thus, each p value, in turn, borrows evidence from other p values across the tests. A new set of accept/reject decisions are reached for each null hypothesis using the empirical Bayes adjusted p values through a resampling based step-down p value calculation that protects the analyst against the overall (familywise) Type 1 error rate. The new procedure is shown to produce further improvement in sensitivity in a number of examples.

Heterogeneity of Variance in Gene Expression Microarray Data

◆ David M. Rocke, University of California, Davis

University of California, Davis, Dept. of Applied Science, One Shields Ave., 2343 Academic Surge Bldg., Davis, CA 95616-8553

dmrocke@ucdavis.edu

Key Words: *empirical Bayes*

One important problem in the analysis of gene expression microarray data is that the variation in expression under constant conditions is not stable from gene to gene. Recently, variance stabilizing transformations have been developed that can remove the systematic dependence of the variance on the mean, but it appears that there is still considerable variance heterogeneity that can interfere with global analysis of expression data. We and others have developed methods to deal with this problem, mostly in an empirical Bayes framework.

Fraction of Missing Information with Selective Genotyping in QTL-mapping

◆ Saunak Sen, University of California, San Francisco

University of California, San Francisco, 94143

sen@biostat.ucsf.edu

Selective genotyping strategies are very effective in reducing genotyping costs at the cost of losing some information. We show, for the backcross, that there is a very simple formula for computing the fraction of missing information at a location that has been incompletely genotyped. Using this formula, we find that for almost any effect size, genotyping a quarter of both extremes of the population gives at least 80% of the information. Additionally, the bias in the Haley-Knott algorithm for computing LOD scores is related to the fraction of missing information.

364 The Lasso and LARS

Section on Statistical Computing

Wednesday, August 11, 2:00 pm–3:50 pm

1-norm Regularization: Efficient and Effective

◆ Saharon Rosset, IBM T.J. Watson Research Center; Ji Zhu, Stanford University

IBM T.J. Watson Research Center, 1101 Kitchawan Rd, Yorktown Heights, NY 10598

srosset@us.ibm.com

Key Words: *L1 regularization, flexible fitting, robust modeling, boosting, LARS, lasso*

We consider the general regularized optimization problem of minimizing loss+penalty, where the loss depends on the data and the model, and the penalty on the model only. We illustrate that the choice of L1 (lasso) penalty, in combination with appropriate loss, leads to several desirable properties: (1) Approximate or exact L1 regularization has given rise to highly successful modeling tools: the lasso, boosting, wavelets and 1-norm support vector machines. (2) L1 regularization creates sparse models, a property that is especially desirable in high-dimensional predictor spaces. We formulate and prove sparsity results. (3) L1 regularization facilitates efficient methods for solving the regularized optimization problem. The LARS algorithm takes advantage of this property. We present a general formulation of regularized optimization problems for which efficient methods can be designed. We show how we can create modeling tools which are robust (because of the loss function selected), efficient (because we solve the regularized problem efficiently) and adaptive (because we select the regularization parameter adaptively). (Joint work with Trevor Hastie and Rob Tibshirani.)

On Algorithms for Solving Least Squares Problems under an L1-type Penalty or Constraint

◆ Berwin A. Turlach, University of Western Australia

University of Western Australia, School of Mathematics and Statistics (M019), 35 Stirling Highway, Crawley, WA 6009 Australia

berwin@maths.uwa.edu.au

Key Words: *constrained least squares problem, constrained regression, convex programming, interior-point algorithm, quadratic programming, subset selection*

Tibshirani (1996) proposed the least absolute shrinkage and selection operator (LASSO) which estimates a vector of regression coefficients by minimizing the residual sum of squares subject to a constraint (penalty) on the sum of the absolute values of the coefficient estimates. We shall discuss several algorithms that can be used to calculate the LASSO solution. We will discuss the individual strengths and weaknesses of these algorithms by commenting, among others, on the accuracy of the calculated solution, the ease of their implementation (a somewhat subjective criterion}, and the ease with which they can be extended to similar but more complex problems. This talk is based on joint work with M.R. Osborne, B. Presnell, W.N. Venables, and S.J. Wright.

Variable Selection via COSSO in Nonparametric Regression

Yi Lin, University of Wisconsin, Madison; ◆ Hao H. Zhang, North Carolina State University

North Carolina State University, 8203 Campus Box, Raleigh, NC 27695-8203

hzhang2@stat.ncsu.edu

Key Words: *smoothing spline ANOVA, variable selection, lasso, shrinkage, nonparametric, soft thresholding*

We propose a new method for model selection and model fitting in nonparametric regression models, in the framework of smoothing spline ANOVA. The "COSSO" is a method of regularization with the penalty functional being the sum of component norms, instead of the squared norm employed in the traditional smoothing spline method. The COSSO provides a unified framework for several recent proposals for model selection in linear models and smoothing spline ANOVA models. Theoretical properties, such as the existence and the rate of convergence of the COSSO estimator, are studied. In the special case of a tensor product design with periodic functions, a detailed analysis reveals that the COSSO applies a novel soft-thresholding-type operation to the function components and selects the correct model structure with probability tending to one. We give an equivalent formulation of the COSSO estimator which leads naturally to an iterative algorithm. A novel one-step update algorithm is proposed, which is computationally fast and efficient. Simulations and real examples show that the COSSO gives very competitive performances when compared with other methods.

365 *JASA* Applications and Case Studies Invited Paper Session ▲

JASA, Applications and Case Studies

Wednesday, August 11, 2:00 pm–3:50 pm

Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation (GOP) Method

◆ Adrian E. Raftery, University of Washington

University of Washington, Center for Statistics and the Social Sciences, Box 354320, Seattle, WA 98195-4320

raftery@stat.washington.edu

Key Words: *climatology, deterministic models, prediction, spatial correlation*

Probabilistic weather forecasting consists of finding a joint probability distribution for future weather quantities. It is typically done by using a numerical weather prediction model, perturbing the inputs to the model in various ways, and running the model for each perturbed set of inputs. The result is then viewed as an ensemble of forecasts, taken to be a sample from the predictive distribution. This is typically not feasible for mesoscale weather prediction carried out locally by organizations without the vast data and computing resources of national weather centers. Instead, we propose a simpler method which breaks with previous practice by perturbing the outputs, or deterministic forecasts, from the model. Forecast errors are modeled using a geostatistical model, and ensemble members are generated by simulating realizations of the geostatistical model. In an experiment with 48-hour temperature forecasts in the Pacific Northwest, our forecast intervals turned out to be empirically well calibrated, sharper than those obtained from approximate climatology, and consistent with the spatial correlation structure of the observations.

366 Spatial Boundary Analysis: Wombling and Beyond ▲ ⊞

ENAR, Section on Statistical Graphics

Wednesday, August 11, 2:00 pm–3:50 pm

Bayesian Wombling for Areal Data

◆ Brad Carlin, University of Minnesota; Haolan Lu, University of Minnesota

University of Minnesota, MMC 303, Division of Biostatistics, School of Public Health, Minneapolis, MN 55455

brad@biostat.umn.edu

Key Words: *posterior distribution, MCMC, spatial data*

Boundary analysis is the name given by geographers to the use of spatially oriented data to determine boundaries separating areas of dissimilar values on a map. "Wombling" is the name of an algorithmic approach (due to Womble, 1951) which uses spatial gradients computed from point level data, after which regional boundaries are constructed "from scratch." With areal (say, zip or county) level data, however, boundary analysis methods must use a subset of the existing set of (typically geopolitical) boundaries. We offer a Bayesian solution to this problem, which allows full inference regarding the existence of a boundary separating any two regions. The analogues of "crisp" and "fuzzy" wombling arise naturally by determining the boundaries using either the posterior means of the metrics themselves, or the posterior probabilities that they exceed a given level, respectively. The Bayesian framework also enables easy inference regarding the uncertainty of any boundary or group of boundaries (the latter via multivariate summaries). We illustrate our approach using a dataset on late detection of colorectal and breast cancer in Minnesota counties.

Gaussian Difference Boundaries: Boundary Detection Using Distribution Theory

◆ Geoffrey M. Jacquez, BioMedware, Inc.

BioMedware, Inc., 516 North State, Ann Arbor, MI 48104-1236

jacquez@biomedware.com

Key Words: *boundary analysis, neutral models*

Geographic boundary analysis is used to detect zones of rapid change in spatially distributed variables. A wide variety of methods have been proposed and applied, including "wombling," approaches based on fuzzy set theory, and those suited for univariate as well as multivariate frameworks, and operating on real, integer and categorical data. Many of these approaches are distribution-free, relying on randomization to construct empirical distributions for evaluating the extent of "boundariness" in a given dataset. This presentation considers two salient problems in boundary analysis. First, the need for techniques based on distribution theory that are computationally fast and based on realistic assumptions. And second, the construction of neutral null models suited for the assessment of probabilities under more realistic assumptions than complete spatial randomness.

How to Determine Cohesive Ecological Boundaries

◆ Marie-Josee Fortin, University of Toronto; Ferenc Csillag, University of Toronto; Mathieu Philibert, Institut national de santé publique du Québec

University of Toronto, Dept. of Zoology, 25 Harbord St., Toronto, ON M5S 3G5 Canada

mjfortin@zoo.utoronto.ca

Key Words: *boundary detection, forest, randomization tests, spatial autocorrelation, scaling, sampling*

In ecology, the development of boundary detection methods has a long tradition in the studies associated to ecotone's delineation and forest edge effects. The ecological and statistical question is, however: what is a cohesive boundary? Indeed, while a boundary can be operationally defined as the spatial location where the highest rate of change occurs, the problem is to determine how high is high enough. Boundary membership is usually determined using an arbitrary threshold. In ecological landscapes, however, the strength of boundaries is affected by sampling design issues (e.g., number of sampling locations and their spacing) as well as the presence of spatial autocorrelation within patches. To maximize the detection of among-patch boundaries, it is recommended to carry out boundaries detection at several spatial resolutions. Such scaling procedures allow identification of the degree of boundaries' persistency across scales to be found. This procedure is illustrated with both simulated landscapes data and forestry data.

On Bayesian Wombling: Estimating Spatial Gradients

◆ Alan E. Gelfand, Duke University; Sudipto Banerjee, University of Minnesota

Duke University, Institute of Statistics and Decision Science, Old Chemistry Bldg., Durham, NC 27708-0251

alan@stat.duke.edu

Spatial process models are now widely used for inference in many areas of application. In such contexts interest often lies in estimating the rate of change of a spatial surface at a given location in a given direction. This problem, known as "wombling," after a foundational paper by William Womble, is encountered in several scientific disciplines. Examples include temperature or rainfall gradients in meteorology, pollution gradients for environmental data, and surface roughness assessment for digital elevation models. Since the spatial surface is viewed as a random realization, all such rates of change are random as well. We formalize the notions of directional finite difference processes and directional derivative processes building upon the concept of mean square differentiability and obtain complete distribution theory results for stationary Gaussian process models. We present inference under a Bayesian framework which, in this setting, presents several modeling advantages. Illustrations are provided with simple and complex spatial models.

367 Legal Requirements and Ethical Behavior: The Case of Certifying Statistical Deidentification ▲ ⌘

Cmte on Privacy and Confidentiality, Section on Government Statistics, Social Statistics Section, Cmte on Professional Ethics, Section on Health Policy Statistics

Wednesday, August 11, 2:00 pm–3:50 pm

Legal Requirements and Ethical Behavior: The Case of Certifying Statistical Deidentification

◆ Barbara Bailar, NORC, University of Chicago; ◆ Alvan O. Zarate, U.S. National Center for Health Statistics; ◆ Alan Zaslavsky, Harvard Medical School; ◆ Laura Zayatz, U.S. Census Bureau

U.S. Census Bureau, Statistical Research Division, STOP 91, Washington, DC 20233-9100

Laura.Zayatz@census.gov

Key Words: *ethics, privacy, confidentiality, deidentification, HIPAA, disclosure limitation*

New federal regulations have been implemented to protect the privacy of health information that identifies individuals. These regulations, in particular the HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule, constrain the use and sharing of health data for research. Researchers may comply with these new regulations by using statistical methods to deidentify the records. "A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" must certify that there is "small risk" the information could be used to identify the individual who is the subject of the information. Yet, the standard of risk is undefined, the nature of the potential threats is unknown, and the technical issues in deidentifying data and assessing risk are complex. Under these circumstances, what does it mean for a statistician to certify a "small risk"? We bring together experts in the areas of data deidentification methods, health research, and statistical ethics to discuss the practical, legal, and ethical implication of providing such certification.

368 Statistical Challenges and Issues in the Analysis of Radiographic Data from Rheumatoid Arthritis Clinical Trials ▲

Biopharmaceutical Section

Wednesday, August 11, 2:00 pm–3:50 pm

An Introduction to the Analysis and Interpretation of Radiographic Data from Rheumatoid Arthritis Clinical Trials

◆ James Whitmore, Amgen, Inc.; Guowen (Gordon) Sun, Amgen, Inc.; Wayne Tsuji, Amgen, Inc.

Amgen, Inc., One Amgen Center Dr., MS 24-2-A, Thousand Oaks, CA 91320-1799

whitmore@amgen.com

Key Words: *clinical trials, rheumatoid arthritis, radiographic data*

Rheumatoid arthritis (RA) is an autoimmune disease that often results in progressive loss of function in affected joints. Serial x-ray films reveal the joint erosions and progressive joint space narrowing that characterize RA. Validated radiographic indices have been developed, with progression generally assessed over periods of at least 12 months. Slowing radiographic progression has become an established surrogate marker for overall patient benefit. Advances in RA treatment have led to therapies that may inhibit the progression of structural damage. The analysis and interpretation of the radiographic data from clinical trials of these therapies is complicated by missing data, inter- and intra-reader variability, and the highly skewed distribution of change. Further, placebo treatment arms may be unethical in these studies, leading to active comparators for which historical data on radiographic progression versus untreated patients is scant. This paper will provide an overview of the rationale and methods for characterization of progression and illustrate statistical issues in the analysis of radiographic data from RA clinical trials.

Handling Missing Radiographic Data in Rheumatoid Arthritis Clinical Trials

◆ Kimberly L. DeWoody, Centocor, Inc.

Centocor, Inc., 200 Great Valley Pkwy., Malvern, PA 19355

kdewoody@cntus.jnj.com

Key Words: *missing data, radiographic data, rheumatoid arthritis*

In rheumatoid arthritis clinical trials, radiographs of the hands and feet are taken to assess structural damage. Using a standardized scoring system, independent and blinded readers score each of the joints for erosions and joint space narrowing. Missing radiographic data can occur on a joint-level for reasons such as inadequate image quality or poor patient positioning. Missing data can also occur on a patient-level due to the patient missing a visit for various reasons, including patient withdrawal from the study or being lost to follow-up. Approaches to handling the missing radiographic data will be discussed.

Comparison of Methods for Missing Data Imputations in the Analysis of Radiographic Data from a Large Phase III Trial

◆ Guowen (Gordon) Sun, Amgen, Inc.; Pilita Canete, Amgen, Inc.; James Whitmore, Amgen, Inc.; Gary Aras, Amgen, Inc.; Nick Fotheringham, Amgen, Inc.

Amgen, Inc., 2806 Arbella Lane, Thousand Oaks, CA 91362

gsun@amgen.com

Key Words: *clinical trial, missing data, simulation, estimation, sensitivity analysis*

The treatments for rheumatoid arthritis have advanced dramatically in the last few years. These treatments are indicated not only for reducing signs and symptoms but also for inhibiting the progression of structural damage in patients with active rheumatoid arthritis. The progression of structural damage due to rheumatoid arthritis is generally slow and a validated index/composite score is commonly used to evaluate the progression based on the radiographs taken at different time-points. Since the index/composite score is based on the radiographs that include more than 80 individual joints in the hands, wrists, and feet, handling missing radiographic joint scores is a challenge. In addition, missing radiographic data would also occur due to missed visits, early withdrawal, or lost to follow-up. These types of missing data pose challenges in estimating the true treatment effect of new therapies. In a large phase III clinical trial, analyses of the radiographic data were performed using four missing data imputation methods (linear interpolation or extrapolation, worst-case imputation, multiple-imputation, and last-observation-carried-forward).

Issues Scoring Radiographs and the Use of Multiple Readers

◆ Ronald D. Pedersen, Wyeth Pharmaceuticals

Wyeth Pharmaceuticals, 500 Arcola Rd., Collegeville, PA 19426

pedersr@wyeth.com

Key Words: *intraclass correlation, radiograph*

When radiographs are used in clinical trials multiple readers are often used with the intent of both saving time and improving the precision of estimation. A design objective of clinical trials including radiographic endpoints is to find the right balance of the number of readers and the number of patients to maximize statistical power while minimizing completion time and managing costs. Increasing the number of replicate readings to achieve greater precision increases both costs and time required. Methods to help decide the number of replicates will be reviewed. With multiple readers, inter-reader variability must be controlled, as well as intra-reader variability. Statistical methods to guide training using intraclass and interclass correlation evaluation will be reviewed. Approaches to minimize reader bias through designed distribution of radiographs to readers will be reviewed and statistical models to address reader effect will be discussed.

Regulatory Guidelines and FDA's Perspective on the Collection and Analysis of the Radiographic Data in Rheumatoid Arthritis Clinical Trials

◆ Boguang A. Zhen, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 11413 Duryea Dr., CDER/OPaSS/OB/BTSS, Potomac, MD 20854

zhenb@cber.fda.gov

Key Words: *clinical trial, regulatory guidance, radiographic data, missing data, repeated measurements, statistical method*

Rheumatoid arthritis (RA) is a chronic, symptomatic disease that can result in a variety of outcomes with different chronologies, severities, and overall patient effects. Over the past decade, there has been a search for better measures to describe patient outcomes in RA clinical trials. One of the important outcome variables for RA therapy is the structure damage as measured by x-ray. However, collection, analysis, and presentation of the radiographic data are challenging. One often needs to deal with issues such as repeated measurements over a long period of time, multiple radiographic readers, multiple co-primary endpoints, selection of analysis method, handling of missing radiographic data, presentation of data in labeling, etc. The author will introduce the regulatory guidance for clinical development programs seeking claim of inhibiting progression of structure damage for products used in RA treatment. The author will also present and discuss concerns raised during the review of submissions with radiographic data.

369 Small-area Estimation IV ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Wednesday, August 11, 2:00 pm–3:50 pm

New Developments in SAIPE County Median Household Income Models

◆ Geoffrey M. Gee, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill R., Washington, DC 20233-8500

Geoffrey.M.Gee@census.gov

Key Words: *small-area estimation, hierarchical Bayes models*

The U.S. Census Bureau Small Area Poverty and Income Estimates (SAIPE) program provides select poverty and income estimates for states, counties, and school districts. Recent research on the SAIPE county median household income (MHI) model has focused on two areas. The dependent variable in SAIPE county MHI production models has been three years of Current Population Survey (CPS) data that are centered on the target year. The primary justification for combining data across time is that the CPS single-year sample size for most counties is quite small. Including additional data from neighboring years improves the variance of the estimates; however, it also introduces unmeasured error due to differences in income across time. In response to the State Children's Health Insurance Program (SCHIP), the CPS sample size was significantly increased, opening the possibility that SAIPE county MHI estimates based on a CPS single-year estimate might be feasible. Two alternative models have been developed. The first model is similar in form to the present SAIPE production model with some exceptions.

An Empirical Study on Using ACS Supplementary Survey Data in SAIPE State Poverty Models

◆ Elizabeth T. Huang, U.S. Census Bureau

U.S. Census Bureau, Washington, DC 20233

elizabeth.t.huang@census.gov

Key Words: *small-area estimation, Fay-Herriott model, American Community Survey*

The Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces state poverty estimates from a Fay-Herriot model applied to direct state estimates from the Current Population Survey (CPS). In recent years supplementary surveys for the American Community Survey (ACS) have also produced state poverty estimates. While their different data collection procedures mean that these estimates should have different nonsampling errors than the CPS, they have the advantage of being based on much larger samples. We take data from the 2001 supplementary survey of ACS (SS01) and examine alternative models that use the SS01 data in different ways reflecting different assumptions about the relative biases between the SS01 and CPS estimates. We then examine the prediction error variances from the various models including the SAIPE production models without the SS01 data. The overall conclusions are that there is significant potential for improving the state poverty estimates by using the supplementary survey data in the models, and that the amount of improvement depends on what one assumes about relative biases in the estimates.

Investigating the Use of IRS Tax Data in the SAIPE School District Poverty Estimates

◆ Jerry J. Maples, U.S. Census Bureau

U.S. Census Bureau, SRD Room 3132 Bldg. 4, Washington, DC 20233

jerry.j.maples@census.gov

Key Words: *administrative records, SAIPE, small-area estimation*

The Small Area Income and Poverty Estimates (SAIPE) program provides estimates for selected income and poverty statistics for states, counties, and school districts. School districts are decomposed into pieces to avoid overlapping county boundaries. In the current school district poverty model, within-county shares of poverty are created for each school district piece based on the most recent census long-form data. These shares allocate the county's child poverty to each school district piece and the estimate of the school district is the sum of the poverty in its pieces. Recently, tabulations of the IRS income tax table at the school-district-piece-level have become available. IRS income tax return data, number of child exemptions on returns with income below the poverty line, can provide information about poverty that is more temporally relevant than the decennial data during noncensus years. However, some school district pieces do not have IRS income tax data and this disproportionately affects the smaller sized pieces. We will investigate two types of models which will use both the IRS tax data and decennial data.

School District Population Estimates Methodology Research Using Administrative Data Sources

◆ Monique M. Oosse, U.S. Census Bureau

U.S. Census Bureau, 8751 Cedarcrest Dr., Traverse City, MI 49684

moniqueo1@aol.com

Key Words: *school districts, population estimates, school-age children, public school enrollment, census data, regression models*

This paper describes efforts by the U.S. Census Bureau's Population Division to develop new methods for producing school district population estimates. The current method is not accurate for school districts that experience rates of population change that are substantially different from the corresponding county-level population changes. To improve the accuracy of the school district population estimates, several methods using the National Center for Education Statistics (NCES) public school enrollment data and the 1990 Census were developed. One method treats school enrollment data as direct indicators of school district population and population change. The NCES data are used in regression models as predictor variables for the school district population estimates. A third approach combines information about population changes from the NCES data with preliminary estimates developed from the shares method currently in use. The test estimates are evaluated by school district size, type, location, and other characteristics using Census 2000 data. This paper also describes the geographic coverage for the school-age population in the 1989-1990 and 1999-2000 NCES data.

Small-area Estimation of Health Insurance Coverage from the Current Population Survey's Annual Social and Economic Supplement and the Survey of Income and Program Participation

◆ Joanna M. Turner, U.S. Census Bureau; Robin C. Fisher, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233-8500

joanna.m.turner@census.gov

Key Words: *small-area estimates, health insurance coverage, ASEC*

The U.S. Census Bureau's Small Area Estimates program is investigating the feasibility of producing model-based health insurance coverage statistics. Small-area estimates, such as those for states and counties, are increasingly in demand, and in fact necessary with the creation of the State Children's Health Insurance Program. The Current Population Survey's Annual Social and Economic Supplement is the most widely cited source of data on the uninsured, and the exclusive source for a comprehensive set of state estimates, due to the state-based sampling design. However, no comprehensive set of annual county estimates is available. The Small Area Estimates program currently produces income and poverty estimates for states and counties, where the relationship between administrative records and ASEC direct estimates is modeled to get more precise estimators and estimated standard errors. We extend that work and the work of Fisher and Campbell (2002) and Fisher and Turner (2003) to develop and refine the estimation methods for health insurance coverage and consider the incorporation of new sources of data.

370 Statistical Partnerships between Academia and Industry Government ▲ ⌘

Section on Government Statistics, Section on Survey Research Methods, Section on Statistical Education

Wednesday, August 11, 2:00 pm–3:50 pm

Role of SPAIG in Encouraging and Developing Partnerships

◆ Robert R. Starbuck, Wyeth Research

Wyeth Research, D-5223, PO Box 42528, Philadelphia, PA 19101-2528

StarbuR@wyeth.com

A brief history and the mission of the Statistics Partnerships among Academe, Industry, & Government (SPAIG) ASA initiative will be presented. The accomplishments and current activities of SPAIG to encourage and develop partnerships between academe and business/industry/government organizations will be reviewed.

The NISS Affiliates Program: Furthering Collaboration between Academe and Industry/Government

◆ Alan F. Karr, National Institute of Statistical Sciences

National Institute of Statistical Sciences, PO Box 14006, Research Triangle Park, NC 27709-4006

karr@niss.org

Key Words: *NISS, affiliates program, collaboration*

Together, the NISS Affiliates Program and NISS/SAMSI University Affiliates Program have as members more than 50 corporations, government agencies, national laboratories, and university departments. Catalyzing research collaborations between academe and industry/government is an explicit component of the affiliates programs. Mechanisms used by NISS to stimulate such collaborations will be described, as will a variety of lessons learned. The mechanisms range from postdoctoral appointments to summer internships for graduate students to Affiliates Problem Days. The lessons learned concern which of these programs have been relatively more successful (all have succeeded to some degree) and why.

Impact of the NSF/ASA Research Fellowship Program on the Research Program of Federal Statistical Agencies

◆ Tommy Wright, U.S. Census Bureau; Cynthia Z.F. Clark, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Statistical Research Division, Washington, DC 20233-9100

Tommy.Wright@census.gov

Key Words: *NSF/ASA Fellowship*

The ASA currently administers the ASA/NSF Research Fellowship Program with support from the National Science Foundation (NSF) and three federal statistical agencies/division: NSF's

Division of Science Resources Statistics, the U.S. Bureau of Labor Statistics, and the U.S. Census Bureau. With some background, we highlight experiences of past participants and note some of their impacts on the research programs of the respective government agencies. We also include some experiences and impacts by individuals who participated in the related research programs including: ASA/EIA Research Program; ASA/NCHS Research Fellowship Program; and ASA/USDA-NASS Research Fellow and Associate Program.

The Harvard/Schering-Plough Partnership: A Broad and Beneficial Relationship

◆ Kenneth J. Koury, Schering-Plough Research Institute; L. J. Wei, Harvard School of Public Health

Schering-Plough Research Institute, 2015 Galloping Hill Rd., K-15-2,2435, Kenilworth, NJ 07033

kenneth.koury@spcorp.com

Key Words: *Harvard/Schering-Plough Partnership, Harvard School of Public Health, pharmaceutical industry*

The Harvard/Schering-Plough Partnership provides a mechanism for capitalizing on the complementary nature of the roles of industry and academia in research related to pharmaceutical product development. The need to accurately assess the health effects of new and existing therapies represents a broad area of overlap between the pharmaceutical industry and the Department of Biostatistics at the Harvard School of Public Health. The partnership facilitates the exploration of mutual research interests and the generation of new research projects which affect the practice of quantitative methods in the industry. The centerpiece of the partnership is the annual Harvard/Schering-Plough Workshop. This workshop has emerged as a premier conference for clinical sciences in the pharmaceutical industry, attracting participants from industry, government, and academic institutions to discuss timely and emerging issues in drug development and evaluation. Structured activities, such as visits between the two institutions and the funding of junior faculty to support relevant methodological research, provide a framework that allows Schering-Plough and Harvard to benefit from each other's resources.

371 Income and Wealth Distributions and Human Capital Measurement ▲ ▚

Business and Economics Statistics Section

Wednesday, August 11, 2:00 pm–3:50 pm

The Estimate of Human Capital from Two Sets of Observable Indicators: Formative and Reflective

◆ Vittadini Giorgio, University of Bicocca, Milan; Lovaglio Pietro Giorgio, University of Bicocca, Milan

University of Bicocca, Milan, 1, p.zza ateneo Nuovo, Milan, 20123 Italy

giorgio.vittadini@unimib.it

Key Words: *human capital, latent variables, structural model, path diagram, formative indicators, reflective indicators*

A situation well known in literature is a Latent Variable (LV) linked to its Manifest Variables both in a reflexive block and a formative block, where a block is said to be formative if the manifest variables cause the latent variables, whereas the block is reflexive if the latent variables cause manifest variables. The aim of the present paper is to generalize the unique-reflective indicator approach proposed to the estimate of Human Capital to the case of a set of indicators specifying a measurement model coherent with the economic issues. The importance of the income as reflective indicator for the HC justifies the research of different aspect of this variables (wage, financial, etc.) and different sets of formative indicators (educational variables and working experience variables). In this way the dimension of HC can be extended to a multiple dimension (each with possibly different sets of formative or reflective indicators).

Some Aspects About the Quality of Income Data from Sample Surveys

◆ Rosalia Castellano, Parthenope University of Naples; Claudio Quintano, Parthenope University of Naples; Andrea Regoli, Parthenope University of Naples

Parthenope University of Naples, Dept. of Statistics and Mathematics, Via Medina 40, Napoli, 80133 Italy

lia.castellano@uniparthenope.it

Key Words: *missing data, imputation, validity*

We aim at investigating some dimensions of the quality of income data from two Italian sample surveys, namely the Survey of Household Income and Wealth conducted by the Bank of Italy and the national component of the European Community Household Panel. In addition, the U.S. Panel Study of Income Dynamics is considered, which is a long-standing longitudinal survey. The data quality on sensitive topics such as income depends crucially on the control of nonsampling errors that are caused by nonresponses and response errors. As a consequence the estimates of the resulting income distribution, as well as any measure of the correlation between income and other variables, may suffer from inefficiency and bias. The main dimensions of the quality concept that we deal with are the accuracy and the validity of the information that is produced and disseminated. Special attention is devoted to the editing, weighting, and imputation methods currently adopted for these surveys, through which any erroneous value is corrected and any missing value is compensated for.

Specification and Analysis of Wealth Distribution Models with Applications

◆ Camilo Dagum, University of Ottawa

University of Ottawa, 200 Wilbrod St., Ottawa, ON K1N 6N5 Canada

dagum@stat.unibo.it

Key Words: *wealth distribution properties, Pareto model, lognormal distribution, log-logistic model, Pearson Type V, Dagum wealth distribution model*

A large number of income distribution models have been developed after Pareto's Type I, introduced in 1895. Nevertheless, the important issue of wealth distribution only called the attention of researchers more than 60 years later with the pioneering works by Wold and Whittle, and Sargan, both published in 1957. The former authors proposed the Pareto Type I model and the latter the lognormal distribution but did not empirically validate them. Later, other models were proposed such as, the Pareto Type II by Stiglitz in 1969, the log-logistic by Atkinson in 1975, and the Pearson Type V by Vaugham, also in 1975. In 1990, Dagum developed his Type II model of income and wealth distribution that was validated with real life data from the United States, Canada and Italy. This study discusses the properties and the relative power of the above models to accurately fit observed wealth distributions.

The Effects of Imputation on Income Distribution as Estimated by Means of Dagum Models

◆ Andrea Regoli, Parthenope University of Naples; Rosalia Castellano, Parthenope University of Naples; Claudio Quintano, Parthenope University of Naples

Parthenope University of Naples, Dept. of Statistics and Mathematics, Via Medina 40, Napoli, 80133 Italy

andrea.regoli@uniparthenope.it

Key Words: *missing data, parametric income distribution models, inequality indices*

The problem of missing data is a crucial point for the quality of the statistical information. In the presence of item nonresponses, one can choose to carry out the analysis only on the valid cases or to fill up the nonresponses through imputation methods in order to reconstruct the complete dataset. We have studied the effects of the imputation of missing income values on the main income descriptive statistics and the measures of association with income. Our interest now lies in the evaluation of the consequences that different imputation procedures may have on the whole income distribution, with special attention to the inequality measurements. Starting with the estimation of the Dagum income distribution model, we want to analyze to what extent a given missing income treatment can modify the parameters of the model. Moreover, since each parameter has its own economic interpretation, we try to interpret the meaning of such changes. The study is carried out on the income data from the Survey of Household Income and Wealth by the Bank of Italy. The missing responses are artificially generated through a simulation technique.

372 Sequential and Multistage Methods in Clinical and Biological Sciences ▲ ▨

Biopharmaceutical Section, Biometrics Section

Wednesday, August 11, 2:00 pm–3:50 pm

Adaptive Stagewise Designs in Clinical Trials

◆ Shelemyahu Zacks, Binghamton University

Binghamton University, Dept. of Math and Science, Binghamton, NY 13902-6000

shelly@math.binghamton.edu

Key Words: *adaptive procedures, search for optimal dosage, cancer phase I clinical trials, quantal response analysis, Bayesian procedures, randomized designs*

Adaptive stagewise procedures are those which incorporate in the design of each stage all the information gathered in previous stages. In contrast with sequential procedures, there are no stopping rules. Adaptive stagewise designs are applied in different fields of treatments and data collection. We focus attention on cancer phase I clinical trials (CPICT), in which the toxicity of new drugs is tested. The objective in these trials is to find the Maximum Tolerated Dose (MTD), and give all patients this dose. We discuss parametric and nonparametric models relating the observed toxicity to the MTD. We discuss Bayesian and non-Bayesian procedures and their efficacy. Randomized play the winner procedures are compared with parametric Bayesian procedures.

Efficient Multistage Designs

◆ Robert W. Keener, University of Michigan

University of Michigan, Dept. of Statistics, 458 West Hall, Ann Arbor, MI 48109-1092

keener@umich.edu

Key Words: *two-stage designs, sequential testing*

Two-stage designs will be considered for processes in which mean response depends in an additive fashion on settings for many binary factors. The experimental goal will be to identify active factors or signs of main effects, and procedures will be judged by risk in a Bayesian model. Efficient designs will be derived in a limit in which sample sizes and the number of factors both tend to infinity.

Sequential Testing in the Agricultural Sciences

◆ Linda J. Young, University of Florida

University of Florida, Statistics Dept., College of Medicine, Gainesville, FL 32610-0212

LYoung@biostat.ufl.edu

Key Words: *SPRT, 2-SPRT, ecological models, binomial, negative binomial*

Sequential procedures have been used to help make management decisions concerning control of pests, such as insects or weeds, within agricultural systems. While Wald's Sequential Probability Ratio (SPRT) Test is still used, the 2-SPRT has been suggested as an alternative. Presence/absence sampling and binomial sequential tests have been suggested to reduce the total sampling effort. Recent efforts have focused on using the modeled relationship between the proportion infested and the sample mean to establish sequential testing procedures. The strengths and weaknesses of these various approaches are reviewed. Open questions are discussed.

Sequential Estimation in the Agricultural Sciences

◆ Madhuri S. Mulekar, University of South Alabama; Linda J. Young, University of Florida

University of South Alabama, Dept. of Mathematics & Statistics, 307 Univ. Blvd, ILB 325, Mobile, AL 36688

mmulekar@jaguar1.usouthal.edu

Key Words: *coefficient of variation, ecological model, Taylor's power model, Iwao's patchiness regression, crowding index, pest population density*

Pest populations, such as insects and weeds, reduce yield, but control of these pests can be expensive. An understanding of the population dynamics of each species in the field is needed before optimal use of pesticides can occur. One element in the study of population dynamics is precise estimation of the density of pest and beneficial species. Sequential sampling has proven to be a useful tool in this setting. The biological meaning of term "density" is the average number of insects in a sampling unit, which differs from the traditional statistical use of density as a term for the probability density function. The most commonly used measure of precision in sequential estimation of the mean is the coefficient of variation D of the sample mean. Since ease of field implementation is a primary concern in agricultural sciences, often a less than optimal but easy to obtain estimator, is preferred to an optimal estimator requiring computations in the field. Methods commonly used within agriculture to sequentially estimate the mean with a specified coefficient of variation D will be reviewed, and some new approaches will be considered.

A New Two-stage Sampling Design for Estimating the Latest Flowering Time

◆ Nitis Mukhopadhyay, University of Connecticut

University of Connecticut, Dept. of Statistics, U 4120, CLAS Bldg, 215 Glenbrook Rd., Storrs, CT 06269-4120

mukhop@uconnvm.uconn.edu

Key Words: *largest mean, fixed-width, confidence interval, implementation, data analysis*

A horticulturist was considering the number of days (X) each variety took from planting seeds to reach the stage when the first bud appeared for three local marigold varieties. The data X could be recorded with accuracy of one-half day. The primary interest was to estimate the maximum waiting time between "seeding" and "first budding" among three varieties under consideration. The information that was made available guided us to use unequal pilot sample sizes. So, a new two-stage sampling design had to be developed and implemented. Important exact as well as some large-sample properties of the methodology are highlighted. The present approach is proven theoretically superior to the existing methodology that works only with equal pilot sample sizes. With the help of the data gathered from our designed experiment, the advantages of the new methodology will be indicated. This presentation will include some joint work with professors Mun S. Son and Y.C. Ko.

373 Bayesian Modeling with Process Convolutions

Scientific and Public Affairs Advisory Committee, Section on Bayesian Statistical Science

Wednesday, August 11, 2:00 pm–3:50 pm

A Spatio-temporal Framework for Modeling Ambient Particulate Matter Concentration Levels

◆ Catherine A. Calder, The Ohio State University

The Ohio State University, 1958 Neil Ave., Columbus, OH 43210

calder@stat.ohio-state.edu

Key Words: *dynamic process convolutions, Bayesian, latent variable, multivariate, fine particulate matter, coarse particulate matter*

Elevated levels of particulate matter (PM) in the ambient air have been shown to be associated with certain adverse human health effects. As a result, monitoring networks that track PM levels have been established across the country. Some of the older monitors measure PM less than 10um in diameter (PM10) while the newer monitors track PM less than 2.5um in diameter (PM2.5); it is now believed that this fine component of PM is more likely to be related to the negative health effects associated with PM. We propose a bivariate dynamic process convolution model for PM2.5 and PM10 concentrations. Our aim is to extract information about PM2.5 from PM10 monitor readings using a latent variable approach and to provide better space-time interpolations of PM2.5 concentrations compared to interpolations made using only PM2.5 monitoring information. We illustrate the approach using PM2.5 and PM10 readings taken across the state of Ohio in 2000.

Space-time Analysis of Extreme Values

◆ Gabriel Huerta, University of New Mexico; Bruno Sanso, University of California, Santa Cruz

ghuerta@stat.unm.edu

Key Words: *space-time process, extreme values, GEV distribution, process convolutions, MCMC, ozone levels*

We propose a new approach for modeling extreme values that are measured in time and space. First we assume that the observations follow a Generalized Extreme Value (GEV) distribution for which the location, scale, or shape parameters define the space-time structure. The temporal component is defined through a Dynamic Linear Model (DLM) or state-space representation that allows to estimate the trend or seasonality of the data in time. The spatial element is imposed through the evolution matrix of the DLM where we adopt a process convolution form. We show how to produce temporal and spatial estimates of our model via customized Markov chain Monte Carlo (MCMC) simulation. We illustrate our methodology with extreme values of ozone levels produced daily in the metropolitan area of Mexico City.

Discrete Convolutions for Space-time Data

◆ Bruno Sanso, University of California, Santa Cruz; Alexandra M. Schmidt, Universidade Federal de Rio de Janeiro

University of California, Santa Cruz, School of Engineering, Santa Cruz, CA 95064

bruno@ams.ucsc.edu

Key Words: *spatio-temporal models, Bayesian statistics, process convolutions*

We consider a class of models for spatio-temporal processes based on convolving independent processes with a discrete kernel that is represented by a lower triangular matrix. We consider two separate families: one obtained by convolving spatial Gaussian processes with isotropic correlations and using the kernel to provide temporal dependencies. A second one is based on considering convolutions of AR(p) processes and using the kernel to provide spatial interactions. We find that the proposed families of models provide a rich variety of covariance structures. These include covariance functions that are stationary and separable in space and time as well as time dependent nonseparable and nonisotropic ones.

Nonstationary Spatial Modeling of Environmental Data Using a Process Convolution Approach

◆ Jenise Swall, NOAA/EPA

NOAA/EPA, MD E243-01, RTP, NC 27711

Swall.Jenise@epa.gov

Key Words: *Bayesian statistics, spatial modeling, environmental modeling*

Traditional approaches to modeling spatial processes involve the specification of the covariance structure of the field. Although such methods are straightforward to understand and effective in some situations, there are often problems in incorporating nonstationarity and in manipulating the large covariance matrices that result when dealing with large datasets. Our approach takes a different perspective, modeling a process as a convolution of a Gaussian white noise process and suitable kernels. Depending on the particular parameterization, this approach can allow flexibility in modeling nonstationary processes, while avoiding the task of working directly with the covariance matrix. We discuss some relevant approaches, and present an application involving environmental monitoring. In particular, we focus on such practical issues as computational efficiency and methods for assimilating data from differing sources.

Statistical Inference for a Particle Simulator

◆ Herbert Lee, University of California, Santa Cruz; Bruno Sanso, University of California, Santa Cruz; Weining Zhou, University of California, Santa Cruz; David Higdon, Los Alamos National Laboratory

University of California, Santa Cruz, School of Engineering, 1156 High St., Santa Cruz, CA 95064

herbie@ams.ucsc.edu

Key Words: *inverse problem, proton accelerator, Gaussian process*

A beam of protons is produced by a linear charged particle accelerator, then focused through the use of successive quadrupoles. The initial state of the beam is unknown, in terms of particle position and momentum. Wire scans are used to collect data on the current state of the beam as it passes through and beyond the focusing region, and the goal is to infer the initial state from the wire trace data. This setup is that of a classic inverse problem, in which a computer simulator is used to link an initial state configuration to observable values (wire traces), and then inference is performed for the distribution of the initial state. We model the initial distribution (position and momentum) as two bivariate Gaussian processes, one for each of the x and y directions. The process convolution approach improves the computational efficiency.

374 Don’t Get Bitten by the Sound Bite

General Methodology

Wednesday, August 11, 2:00 pm–3:50 pm

Don’t Get Bitten by the Sound Bite

◆ G. David Williamson, Centers for Disease Control and Prevention; ◆ Arlene S. Ash, Boston University; ◆ Betty K. Skipper, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, ATSDR/NCEH, 1600 Clifton Rd., NE, MS E29, Atlanta, GA 30333

bos1@cdc.gov

Key Words: *interviews, media*

Whether it’s an on-camera broadcast interview or an informal phone interview with a local newspaper, preparation is the key. This course gives participants an opportunity to learn more about working with the news media as an effective tool in communicating information to the public. The course provides attendees tips for handling interviews and an understanding of what news is, what reporters need, and why it is important to craft key messages in preparation for communicating with the media. There will be an introduction to message development techniques, including the use of message maps and message boxes. Attendees will learn how to use these tools as the cornerstone of interviews by staying on point with the key messages, and bridging back to them when necessary. The session will also include video clips of actual media interviews so that participants can observe and critique the use of the techniques discussed.

375 Rethinking the First Statistics Course for Math Majors: From Easy Additions and Changes to a Complete Metamorphosis

Section on Statistical Education

Wednesday, August 11, 2:00 pm–3:50 pm

Rethinking the First Statistics Course for Math Majors: From Easy Additions and Changes to a Complete Metamorphosis

◆ Andre M. Lubecke, Lander University; ◆ Allan Rossman, California Polytechnic State University; ◆ Beth Chance, California Polytechnic State University; ◆ Ann Cannon, Cornell College; ◆ Deborah Nolan, University of California, Berkeley; ◆ Mary Parker, Austin Community College/University of Texas

Austin Community College/University of Texas, 11928 Stonehollow Dr., Austin, TX 78758-3101

mparker@austincc.edu

Key Words: *class activities, mathematical statistics, statistics education, calculus-based course*

Over the last decade, a great deal of time and energy has been expended redesigning the content and pedagogy used in the algebra-based introductory statistics course. During that time, a number of educators have also been thinking about enhancing the experiences of the more mathematically sophisticated student who is introduced to statistics through a course with a calculus prerequisite. Panelists will describe the calculus-based introductory statistics course that was being taught at their college/university, their perception of why change(s) would benefit the student, what they hoped to accomplish through change(s), and what the response to their change(s) has been. Each panelist will also describe one or two enhancements, activities, or changes that they have made, and, in one case, describe an alternative approach to the traditional calculus-based introductory statistics course.

376 Diagnostics and Robustness

Section on Statistical Computing

Wednesday, August 11, 2:00 pm–3:50 pm

Robust Testing of Level Changes in Interrupted Time-series Analysis

◆ Philip H. Ramsey, Queens College of CUNY

Queens College of CUNY, Psychology Dept., Flushing, NY 11367

phramsey@qc1.qc.edu

Key Words: *Mann-Wald, simulation, regression lines*

An ongoing series of studies has been aimed at various problems arising in the detection of intervention effects in time-series data. In the most recent published work (Ramsey & Ramsey, 2003), it has been shown that a composite procedure can provide accurate tests of level changes in interrupted time-series analysis. The comparison of least squares lines has been shown to greatly improve the accuracy of Type I error control in testing level changes in interrupted time-series analysis. The composite procedure combines the Mann-Wald procedure with that of the Kromrey and Foster-Johnson (1996) regression procedure allowing for testing changes in level as well as changes in both trend and level in the presence of trend effects in time series data. Procedures for evaluating changes in level in the presence of autocorrelation and slope changes are suggested and evaluated.

Flexible Fitting and Diagnostics of Finite Mixture Models

◆ Guan Xing, Case Western Reserve University; Ramani S. Pilla, Case Western Reserve University

Case Western Reserve University, Dept. of EPBI, Cleveland, OH 44106

gxx4@po.case.edu

Key Words: *mixture model, EM algorithm*

Mixture models, which can be viewed as clustering techniques, have become one of the most widely used statistical methods for the analysis of heterogeneous data. Mixture models enable flexible fitting of data arising from diverse fields such as astronomy, bioinformatics, genetics, hyperspectral imaging, medical imaging, and minefield detection. We first develop various functions for fitting the finite mixture models using the EM algorithm and the recent alternative EM methods we proposed. Moreover, a function for assessing the mixture fit based on the directional derivatives is proposed. The applicability and the flexibility of these simple and yet powerful functions is illustrated by fitting a finite mixture of normal distributions to the galaxy data. The proposed functions are quite general and allow fitting of mixtures of distributions from exponential families. The problem with starting values for the EM algorithm and an algorithmic termination criterion are also addressed. Model-building through the rotated, hierarchical, and composite EM algorithms and selection of the number of mixture components is illustrated through the galaxy dataset.

Robust Transformation for Linear Regression with Both Continuous and Binary Regressors

◆ Tsung-Chi Cheng, National Chengchi University

National Chengchi University, 64 Chih-Nan Rd., Section 2, Dept. of Statistics, Taipei, 11623 Taiwan

chengt@nccu.edu.tw

Key Words: *Box-Cox transformation, regression diagnostics, robust estimator*

The problem of non-normality may often be simplified by an appropriate transformation, such as the parametric family of power transformations in Box and Cox (1964). However, the evidence for transformations may sometimes depend crucially on one or a few observations. Several authors have pointed out that data transformations are very sensitive to outliers. A robust estimate of the transformation parameter is proposed for the cases when both continuous and binary regressors exist in the linear regression model. The proposed procedure extends the ideas of Hubert and Rousseeuw (1997), in which the robust distance of the continuous regressors is used to be the weight for L1 regression. The purpose of the present article intends to avoid the influence from potential outliers during the process of data transformations. Simulation study and real data analysis show the performance of the resulting approach.

A Robust Approach to Longitudinal Data Analysis Using the T Linear Mixed Model with AR(1) Serial Correlation

◆ Tsung I. Lin, Tunghai University

Tunghai University, Dept. of Statistics, Box 823, Taichung, 40704 Taiwan

tilin@mail.thu.edu.tw

Key Words: *ECME algorithm, Fisher information matrix, longitudinal data, maximum likelihood estimation, random effects, score test*

We discuss a robust extension of linear mixed model based on multivariate t distributions. Since longitudinal data are collected over time and hence tend to be serially correlated, we employ a parsimonious first-order autoregressive [AR(1)] dependence structure for the within-subject errors. For parameter estimation, we present a hybrid ECME/scoring algorithm that combines the stability and rapid convergence with standard errors as a by-product. Moreover, we offer a score test statistic for testing the existence of autocorrelation in the within-subject errors. The techniques for estimating random effects and predicting future responses of a subject are also investigated. Numerical results are illustrated with a real clinical trial data.

A Note on Diagnostic Statistics for Ridge Regression

◆ Anwar M. Hossain, New Mexico Tech; Yao Wang, Roosevelt University

New Mexico Tech, 801 Leroy Place, Socorro, NM 87801

hossain@nmt.edu

Key Words: *multicollinearity, diagnostics, leverage, shrinkage estimators*

Very little work has been done on the effects that collinearity can have on the influence of an observation. We propose some diagnostic statistics when ridge regression is applied in the presence of multicollinearity. Some approximate distributions of diagnostic statistics are propossed for ridge regression.

Robust and Misspecification Resistant Model Selection with Information Complexity and Genetic Algorithms

◆ Yan Liu, University of Tennessee; Hamparsum Bozdogan, University of Tennessee

University of Tennessee, 331 Stokely Management Center, Dept. of Statistics, Knoxville, TN 37996-0532

yliu7@utk.edu

Key Words: *robust regression, misspecification, information complexity, genetic algorithms, subset selection*

We introduce and develop a new unified theory of model selection, which is robust and at the same time misspecification-resistant based on Bozdogan's information-theoretic measure of complexity (ICOMP) criterion. In developing ICOMP in robust regression, we take into account several critical issues, which include the model misspecification, the presence of outliers and the existence of autocorrelation and heteroscedasticity. With ICOMP in robust regression, we use the genetic algorithms (GA) to select the optimal subset of variables. The genetic algorithms enable the rapid computation of model subset selections that would otherwise be impossible in a reasonable amount of time. As a result, it is now feasible to automatically and dynamically develop the best-fitting models with many different combinations of variables.

Testing and Detecting Non-normalities

◆ Toshinari Kamakura, Chuo University

Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 Japan

kamakura@indsys.chuo-u.ac.jp

Key Words: *normality, testing, skewness, kurtosis, changepoint*

Many statistical models are highly dependent on assumption of normal distribution. However, practicioners sometimes use the statistical model based on the normal assumption without any care of goodness of fitting observed data, so it is important to testing the normality. We will investigate the several methods of testing for normality of observations such as skewness measure test, D'Agostino test, Shapiro & Wilk test, and LM test based on simulation studies. The LM test may have good power over Pearson family, but it is still insufficient for detecting non-normality in case of structure change like change point problems. We will propose modified LM test and investigate its property of testing and detecting non-normality against not only Pearson family but also other group of distribution family used in the field of survival analysis.

377 Sampling and Modeling Ecological and Natural Resources ▲

Section on Statistics and the Environment

Wednesday, August 11, 2:00 pm–3:50 pm

Bayesian Methods for Dealing with Response Uncertainty and Missing Observations in Longitudinal Habitat Preference Studies

◆ Penelope S. Pooler, Virginia Polytechnic Institute and State University; Eric P. Smith, Virginia Polytechnic Institute and State University; David R. Smith, USGS Biological Resources Division

Virginia Polytechnic Institute and State University, Dept. of Statistics, Mail Code 0439, Blacksburg, VA 24060

ppooler@vt.edu

Key Words: *uncertainty, Bayesian, longitudinal, habitat, hierarchical*

When sampling to determine species distribution and habitat preferences, many factors such as harsh weather conditions, poor visibility, and animal defense mechanisms contribute to response uncertainty. Additionally, when habitat preference studies involve a longitudinal component due to repeated measures, there is the added complication of missing observations over time. Both response uncertainty within a single sample period, and missing observations over time can lead to model bias if these factors are correlated with the response. We will discuss the model implications for ignoring such possible sources of model bias and outline methods for improving longitudinal habitat models using Bayesian methods.

A Sampling Protocol for Design-unbiased Estimation in Line Intersect Sampling Using Segmented Transects

◆ David L.R. Affleck, Yale University

Yale University, 210 Prospect St., School of Forestry & Environmental Studies, New Haven, CT 06511

david.affleck@yale.edu

Key Words: *line intersect sampling, design-unbiasedness, intersection, segmented transects*

In many applications of line intersect sampling, transects consisting of multiple, connected segments in a prescribed configuration are employed. Although the transect configuration influences the selection probabilities of discrete population elements, estimators for straight-line transects are not, in general, design-unbiased when applied to data from segmented transects. An amended sampling protocol, applicable to populations of arbitrarily shaped particles, is therefore proposed. The protocol obviates the arbitrary practice of treating multiple intersections of a single element as independent probabilistic events and preserves the design-unbiasedness of Kaiser's conditional and unconditional estimators, suitably generalized to multisegmented transect designs. Although design-unbiasedness can thus be ensured, the efficiency and utility of segmented transect designs from a fixed population perspective remains an open question.

Light Availability and Juvenile Tree Distributions in a Temperate Rainforest: Why Do Different Species Coexist on the Same Resources?

◆ Glenn Hofmann, Household International

glennhofmann@yahoo.com

Key Words: *bounded domain, bounded null models, mid-domain effect, competition, gap-partitioning, rainforest*

Ecologists have proposed that tree species may coexist by specializing on light environments associated with forest gaps of different sizes. Remarkably few studies, however, have actually examined juvenile tree distributions along light availability gradients. We describe distributions of juvenile trees in relation to canopy openness in a temperate rainforest, and test the hypothesis that competitive sorting causes coexisting species to overlap less in light environment occupancy than would be expected by chance. Average overlap of species' interquartile ranges on the canopy openness gradient was tested against a bounded domain null model of community structure which used range-size criteria to constrain random placement of species optima within the domain. The null model assumes that proximity to domain boundaries depresses species density through its influence on range size. This mathematical phenomenon, known as the mid-domain effect in studies of geographic gradients, has not previously been applied to resource gradients. It concurs surprisingly well with our data.

Accounting for Spatial Autocorrelation in Loblolly Pine Stands Using a Simultaneous Autoregressive Model

◆ Bronson Bullock, North Carolina State University

North Carolina State University, 3102 Jordan Hall, Box 8008, Dept. of Forestry, Raleigh, NC 27695-8008

Bronson_Bullock@ncsu.edu

Key Words: *simultaneous autoregressive, SAR, autocorrelation, Pinus taeda, competition, forestry*

A simultaneous autoregressive model will be employed to account for the spatial autocorrelation in loblolly pine stands. Measurements of annual diameter growth will be used as the vehicle through which tree interdependence is expressed. The ability to explicitly characterize the spatial relationships between loblolly pine trees will allow for more accurate representation of stand conditions (current and projected). This research will expound on current knowledge of spatial trends in loblolly pine forest stands by determining the spatial structure of stands using incremental diameter growth measurements.

Spatially Balanced Designs for Experiments on Autocorrelated Fields

◆ Harold M. van Es, Cornell University; Cindy L. van Es, Cornell University; Carla Gomes, Cornell University; Meinolf Sellmann, Cornell University

Cornell University, 1005 Bradfield, Crop and Soil Sciences, Ithaca, NY 14853-1901

hmv1@cornell.edu

Key Words: *design, autocorrelation, agronomy, spatial*

Despite the availability of a multitude of advanced design, most agronomic field experiments (at least 90%) are implemented through randomized complete block designs (RCBD). Use of blocks is generally justified by spatial autocorrelation of soil fertility patterns due to regional variations in drainage, erosion, etc. We demonstrated in earlier work that randomized treatment allocation to plots can cause biases and imprecision under most field conditions due to autocorrelation and trends. We propose spatially balanced experimental designs that are inherently robust to nonrandom field variability (autocorrelated or trended), and use standard designs that are optimized in terms of average spatial separation of treatments. We are using two approaches to develop spatially balanced CBDs: a heuristic local search approach using simulated annealing, and an exact method using combinatorics (for Latin Squares only at this time). We discovered that local search methods provided satisfactory results for smaller designs only. Results with a combinatoric approach are promising and we generated totally spatially balanced Latin Squares up to order 18.

378 Semiparametric and Nonparametric Models ⌘

General Methodology

Wednesday, August 11, 2:00 pm–3:50 pm

Functional Response Models and Their Applications

◆ Xin M. Tu, University of Rochester

University of Rochester, Dept. of Biostatistics and Computational Bio., 601 Elmwood Ave., Rochester, NY 14642

xin_tu@urmc.rochester.edu

Key Words: *distribution-free, functional response, generalized estimating equation, regression analysis, second-order moments, U-statistics*

I will discuss a new class of semiparametric (distribution-free) regression models with functional responses. This class of functional response models (FRM) generalizes the traditional regression models by defining the response variable as a function of several responses from multiple subjects. By using such multiple-subjects-based responses, the FRM addresses a fundamental flaw in existing regression models that limits their applications to modeling the mean or the first-order moment of a response variable and makes it possible to model complex higher-order moments as in many popular non- and semiparametric approaches. For example, under FRM, we can derive regression models to perform inference for nonparametric inference, such as two-way contingency analysis and the Mann-Whitney-Wilcoxon (MWW) rank-based tests. For semiparametric regression analysis, the FRM can be used to model complex variance components such as intraclass correlation in mixed-effects models that are extremely difficult to accomplish using existing semiparametric models. For inference of FRM, we discuss a novel approach by integrating the U-statistic theory with the generalized estimating equations.

Semiparametric Models for Missing Covariates in Generalized Linear Mixed Models

◆ Qingxia Chen, University of North Carolina, Chapel Hill; Joseph G. Ibrahim, University of North Carolina, Chapel Hill

University of North Carolina, Chapel Hill, Dept. of Biostatistics, School of Public Health, Chapel Hill, NC 27599

qchen@bios.unc.edu

Key Words: *generalized additive model, EM algorithm, Gibbs sampling, Monte Carlo EM*

We consider a class of semiparametric models for the covariate distribution and missing data mechanism for missing covariate and/or response data for generalized linear models and generalized linear mixed models. Ignorable and nonignorable missing covariate and/or response data is considered. A generalized additive model (GAM) is considered for the covariate distribution and/or the missing data mechanism. Penalized regression splines are used to express the GAM's as a generalized linear mixed effects model, in which the variance of the corresponding random effects provides an intuitive index for choosing between the semiparametric and parametric model. Maximum likelihood estimates are obtained via the EM algorithm. Simulations are given to demonstrate the methodology and two real datasets from a melanoma cancer and breast cancer clinical trials are analyzed using the proposed methods.

A Completely Nonparametric Approach to the Analysis of Longitudinal Data via a Set of Level-crossing Problems with Application to the Analysis of Longitudinal Microarray Experiments

◆ Cavan Reilly, University of Minnesota

University of Minnesota, Division of Biostatistics, A460 Mayo Bldg., MMC 303, 420 Delaware St. SE, Minneapolis, MN 55455

cavanr@biostat.umn.edu

Key Words: *level-crossing problems, longitudinal analysis, microarrays, nonparametric tests, survival analysis*

We develop a completely nonparametric method for comparing two groups on a set of longitudinal measurements. No assumptions are made about the form of the mean response function, the covariance structure, or the distributional form of disturbances around the mean response function. The idea of the method is quite simple: fix a set of levels and, for each subject, determine the first time the subject has an upcrossing and a downcrossing of this level. For each level one then computes the log rank statistic and uses the maximum in absolute value of all these statistics as the test statistic. By permuting group labels we obtain a permutation test of the hypothesis that the joint distribution of the measurements over time doesn't depend on group membership. Simulations are performed to investigate the power and it is applied to the area that motivated the method-the analysis of microarrays. In this area small sample sizes, few time points, and far too many genes to consider genuine gene level longitudinal modeling have created a need for a simple, model-free test to screen for interesting features in the data.

Penalized Spline Estimation for Single-Index Coefficient Models

◆ Zhou Wu, University of Cincinnati; Yan Yu, University of Cincinnati

University of Cincinnati, 534 Carl H. Lindner Hall, PO Box 210130, Cincinnati, OH 45221-0130

wuzu@email.uc.edu

Key Words: *dimension reduction, nonparametric estimation, nonlinear time series, simulation, varying-coefficient*

The single-index coefficient model, where the coefficients are functions of an index of a covariate vector, is a new powerful tool to model nonlinearity for time series. By reducing the covariate vector to an index, the single index coefficient model overcomes the "curse of dimensionality" that arises in nonparametric estimation. We propose a penalized spline approach to estimate coefficient functions of the single-index coefficient models. Each coefficient function is approximated by a spline with penalty terms to balance the goodness-of-fit of the model and the smoothness of coefficient functions. Implementation details are discussed. We show that the penalized least squares estimators are consistent and asymptotically normal under mild conditions. Moreover, our method has certain advantages over other methods. Different coefficient functions are allowed to have different smoothness with the aid of different penalty terms. The proposed p-spline approach is a global smoothing method which yields a parsimonious model and thus multistep forecasting based on the fitted model is straightforward. An application to GNP and simulated examples are presented to illustrate the approach.

Rao-type Tests for B-spline Estimators of Conditional Quantiles with Varying Coefficients

◆ Mi-Ok Kim, University of Kentucky

University of Kentucky, 817 Patterson Office Tower, Lexington, KY 40506-0027

miokkim@ms.uky.edu

Key Words: *quantile regression, lack-of-fit test for regression quantiles, B-splines*

When a nonparametric regression model is in use, model validation is warranted by the principle of parsimony. In case of varying-coefficient models, whether the coefficients indeed vary needs to be checked. While there exist many studies on the estimation of the model, just a few discuss model validation. For the conditional mean regression, the confidence interval method and a likelihood-ratio-type test has been suggested. We propose a Rao-type test for validating the varying-coefficient model in estimating conditional quantiles with B-spline estimators. The proposed test is consistent against all alternatives with the number of knots growing, while when the number of knots is fixed, the proposed test is inconsistent against a certain form of alternatives. We identify the set of the alternatives against which the test is inconsistent and observe the limiting distribution of the proposed test statistic. A simulation study is conducted to assess the the performance of the proposed test.

On Nonparametric Density Estimation Near the Boundary

◆ R J. Karunamuni, University of Alberta

University of Alberta, Dept. of Math. & Stat. Sciences, Edmonton, AB T6G 2G1 Canada

R.J.Karunamuni@ualberta.ca

Key Words: *density estimation, mean squared error, kernel estimation, reflection, transformation*

The kernel method of estimation of curves is now popular and widely used in statistical applications. Kernel estimators suffer from boundary effects however, when the support of the function to be estimated has finite endpoints. Boundary effects seriously affect the overall performance of the estimator. A number of solutions (adjustments and modifications) have been proposed with some success. We develop a new general method of boundary correction for kernel density estimation. Our technique is a kind of generalized reflection involving transformed data. The proposed method generates a class of boundary corrected estimators. They all possess desirable properties such as local adaptivity and the non-negativity. The theory behind the new methodology along with the bias and variance of the proposed estimators are presented. Results of a data analysis are also given.

Consistency and Generalization Error Bound of Neural Network Models

◆ Indrias G. Berhane, University of Kentucky

University of Kentucky, 4070 Victoria Way Apt. 74, Lexington, KY 40515

iberhane@ms.uky.edu

Key Words: *generalization error, empirical mean square error, uniform stability, smoothing regularization*

Estimation using flexible models based on finite and noisy datasets is a challenging task. The main concern in using flexible models is the possibility of overfitting the training data. Feed-forward neural networks are class of approximating functions with a universal approximation capabilities. For a class of functions with known approximation capabilities the main concern is performance on future datasets. In real-world applications the availability of training samples are often limited and noisy. Since it is possible to find a neural network model which minimizes the empirical error but may not minimize the generalization error we restrict our focus to stable models only. We use a smoothing regularization as a method of stabilizer. We provide an upper bound of the generalization error of a neural network, estimated with a smoothing regularization. We also prove possible generalization error of a neural network estimate converges to the minimum achievable by the class as the training sample increases, with high probability. Our numerical experiments show neural network models estimated with smoothing regularization give similar or less prediction errors.

379 Statistical Inference

IMS

Wednesday, August 11, 2:00 pm–3:50 pm

Bilateral Symmetry in Landmark-based Shape Analysis

◆ John T. Kent, University of Leeds

University of Leeds, Woodhouse Lane, Leeds, LS2 9JT UK

j.t.kent@leeds.ac.uk

Key Words: *shape, landmark, symmetry, bilateral*

Geometric objects possessing bilateral symmetry have the property that their shape is unchanged after a reflection about a line (in two dimensions) or a plane (in three dimensions). Biology provides many examples; indeed most animals possess approximate bilateral symmetry for many of their features. In order to study bilateral symmetry for landmark data, the first task is to construct a suitable coordinate system, a nontrivial exercise due to the lack of a fixed registration system. With this coordinate system, it is possible to quantify any asymmetry and to construct statistical tests for symmetry. The human face provides a nice setting in which to illustrate these ideas.

Some Recent Advances in Multivariate Analysis

◆ Samuel S. Wu, University of Florida

University of Florida, 1600 SW Archer Rd. Room C2-22, PO Box 100212, Gainesville, FL 32610

samwu@biostat.ufl.edu

Key Words: *Fisher's method of combining tests, Bahadur efficiency, average power, multivariate linear hypotheses, multivariate location problem*

We discuss some new findings on combining univariate tests for multivariate problems. It is well known that in a general multiparameter setting, there may not exist any unique best test. More importantly, unlike the univariate case, the power of different test procedures could vary remarkably in multivariate analysis. A Monte Carlo study indicate that the small-sample power of the combination methods is much better than the Hotelling-type tests in many cases. Relationship is established between Fisher's method of combining tests and a new class of tests that have best average power for multivariate linear hypotheses. And asymptotic efficiencies of three combination methods, when both n and p may go to infinity, help to explain when and why they perform better than Hotelling's tests. Applications in microarray analysis, QTL detection, and composite analysis of five aphasia trials are also considered.

Empirical Likelihood Estimation for a Common Mean with Heteroscedastic Variances

◆ Changbao Wu, University of Waterloo

University of Waterloo, Dept. of Statistics and Actuarial Science, Waterloo, ON N2L 3G1 Canada

cbwu@uwaterloo.ca

Key Words: *empirical likelihood, heteroscedasticity, confidence intervals, weighted EL approach*

Estimation of a common mean using information from k independent samples with unequal variances is a classical statistical problem, with applications found in many areas of statistics, such as designed experiments and measurement error models. We show that the recently proposed empirical likelihood method can be a very attractive approach to handling the related inference problems. The maximum empirical likelihood estimator and the empirical likelihood ratio confidence interval perform well under relatively large samples. Two alternative approaches, namely, the weighted empirical likelihood method and the augmented empirical likelihood method, are proposed and are shown to have much improved performance for small samples.

A Note on Multiple Tests for Gene Expression Data

◆ Yuhong Yang, Iowa State University

Iowa State University, Dept. of Statistics, 315B Snedecor Hall, Ames, IA 50011-1210

yyang@iastate.edu

Key Words: *gene expression data, multiple testing, weak control, strong control, subset pivotality*

The issue of multiple testing is commonly encounterted in the work of identifying differentially expressed genes in the analysis of gene expression data. We carefully examine some of the key concepts in the multiple-testing literature such as weak control, strong control, and subset pivotality. We also study the applicability of the permutation-based Westfall and Young procedure for controling the familywise error rate. Our goal is to clarify some misunderstandings about multiple testing in the literature.

Bootstrap Fictitious Play in Repeated Games

◆ Dennis Gilliland, Michigan State University

Michigan State University, Dept. of Statistics and Probability, IMS, East Lansing, MI 48824

gillilan@msu.edu

Key Words: *repeated play, fictitious play, online learning, Hannan consistency*

In repeated games, a player may use opponent's past moves in selecting a move for the present game. With fictitious play, the player plays Bayes against the empirical distribution Gn of those past moves. With bootstrap fictitious play, the player plays Bayes against the empirical distribution of an iid sample from those moves. Let regret denote the player's excess average loss over R(Gn), the component game Bayes envelope risk evaluated at Gn. For a class of component games, we show that the bootstrap fictitious play strategy has regret converging to zero uniformly in opponent's move sequences. The rate of this Hannan consistency is established.

Modifications of Stein's Effect in Shape-restricted Inference

◆ Anna Amirdjanova, University of Michigan; Michael Woodroofe, University of Michigan

University of Michigan, 439 West Hall, 550 E. University, Dept. of Statistics, Ann Arbor, MI 48109-1092

anutka@umich.edu

Key Words: *Stein's identity, inadmissible estimate, shape-restricted inference*

We construct a wide class of shrinkage estimators for a multivariate normal mean when the latter is known to belong to a convex polyhedral cone. It is shown that shrinking the maximum likelihood estimator towards suggested modifications reduces the mean squared error. As an illustration, the case of isotonic regression is considered and corresponding class of target estimators is investigated.

380 Pharmaceutical Theory

Biopharmaceutical Section

Wednesday, August 11, 2:00 pm–3:50 pm

Simulation Studies for Additive-multiplicative Hazard Model and Mixed Continuous and Discrete Cox Regression Model via Empirical Likelihood

◆ Zhiheng Xu, Georgia State University; Yichuan Zhao, Georgia State University

Georgia State University, 1437 Willow Lake Dr., Apt. C, Atlanta, GA 30329

zxu2@gsu.edu

Key Words: *Cox's proportional hazard model, empirical likelihood, normal approximation, counting process*

In contrast to the Cox's proportional hazard model, the additive-multiplicative hazard model and mixed discrete and continuous Cox regression model become more plausible for many applications in biomedical settings. The additive-multiplicative hazard model includes both Cox's proportional hazards model and additive risk model as special cases, and mixed discrete and continuous Cox regression model incorporates the continuous and discrete components, instead of the multiplicative form and absolute continuous failure time in the Cox's proportional hazard model. We apply the empirical likelihood method to these two models to analyze the regression parameter. Two simulation studies are carried out to compare the empirical likelihood ratio method with the normal approximation method.

Cox Regression Methods for Two-stage Randomization Designs

◆ Yuliya Lokhnygina, North Carolina State University; Jeffrey Helterbrand, Genentech, Inc.

North Carolina State University, Dept. of Statistics, Campus Box 8203, Raleigh, NC 27695

ylokhny@ncsu.edu

Key Words: *two-stage randomization designs, proportional hazards, inverse probability weighting, clinical trials*

Two-stage randomization designs are becoming common in oncology and AIDS clinical trials. In these designs patients are initially randomized to an induction treatment, followed by randomization to a maintenance treatment conditional on their remission and consent to further participation in the trial. The goal of the study is usually the comparison of different combinations of induction and maintenance therapies to find the most beneficial treatment policy. Recently, Lunceford, Davidian, and Tsiatis (2002) and Wahed and Tsiatis (2003) developed some nonparametric approaches to the analysis of such designs. Our analysis methods are based on the Cox proportional hazards model. We consider a special situation where at the second stage patients can be randomized to either observation or active drug, and the primary objective is to compare the induction therapies with respect to a time-to-event endpoint. We propose reweighted versions of the usual score estimating equation and the score test in the Cox model. Large sample properties are derived and illustrated via a simulation study. Application of the proposed methods is demonstrated on the example of E4494 clinical trial.

Bayesian Analysis with Mixed Model in Unbalanced Case

◆ Ruoyong Yang, Pfizer Inc.

Pfizer Inc., 235 E 42nd St., New York, NY 10017

ruoyong.yang@pfizer.com

Key Words: *mixed model, repeated measures, random effect, Jeffreys prior, reference prior*

Mixed model is investigated from a Bayesian perspective with non-informative prior approach. Jeffreys prior and reference priors for this model are derived. A clinical trial data is analyzed by this approach. Comparison with standard classical approach shows the similarity between Bayesian and classical approach.

The Impact of Rater Reliability on Statistical Inferences in Ordinal Ratings

◆ Paul Zhang, Pfizer Inc.

Pfizer Inc., 201 Tabor Rd., Morris Plains, NJ 07950

paul.zhang@pfizer.com

Key Words: *rater reliability, statistical power*

Rater reliability of ordinal ratings cause concerns as to the validity and efficiency of statistical inferences on the mean ratings. We discuss the approaches to assess the rater reliability and its impact on the statistical inferences on the group mean ratings. The criterion for reducing these impacts in clinical trials using ordinal ratings are proposed.

Statistical Issues in Toxicology Cardiovascular Safety Pharmacology Studies

◆ Alan Y. Chiang, Eli Lilly and Company

Eli Lilly and Company, Greenfield Research Laboratories, DC #GL43, PO Box 708, Greenfield, IN 46140

chiangay@lilly.com

Key Words: *toxicology, safety pharmacology, design, power evaluation*

The goal of safety pharmacology studies is to investigate the potential undesirable pharmacodynamic effects of a substance on physiological functions in relation to exposure in the therapeutic range and above. The cardiovascular system is one of the vital organ systems acutely critical for life and, therefore, most important. General recommendations for testing strategies in nonclinical studies include: in-vivo evaluation of blood pressure, heart rate, and the electrocardiogram, and in-vitro analysis of ionic current assays and repolarization assays. This talk focuses on some of the key statistical issues, including design, power evaluation, and data analysis, in toxicology cardiovascular safety pharmacology studies. Statistical considerations arisen from ICH S7A and S7B guidance documents for nonclinical studies will also be discussed.

381 Bayesian Biomedical Applications ▲

Section on Bayesian Statistical Science

Wednesday, August 11, 2:00 pm–3:50 pm

Hierarchical Bayesian Analysis of Competing Risk Frailty Models for Duration Data with Correlated Failure Events

◆ Getachew A. Dagne, University of South Florida; James Snyder, Wichita State University

University of South Florida, Dept. of Epidemiology and Biostatistics, 13201 Bruce B. Downs Blvd., MDC 56, Tampa, FL 33612

gdagne@hsc.usf.edu

Key Words: *behavioral observation, random effects, survival analysis, Bayesian method, competing risks*

Competing risk duration models are used to analyze the time patterns of interactions of parents and children, as well as the effects of covariates on transition rates from one state to another. Such duration data often contain repeated events in addition to competing risks. We propose statistical models that analyze the events simultaneously taking dependencies among the separate competing events into account. We specify a flexible parametric proportional competing risks model which permits unrestricted correlations among the risks using Bayesian approach. The methods will be applied to event history of child emotion display in parent-child interactions.

An Application of Markov Chain Monte Carlo Algorithm to Fit a Bayesian Frailty Model

◆ Adriana Lopez, San Diego State University; Richard A. Levine, San Diego State University; Juanjuan Fan, San Diego State University

San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-0001

lopez20@rohan.sdsu.edu

Key Words: *Bayesian analysis, Gibbs sampler, Metropolis-Hastings sampler, gamma frailty, ophthalmology application*

Using a frailty model for correlated observations due to a common environment has become an important research topic. We propose a Markov chain Monte Carlo algorithm to fit a Bayesian frailty model having a prior gamma distribution for the frailty parameter. The algorithm is a Metropolis-Hastings sampler, incorporating a Gibbs sampling routine, with priors on the gamma parameters and frailty parameters and an empirical Bayes estimator for the hazard function. The dataset we analyzed considered treatment effect on glaucoma onset. The dataset has two failure times for each patient, the right and the left eye. Each pair of failure times is naturally correlated. The frailty parameter takes account of variables that had not been measured, but could affect the onset of primary open angle glaucoma.

Bayesian Analysis for Assessing Diagnostic Test Accuracy and Optimal Design in a Cardiology Clinical Trial

◆ Manuela Buzoianu, Carnegie Mellon University; Joseph B. Kadane, Carnegie Mellon University

Carnegie Mellon University, Dept. of Statistics, Baker Hall 132, 5000 Forbes Ave., Pittsburgh, PA 15213

manuela@stat.cmu.edu

Key Words: *verification bias, missing data, MCMC algorithms, utility function, optimal design, stepwise sampling*

Evaluating new diagnostic tests is necessary when the gold standard, which provides definitive verification of a disease, has a risk of mortality. This is the motivation for the current cardiology trial, in which a noninvasive test for coronary artery disease diagnosis has to be evaluated against the gold standard coronary angiography. For this, we studied the test sensitivity and specificity. When true disease status is missing for many patients, the estimates of the test parameters are likely to be biased. We performed a Bayesian analysis to correct for verification bias, which is the bias resulting from this missing data mechanism, and compare it with a likelihood-based approach. An important implication of this analysis is to make decisions about future patients so that the test properties estimates can be improved. Thus, the paper also discusses designing experiments, in which new patients are carefully selected for disease verification. Several experiments are possible and the design choice is regarded as a decision problem, which is based on experiment utility maximization. Simulation-based methods and stepwise sampling algorithms are developed to obtain the optimal design.

Applications of Predictive Probabilities in Phase II Cancer Clinical Trials

◆ Diane D. Liu, University of Texas M.D. Anderson Cancer Center; J. Jack Lee, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030

dianeliu@mdanderson.org

Key Words: *Bayesian methods, flexible monitoring scheme, early termination, efficiency, robustness*

A Bayesian predictive probability approach is proposed as an alternative to the commonly used two- or three-stage designs in phase II cancer clinical trials. The new design allows flexible and more frequent monitoring of the trial outcomes. Early termination of the trial can be reached when the interim data indicate that the experimental regimen is not promising. Even with more frequent monitoring, the predictive probability approach can still possess good frequentist's properties by controlling Type I and Type II errors. A bi-dimensional search algorithm is implemented to determine the design parameters. Exact computation and simulation studies demonstrate that the predictive probability approach is more efficient than the traditional multistage designs. The predictive probability design not only is more adaptable in evaluating the study outcome but also remains robust even when deviations occur in monitoring cohorts of patients as specified in the trial design. Examples will be given to illustrate the statistical properties of various design settings.

A Variance Components Analysis of a Multisite fMRI Study

◆ Hal Stern, University of California, Irvine; David A. van Dyk, University of California, Irvine; Padhraic Smyth, University of California, Irvine; Se-Young Kim, University of California, Irvine; First Birn, Bioinformatics Resource Network

University of California, Irvine, Dept. of Statistics, 4800 Berkeley Place, Irvine, CA 92697-1250

sternh@uci.edu

Key Words: *reproducibility, region of interest*

A developing approach to neuroscience research attempts to combine and collate imaging data from a number of sites. This requires methods for standardizing data collection and calibrating the resulting images to account for site-to-site differences including differences in scanner models. As a first step five healthy subjects visited 11 sites and were scanned using the same protocol. Each subject made two visits to each site; each visit comprised a series of 10 functional imaging runs. This preliminary analysis estimates the contribution to observed variance from run-to-run variability within a visit, visit-to-visit variability within a site, and variation across sites and subjects. Analyses are carried out on brain regions of varying size down to the level of a single voxel.

Non-informative Priors for the Interval Estimation of Attributable Risk

◆ Bin Huang, Children's Hospital Medical Center; Siva Sivaganesan, University of Cincinnati; Gail Slap, Cincinnati Children's Hospital Medical Center

Children's Hospital Medical Center, 3333 Burnet Ave., MLC 4000, Cincinnati, OH 45229-3039

bin.huang@cchmc.org

Key Words: *attributable risk, Jeffery's prior, entropy prior, interval estimation, simulation study*

Attributable risk (AR) is one the most important epidemiological measure of the impact of a risk factor on an outcome. While there is a vast amount of literature on the interval estimation of the AR, very little has been done in the Bayesian approach to the interval estimate of AR. Our earlier study suggested a need for "better" noninformative priors for Bayesian interval estimation of AR. We derived an entropy prior, and assessed its performance and compared it with Jeffery's prior and other existing approaches. The study suggested that for the interval estimation of AR in general problem settings, performance of the entropy prior with regard to the frequentist coverage rate and length is better than that of Jeffery's, while these two priors are the same in the simple 2 by 2 contingency table setting.

382 Odds, Ratio, Risk, Matching, and Case Control Studies ▲

Section on Statistics in Epidemiology

Wednesday, August 11, 2:00 pm–3:50 pm

Case Control Studies after RDD: New Sample Designs and Analysis Strategies

◆ Ralph DiGaetano, Westat; Barry I. Graubard, National Cancer Institute; Sowmya R. Rao, National Cancer Institute; Jacqueline Severynse, Westat; Sholom Wacholder, National Institutes of Health

Westat, 1650 Research Blvd., Rockville, MD 20850

digaetr1@westat.com

Key Words: *case-control, sample frame, oversampling, design effects*

Sample selection of population-based controls for case-control studies is often done using an RDD screener to obtain stratification variables such as age, race, and sex, and then subsampling among study eligible persons within strata. The sample distribution of controls is commonly matched to the expected sample case distribution on these variables. With RDD response rates dropping over the past decade, alternative sample designs are being explored. Department of Motor Vehicle lists could serve as sample frames for controls, but coverage issues and the general absence of race data have posed concerns. Using geocoding to link addresses to Census areas permits the oversampling of high-density black areas, helping to increase black sample sizes efficiently. This paper examines available results from a study using such an approach, increasing targeted sample sizes to compensate for design effects so that precision and power are not adversely affected. Coverage of the general population seems fine, but issues related to coverage of subgroups by DMV lists as well as logistical concerns remain. Results on the sample distribution by race and "area density" strata are provided.

Matched Samples Logistic Regression in Case Control Studies with Missing Values: When to Break the Matches

◆ Harry J. Khamis, Wright State University; Lisbeth Hansson, Uppsala University

Wright State University, 3640 Colonel Glenn Hwy., Dayton, OH 45435

harry.khamis@wright.edu

Key Words: *conditional, unconditional, maximum likelihood, method bias, root mean square error*

Simulated data sets are used to evaluate conditional and unconditional maximum likelihood estimation in an individual case-control design with continuous covariates when there are different rates of excluded cases and different levels of other design parameters. The effectiveness of the estimation procedures are measured by method bias, variance of the estimators, root mean square error (RMSE), and a goodness-of-fit measure. Conditional estimation leads to higher RMSE than unconditional estimation in the presence of missing observations, especially for 1:1 matching. The RMSE is higher for the smaller stratum size, especially for the 1:1 matching. The fit of the model appears to be insensitive to missing data, but is generally higher for the conditional estimation than for the unconditional estimation. The fit is particularly good for the 1:2 matching design. For minimizing RMSE, a high matching ratio is recommended; in this case, conditional and unconditional logistic regression models yield comparable levels of effectiveness. For maximizing fit, the 1:2 matching design with the conditional logistic regression model is recommmended.

Interval Estimation of the Difference in Proportions under M-to-one Matching

◆ Kung-Jong Lui, San Diego State University

San Diego State University, 5500 Campanile Dr., Dept. of Mathematics and Statistics, San Diego, CA 92182-7720

kjl@rohan.sdsu.edu

Key Words: *simple difference, multiple matching, interval estimation, efficiency, coverage probability*

We focus interval estimation on the difference in proportions for m-to-one matching. We consider four asymptotic interval estimators, including the estimator directly using the Mantel-Haenszel (MH) point estimator, the estimator using the transformation to improve the normal approximation, the estimator derived from the Cochran-Mantel-Haenszel (CMH) statistic for stratified randomized trials, and the estimator derived from a quadratic equation. To evaluate and compare the performance of these estimators, we employ Monte Carlo simulation. We find that the estimator directly using the MH estimator can have the coverage probability less than the desired confidence level when the number of matched sets is small. We note that the estimator derived from the quadratic equation can actually perform well when the underlying difference is close to 0 even for a small number of matched sets. However, this estimator tends to have the coverage probability less than the desired confidence level as well when the underlying difference in porportons is large. We note that the estimator using the transformation consistently outperforms the interval estimator directly using the MH estimators.

Odds Ratio Estimates for Case-control Studies Subject to Informatively Missing and Misclassified Exposure

◆ Robert H. Lyles, Emory University; Andrew S. Allen, Duke University; W.D. Dana Flanders, Emory University; Lawrence L. Kupper, University of North Carolina, Chapel Hill; Deborah L. Christensen, Emory University

Emory University, Dept. of Biostatistics, 1518 Clifton Rd., NE, Atlanta, GA 30322

rlyles@sph.emory.edu

Key Words: *bias, maximum likelihood, reassessment, validation*

In case-control studies, it is common for a categorical exposure variable to be misclassified. It is also common for exposure status to be informatively missing for some individuals, in that the probability of missingness may be related to exposure. Procedures for addressing the bias due to misclassification via validation data have been extensively studied, and related methods have been proposed for dealing with informative missingness based on supplemental (“reassessment”) sampling of some of those with missing data. We discuss study designs and analytic methods for dealing with both problems simultaneously in a 2X2 analysis. The motivating example comes from a case-control study of the association between low birthweight and the diagnosis of breast cancer later in life, where self-reported birthweight for some women is supplemented by accurate information from birth certificates.

The Odds-risk Inequality and Odds-risk Conversion

◆ Peng T. Liu, U.S. Food and Drug Administration

U.S. Food and Drug Administration, 5100 Paint Branch Pkwy., College Park, MD 20740

Peng.Liu@cfsan.fda.gov

Key Words: *relative risk, odds ratio, inequality, biased estimator*

The odds ratio calculated from a retrospective study could be used for estimating the relative risk when disease incidence is rare. However, the accuracy of estimation is determined not only by the incidence rate but also by the exposure rate and relative risk. The purposes of this article are: (1) to examine the functional relation between the odds ratio and relative risk, (2) to define and verify the odds-risk inequalities, and (3) to generate tables for odds-risk conversion.

The Drane-Wang Modification of the Logistic Dose Response Function

◆ J. Wanzer Drane, University of South Carolina; Huei Wang, Eli Lilly and Company; Walter W. Piegorsch, University of South Carolina; James Hussey, University of South Carolina; Francisco Sy, Centers for Disease Control and Prevention

University of South Carolina, Arnold School of Public Health, 205, Columbia, SC 29208

wdrane@gwm.sc.edu

Key Words: *dose-response, logistic, log odds ratio, Drane-Wang model*

In epidemiology, in the limit the dose-response function or positive predictive value is bounded by the prevalence of the disease. All, or almost all, risk factors when expressed in numeric form are non-negative. Based on this theory the Drane-Wang (Lixia Wang) modification of the logistic dose response function for continuous risk variables is constructed. Constrained maximum likelihood estimation is used to estimate its parameters. An example is given to illustrate properties of the modification of the logistic and also the fundamental theorem of epidemiology—namely, the limit of the cumulative dose response function (from above or below) is the prevalence of the disease it is modeling.

Limitations of the Log Binomial Regression Model for Estimating Risk or Prevalence Ratios

◆ Leigh Blizzard, University of Tasmania; David W. Hosmer, University of Massachusetts

University of Tasmania, Menzies Research Instiute, Private Bag 23, Hobart Australia 7008, 7008 Australia

Leigh.Blizzard@utas.edu.au

Key Words: *binary regression, odds ratio, logistic regression*

An estimate of the risk or prevalence ratio adjusted for confounders can be obtained from logistic regression, but it substantially overestimates when the outcome is not rare. The log binomial model, binomial errors and log link, is increasingly being used for this purpose. Its performance and appropriately modified logistic regression fit tests have not been evaluated. Extensive simulations of artificial data and bootstrap sampling of real data are used to compare the performance of the log binomial, logistic regression and a logistic regression-based method proposed by Schouten et al. (1993). Log binomial regression resulted in "failure" rates (nonconvergence, out-of-bounds predicted probabilities) as high as 59%. Estimates from the Schouten method produced fitted log binomial probabilities greater than one in up to 19% of samples. Coefficient estimates from these two models were similar, but the Schouten method over estimated the standard errors. Rejection rates for the Hosmer-Lemeshow, Stukel goodness-of-link test and the unweighted sum of squares tests were around 5%. No test had high power. Unquestioned use of the log binomial model is not recommended.

383 Methods in Diagnostic Testing

ENAR, Biometrics Section, WNAR, Section on Statistics in Epidemiology

Wednesday, August 11, 2:00 pm–3:50 pm

Receiver Operating Characteristic Surfaces: Inference and Applications

◆ Christopher A. Beck, University of Rochester Medical Center; Michael P. McDermott, University of Rochester

University of Rochester Medical Center, 601 Elmwood Ave., Box 630, Dept. of Biostatistics and Comp. Biology, Rochester, NY 14642

beck@bst.rochester.edu

Key Words: *diagnostic test, diagnostic accuracy, discrimination, compositional data*

Receiver operating characteristic (ROC) curves are commonly used to describe the performance of a diagnostic test in terms of discriminating between healthy and diseased populations. A popular index of the discriminating ability or accuracy of the diagnostic test is the area under the ROC curve. When there are three or more populations, the concept of an ROC curve can be generalized to that of an ROC surface, with the volume under the ROC surface (VUS) serving as an index of diagnostic accuracy. This paper discusses parametric and nonparametric methods for estimating ROC surfaces and VUS's given three-part compositional data that are of the form (p, q, 1-p-q), with each coordinate representing a guessed probability that the experimental unit arose from the corresponding population. This type of compositional data can be plotted in a ternary diagram, an equilateral triangle with each vertex representing one of the three unit vectors. A classification rule partitions this triangle into three regions, with data points classified according to which region they fall into. Several families of classification rules are discussed. Examples are provided to illustrate the proposed methods.

Semiparametric Leasts-squares-based ROC Analysis

◆ Zheng Zhang, University of Washington; Margaret S. Pepe, University of Washington

University of Washington, Dept. of Biostatistics, Campus Mail Stop 357232, Seattle, WA 98195-7232

zhangz@u.washington.edu

Key Words: *accuracy, ROC curve, regression, biomarker, diagnostic test*

Receiver Operating Characteristic (ROC) curve is a standard statistical tool for evaluating the accuracy of a continuous diagnostic test (e.g., biomarker). It provides a complete description of the test performance and a meaningful way to compare the performances of different tests. We have developed regression models for the ROC curve based on least squares. It provides a simple yet elegant way to estimate the ROC curve using standard linear regression algorithms. Asymptotic theory shows the new estimators are asymptotically unbiased and normally distributed. Extensive simulation studies have shown the new estimators have the similar efficiency as the two previously proposed estimators and are 20-30% more efficient than the nonparametric estimators. The method is illustrated in a pancreatic cancer dataset. We have extended our method to ROC curves comparison and covariate effects modeling. For categorical covariates that are available for both diseased and nondiseased population, the method is readily applicable. Asymptotic theories have been developed for independent tests/covariates. Simulation studies and data analysis are also included.

Combining Dependent Tests to Compare the Accuracies of Different Diagnostic Tests

◆ Yuqing Yang, Columbia University

Columbia University, 154 Haven Ave. #707, New York, NY 10032

yy2019@columbia.edu

Key Words: *ROC curve, repeated measurements, combining dependent tests*

In comparing different diagnostic tests, repeated measurements of the same characteristic are often taken under two or more distinct conditions for each experimental subject. The commonly seen example is the multireader ROC studies with multiple tests. To draw an overall conclusion on whether different diagnostic tests have same accuracies, various parametric and nonparametric methods have been proposed. We present a method of combining dependent tests by utilizing the idea in Wei & Johnson. Specifically, a test statistic from each reader is constructed by extending the structural components method of DeLong, DeLong, and Clark-Pearson, then a single test is constructed by combining all the individual test statistics. Numerical studies show that when readers have effect on the measurements, the proposed test yields appropriate size and power while tests based on existing method have inappropriate size and insufficient power.

Multiple Imputation for Correcting Verification Bias in Estimating Sensitivity and Specificity

◆ Ofer Harel, University of Washington; Xiao-Hua A. Zhou, University of Washington

University of Washington, VA Puget Sound Health Care System, 1660 South Columbian Way, 1/424, Seattle, WA 98108

ofer@u.washington.edu

Key Words: *missing data, verification bias, multiple imputation*

Sensitivity and specificity are one of the most widely used statistics to describe a diagnostic test. When all subjects have both test results and true status, the estimation of the sensitivity and specificity is build on two binomial distributions. Although the estimation of binomial confidence interval is a basic task in elementary statistics, it is well documented that this estimation is not trivial. When all subjects are screened using a common test, while only a subset of these subjects are tested using a golden standard test, it is well established that there is a risk for bias, called verification bias. When not all subjects have been verified, we can not any longer estimate the sensitivity and specificity separately, but need to use a special method for this estimation. There are several methods to estimate the sensitivity, specificity, and their standard errors in this kind of situations. The standard methods are very specific scenario oriented. They developed under some special cases of the verification choices. Approaching this problem from a missing data prospective allows us to use multiple imputation (MI) technique in order to impute the data.

Handling Missing Observations in the Estimation of Differences in Proportions from Clustered Data in a Matched-pair Design

◆ Carsten Schwenke, Schering AG; Jaakko Nevalainen, Schering AG

Schering AG, Muellerstrasse 170-178, Berlin, 13342 Germany

carsten.schwenke@schering.de

Key Words: *missing values, matched-pair, clustered data, binary outcomes*

In diagnostic studies, often the difference in the sensitivity or specificity of two diagnostic procedures is to be estimated. We consider studies with a matched-pair design, where all procedures are applied to each patient. With more than one observational unit per patient, e.g., several tumors in the same patient, statistical methods are to be used, which account for the clustering effect. Obuchowski (1998) developed such methods for complete datasets. In terms of clinical studies, this approach is often equivalent to the per-protocol-analysis, where incomplete cases are excluded. For an intent-to-treat analysis, these methods are insufficient. We propose a simple extension, where incomplete datasets can be analyzed without excluding incomplete cases. We divide the full sample into a set of data with complete cases and another set of data with incomplete cases. The differences in proportions are estimated for each subset considering the correlation between patients for the complete data and the clustering effect within the patient for both datasets. Then the two estimated differences are combined by appropriate weights.

Hierarchical Methods to Design Studies of Medical Test Accuracy

◆ Alicia Y. Toledano, Brown University

Brown University, Center for Statistical Sciences, 167 Angell St. 2nd Floor, Box G-H, Providence, RI 02912

Alicia_Toledano@Brown.EDU

Key Words: *sensitivity, screening, ROC curve, hierarchical model, diagnostic test*

Rapid advances in medical technology are leading to an increasing need for prospective, multicenter studies of medical test accuracy. Study planning is complicated by low prevalence of disease, variation in prevalence across centers, and often, variation in test accuracy across centers. This paper presents a framework for designing prospective, multicenter studies of medical test accuracy. We use hierarchical methods to address variation in prevalence and in test accuracy across institutions. We show that when there is low prevalence of disease, a design that specifies accruing a fixed number of positive subjects at each institution is more efficient than one that specifies accruing a fixed total number of subjects. We provide estimators of variation in accuracy that can be applied to pilot data to obtain values for input parameters, and normal approximations to the probability of achieving desired precision in estimating central tendencies in accuracy that can be used to reduce the number of parameter combinations evaluated with Monte Carlo methods. We illustrate use of the proposed method to design a national screening trial.

Prequential Tests in Classification and Regression Trees

◆ Jialu Zhang, University of Maryland, Baltimore County; Francoise Seillier-Moiseiwitsch, University of Maryland, Baltimore County

University of Maryland, Baltimore County, Dept. of Mathematics and Statistics, 709 Tamarack Way, Apt. 3D, Herndon, VA 20170

jzhang8@math.umbc.edu

Key Words: *prequential test, classification, CART, cross-validation*

The data-mining procedure implemented in CART (Classification and Regression Trees) extracts information from the data by constructing binary decision trees. The current CART algorithm exhaustively searches all levels of all covariates and decides on the node-splitting criterion by minimizing node impurity. The node-splitting process goes on until there are too few observations at a node or until the node consists of observations of a single type. Then the large binary tree is pruned back using cross-validation or some other pruning approach. We implement into CART a new model evaluation method, based on prequential testing, to replace the splitting and pruning steps. As a direct consequence, these steps now involve formal statistical tests to validate the decisions. This approach is suited to both parametric and nonparametric decision rules, and assesses models on the basis of the accuracy of their probabilistic predictions for future events. More specifically, data are first divided into a training sample and an evaluation sample. The training sample is extended by one observation at a time to update estimates and make a prediction for the next outcome.

384 Statistics and Cancer △ ⌘

Biometrics Section, ENAR

Wednesday, August 11, 2:00 pm–3:50 pm

Feature Selection with SVM Reformulation and Expression-array-based Cancer Classification

Sunhee K. Ro, Roche Molecular Systems, Inc.; ◆ Wei-min Liu, Roche Molecular Systems, Inc.

Roche Molecular Systems, Inc., 1145 Atlantic Ave., Alameda, CA 94501

wei-min.liu@roche.com

Key Words: *microarray, cancer classification, feature selection, support vector machine*

Gene expression microarray analysis is being used extensively for cancer classification and support vector machine (SVM) is one of the successful classification methods. While SVM can be applied to extremely high-dimensional data such as microarray, the removal of irrelevant and redundant features may produce better results. Classical wrapper methods such as the SVM RFE can be used for feature selection while avoiding the combinatorial explosion with greedy technique; however, it is slow because the classification performance should be calculated at every iteration to find the optimum number of features. Weston et al. (2003) proposed to reformulate the SVM problem as minimizing the number of nonzero elements in the weight vector in SVM decision function. The algorithm amounts to a modification of SVM with multiplicative rescaling of the training data at every iteration so that the weights of useless features approach 0 rapidly. We have extended this algorithm to classify the multiple subclasses of leukemia, which consists of the all-pairwise classifiers each of which was trained with the union of the genes selected pairwisely with the multiplicative rescaling.

Kolmogorov-Smirnov Test for Image Comparison

◆ Eugene Demidenko, Dartmouth College

Dartmouth College, 7927 Rubin, DHMC, Lebanon, NH 03756

eugene.demidenko@dartmouth.edu

Key Words: *Image, nonparametric test, cancer, histology*

An essential task of image analysis is image comparison. Although the histogram is widely used in image processing to represent the distribution of gray levels, no statistical tests are available to compare images. We say that two images are the same if they have the same grayscale distribution. Thus, if F1(g) and F2(g) are two empirical distribution functions of gray levels g=0,1,.,255 from two gray images, the hypothesis reduces to H0:F1=F2. We use the Kolmogorov-Smirnov test, although other nonparametric test can be applied as well. As a graphical tool, distribution functions have an advantage over histograms because several of them can be plotted on the same graph. We illustrate the Kolmogorov-Smirnov test and computation of the respective p value by histology images from two cancer treatments.

Cluster Analysis for Continuous Data with an Excess of Zeros

◆ Kimberly Siegmund, University of Southern California

University of Southern California, 1540 Alcazar St., CHP-220, Los Angeles, CA 90089

kims@usc.edu

Key Words: *class discovery, cluster analysis, mixture models, gene expression*

Abnormal DNA methylation is typical in cancer. Since DNA methylation profiles vary across tumor types and subtypes, it is believed that clustering DNA methylation profiles may uncover novel disease subgroups. DNA methylation data obtained using the MethyLight technology is quantitative with an excess of zeros. For a region of DNA, MethyLight measures the occurrence of fully methylated alleles. For many samples it finds none while for others it finds variable levels. We introduce a Bernoulli-lognormal mixture model for clustering continuous data with an excess of zeros and compare it to standard model-based clustering methods for discrete data and for continuous data. In a simulation study we find the Bernoulli-lognormal mixture model has the lowest misclassification error rate compared to competing approaches. We illustrate the methods using DNA methylation profiles from a study of lung cancer cell lines. The Bernoulli-lognormal mixture model has the lowest cross-validation error for distinguishing lung cancer subtype (nonsmall cell vs. small cell) and allocates samples to classes with the lowest uncertainty.

A Mixed Effects Model for Serial PSA Following Surgery in Men with Prostate Cancer

◆ Mousumi Banerjee, University of Michigan; Julie George, Wayne State University

University of Michigan, Dept. of Biostatistics, 1420 Washington Heights, Ann Arbor, MI 48109

mousumib@umich.edu

Key Words: *random effects, longitudinal data, serial PSA*

Prostate cancer is more prevalent and causes more deaths in African Americans compared to Caucasian men. It is clinically useful to know when PSA levels first begin to rise rapidly after surgery, and to determine if the natural history of PSA progression post-surgery is different in the two races. Follow-up serial PSAs describe a dynamic evolution of the disease, and thus contain important information regarding progression. This article uses a piecewise nonlinear mixed-effects model to describe longitudinal changes in PSA in men after surgery. The model is linear after surgery and exponential nearer the time recurrence is detected. The time at which the PSAs change from linear to exponential phase is unknown but is estimated by including random terms that allow each subject to have his own transition time. The model also accounts for race, age, and stage. Various parameters are allowed to differ between the two races. Our analysis suggests that transition times are different for the two races, and Caucasians have a longer latency period than African Americans. This model may be useful in a variety of research settings where it is necessary to estimate the unknown time of an event.

A Nonparametric Measure to Detect Interactive Information in Discriminant Analysis of Gene Expression

◆ Xin Yan, Columbia University; Tian Zheng, Columbia University; Shaw-Hwa Lo, Columbia University

Columbia University, 542 West 112th St. #2B, New York, NY 10025

xyan@stat.columbia.edu

Key Words: *Gene Profile Association Score, gene expression, breast cancer, informative gene*

Despite biological acknowledge of influence of gene interactions, marginal statistical measures were broadly used on feature selection problems in gene expression data. In our previous paper, we proposed using Gene Profile Association Score (GPAS), a statistical measure to explore higher-order interactive gene information with respect to binary categorical difference. Original GPAS however, was somewhat arguable because of its assumed input of discritized data, thus an potential loss of information. Inspired by the analysis of spatial marked point process, we now introduce our sign-test based Gene Profile Association Score (sGPAS), where all numerical values of expression will be taken into account. Similar to what we did previously, the performance of sGPAS was compared to a variety of predictors. Empirical validation results showed substantial improvement to all marginal predictors and a comparable performance to the original GPAS predictor. SGPAS provides an alternative to GPAS, allowing people to extract differentiation information with continuous expression results, while preserving the gene-to-gene interactions at the same time.

A Parametric Model for the Times to Cessation and Recovery of Menses in Breast Cancer Patients Undergoing Adjuvant Treatment

◆ Suzanne E. Szwarc, Harvard University; Marco Bonetti, Harvard University

Harvard University, Dept. of Biostatistics, 655 Huntington Ave., Boston, MA 02115

sszwarc@hsph.harvard.edu

Key Words: *censoring, distinguishability*

Because some breast cancer adjuvant treatments for premenopausal patients who have undergone surgery cause interrupted menses, many patients express concerns about fertility following treatment. The observation of the resumption of menses after treatment completion is complicated by the fact that treatment-induced amenorrhea is not distinguishable from natural menopause, unless menses are observed to resume post treatment. We discuss a parametric approach designed to model these events, and apply it to data arising from a clinical trial run by the International Breast Cancer Study Group.

Predicting U.S. Cancer Mortality Counts Using Semiparametric Bayesian Techniques

◆ Kaushik Ghosh, George Washington University; Ram C. Tiwari, National Cancer Institute

George Washington University, Dept. of Statistics, Washington, DC 20052

ghosh@gwu.edu

Key Words: *local linear model, local quadratic model, Dirichlet processes, Markov chain Monte Carlo, health statistics, time series*

We present two models for short-term prediction of the number of deaths that arise from common cancers in the United States. For very short-term projections, we use a local linear model in which the slope of the segment joining consecutive death counts is assumed to be randomly distributed. For slightly longer prediction periods, we use a local quadratic model where the series incorporates a random acceleration term. The slope and acceleration terms are assumed to have nonparametric distributions with Dirichlet Process Prior. Through Markov chain Monte Carlo simulations, we can estimate the mean and variance of the predicted number of deaths. We apply the proposed methods to selected common cancer sites and compare the results with those obtained from the currently available methods. It is seen that the local models are very flexible and result in improved predictions.

385 Variance Components and Association in Multivariate Models for Continuous Data

Biometrics Section, WNAR

Wednesday, August 11, 2:00 pm–3:50 pm

Assessing Adequacy of the Covariance Structure in the Generalized Linear Mixed Model

◆ Jean Orelien, SciMetrika

SciMetrika, 2 Davis Dr., Durham, NC 27709

jorelien@scimetrika.com

Key Words: *mixed model, goodness-of-fit, covariance structure*

In the generalized linear mixed model, few tools are available for assessing the adequacy of the covariance structure. Statistics such as the Likelihood Ratio Test (LRT), the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) provide a way to compare the fit of a model against another one. Other statistics such as the concordance correlation coefficient and other pseudo R-squares have been proposed to assess goodness-of-fit for the model at hand without requiring the fit of a second model. None of the statistics mentioned above provide a way to ascertain the appropriateness of the covariance structure. Vonesh and Chinchilli (1986) proposed a test to compare the covariance of the parameter estimates from a generalized linear mixed model to the robust covariance matrix from a generalized estimated equation (GEE) model. However, the performance of this test has not been fully evaluated in simulation (results from a limited simulation of 400 samples were offered) or proven analytically. We present results from a simulation that estimates the empirical Type I error rate and power of this test for various types of covariance structures.

Modeling Continuous Bivariate Longitudinal Data

◆ Robert E. Weiss, University of California, Los Angeles

University of California, Los Angeles, School of Public Health, Biostatistics, Los Angeles, CA 90095-1772

robweiss@ucla.edu

Key Words: *random effects, multivariate, biostatistics, correlation*

Bivariate longitudinal data occurs when we analyze the interrelationships of two response variables both assessed longitudinally on the same subjects. A common issue when analyzing the two variables is whether and how the two variables are correlated. At the big-picture level, we can ask whether the two variables are correlated or not and is the relationship positive or not. At the detail level, we can ask exactly how the variables are correlated. I review several basic models that can be fit with current software, beginning with the bivariate random intercept model where both residual errors are correlated and the random intercepts are correlated. This model has three important nested submodels and can be expanded to a model where one or both variables has a random slope. Time permitting, I consider additional models: the unstructured covariance model, the product correlation model and factor analytic models.

Modeling Multivariate Longitudinal Profiles

◆ Geert Verbeke, Katholieke Universiteit Leuven; Steffen Fieuws, Biostatistical Centre, K.U.Leuven

Katholieke Universiteit Leuven, Biostatistical Centre, U.Z. Sint Rafael, Kapucijnenvoer 35, Leuven, B-3000 Belgium

geert.verbeke@med.kuleuven.ac.be

Key Words: *random effects, multivariate, linear mixed models, association*

Multivariate longitudinal data arise when a set of different outcomes on the same unit is measured repeatedly over time. There are different situations where a joint modeling approach is needed. First, the association structure can be of importance. A possible question might be how the association between outcomes evolves over time or if outcome-specific evolutions are related to each other. In a second situation, the aim is to improve the results of a discriminant analysis by using more than one longitudinally measured outcome. In another situation, interest lies on estimation of the fixed effects. As an example, consider testing the difference in evolution between many outcomes. Linear mixed models are a flexible tool for joint modeling purposes, especially when the gathered data are unbalanced. However, computational problems due to the dimension of the joint covariance matrix of the random effects arise as soon a the number of outcomes and/or the number of used random effects increases. A pseudo-likelihood approach will be presented to circumvent this problem. Real data will be used to illustrate the approach.

Comparison of Statistics Involving Fixed and Random Effects to Discern Inherent Discrepancies in Biomarker Reproducibility

◆ Irene B. Helenowski, Northwestern University; Borko D. Jovanovic, Northwestern University Medical School; Vijayalakshmi Ananthanarayanan, Northwestern University; Peter H. Gann, Northwestern University

Northwestern University, 680 N. Lake Shore Dr., Suite 1102, Chicago, IL 60611

i-helenowski@northwestern.edu

Key Words: *variance components, intraclass correlation coefficient, reproducibility, random effects, fixed effects, biomarkers*

Traditionally, reproducibility statistics, such as the intraclass correlation coefficient (ICC), have involved ratios of variance components. In some circumstances, however, one or more of the factors considered in the ICC are defined as a fixed effect rather than a random effect. Inherent differences that we would wish for a reproducibility statistic to detect can be more effectively discerned in fixed effects from the parameters of the fixed effects themselves than from the variance components of fixed effects. Inherent differences in random effects can be discerned from their variance components efficiently, on the other hand. We consider reproducibility statistics involving the estimated parameters of fixed effects and variance components of random effects. Analyses for this presentation will be applied to p27 indexes coming from the prostate gland.

Frequency of Use of Statistical Techniques for Assessing Agreement between Continuous Measurements

◆ Stephen W. Looney, Louisiana State University; Joseph L. Hagan, University of Louisville

Louisiana State University, Health Sciences Center, School of Public Health, 1600 Canal St., Suite 1100, New Orleans, LA 70112

sloon1@lsuhsc.edu

Key Words: *inter-rater reliability, validity, intra-class correlation, Pearson correlation, Bland-Altman method*

This study examined the frequency of use of statistical techniques that have been proposed for assessing agreement between continuous measurements. A survey of clinical research articles published in 2001 yielded a total of 189 method comparison studies. The most commonly used statistical technique for measuring agreement in these studies was the intraclass correlation coefficient (ICC) (118 studies, 62.4%), followed by the Pearson correlation coefficient (PCC) (53 studies, 28.0%), and the Bland-Altman method (25 studies, 13.2%). These results indicate that the PCC is still commonly used for method comparisons, despite the fact that it has been known to be inappropriate for this purpose for over 30 years. The most commonly used method, the ICC, is also known to have shortcomings as a measure of agreement, and has only limited applicability in method comparison studies. Given the current state of the clinical literature with regard to statistical analyse that are used in method comparison studies, statisticians should be more proactive in promoting the use of appropriate statistical techniques for assessing agreement between continuous measurements.

Analysis of Variance Components in Linear Mixed Effects Models

◆ Miao-Yu Tsai, National Taiwan University; Chu-Hsing Hsiao, National Taiwan University

National Taiwan University, Division of Biostatistics, Institute of Epidemiology, Taipei, Taiwan

miao718@ms41.hinet.net

Key Words: *Bayes factor, correlation, GLMM, longitudinal, Schwarz criterion, variance components*

This research was originally motivated by a longitudinal study of myopia intervention trial. In addition to testing if one treatment is more efficient than the others, it is also of interest to modeling the observed variance and correlation between two eyes and within repeated measurements per subject. The structures of the variance components can be modeled alone or explained via certain random effects. We consider three different structures for discussion. The first model considers correlated residuals, the second regards the correlation as part of a random effect, and the third model assigns correlated random effects to describe the correlation. We explain the relation between these three models and elucidate the situations where they cannot be differentiated. We also evaluate the errors using both estimates and criteria of model selection when incorrect models are fitted. Finally, we consider simulation studies and the myopia intervention trial for illustrations.

386 Mixtures, Clusters, Random Functions, and Decision Trees

Section on Bayesian Statistical Science, Section on Statistical Computing

Wednesday, August 11, 2:00 pm–3:50 pm

Label Switching in Mixture Models

◆ Eric Loken, Pennsylvania State University

Pennsylvania State University, S162 Henderson Bldg., University Park, PA 16802

loken@psu.edu

Key Words: *mixture models, Bayesian inference, label switching*

A well-known identification problem in estimating mixture models is that the observed data likelihood is invariant to permutations of the group labels. In Bayesian estimation, label switching during posterior simulation will distort posterior summaries. A proposal to address mode-switching is offered. If one or more cases are "pre-classified," that is, if the group membership is assumed to be known, label switching can be dramatically reduced. In a mixture of two groups, assuming that a case belongs to group 1 can be considered as simply defining the labeling. The modification is easy to implement in an EM algorithm to maximize the likelihood, and also in MCMC simulation of the posterior distribution. It can be shown that pre-classifying modifies the likelihood by eliminating the nuisance mode and leaving the mode of interest almost perfectly intact. MCMC simulations with a latent class model and with a mixture of two exponentials show that the technique works well and compares favorably with other strategies. The extension of the technique to mixtures with three or more components will be explored.

Bayesian Cluster Analysis for Longitudinal Data

◆ Rolando de la Cruz-Mesía, Pontificia Universidad Católica de Chile

Pontificia Universidad Católica de Chile, Departamento de EstadÌstica, Casilla 306, Correo 22, Santiago, 6904411 Chile

rolandod@mat.puc.cl

Key Words: *cluster analysis, longitudinal data, MCMC methods, nonlinear hierarchical models*

Cluster analysis is an important exploratory tool for analyzing many types of data. In particular, we explore the problem of clustering longitudinal data, under the form of multiple nonlinear profiles. We introduce a general finite mixture of nonlinear hierarchical models for longitudinal data. Parameter estimation is done via appropriate posterior simulation schemes. We illustrate our Bayesian method with the analysis of a real dataset. We use the example to compare our Bayesian approach with a classical method based on the EM algorithm.

Bayesian Regression for Circular Data

◆ Barbara J. George, Defense Manpower Data Center; Kaushik Ghosh, George Washington University

Defense Manpower Data Center, 1600 Wilson Blvd., 4th Floor, Arlington, VA 22209

georgebj@osd.pentagon.mil

Key Words: *directional data, semiparametric Bayes, MCMC, Dirichlet process*

Circular data are observations that are represented as points on a unit circle. Times of day and wind direction are two such examples. The proposed work uses a Bayesian approach to regress a circular variable on a linear predictor. The regression coefficients are assumed to have a Dirichlet process prior with a bivariate normal baseline. Use of this semiparametric Bayesian approach gives added flexibility to the model relative to frequentist approaches. Markov chain Monte Carlo techniques are used to fit the proposed model. This technique is illustrated using simulated data.

Evaluation of Multilevel Decision Trees

◆ Erwann Rogard, Columbia University; Andrew Gelman, Columbia University; Hao Lu, Thales Corporation

Columbia University, New York, NY 10027

er317@columbia.edu

Key Words: *decision analysis, hierarchical Bayes, jackknife, nested computation*

The evaluation of decision trees under uncertainty is difficult because of the required nested operations of maximizing and averaging. Pure maximizing (for deterministic decision trees) or pure averaging (for probability trees) are both relatively simple because the maximum of a maximum is a maximum, and the average of an average is an average. But when the two operators are mixed, no simplification is possible, and one must evaluate the maximizations and averagings in a nested fashion, following the structure of the tree. Nested evaluation requires large sample sizes (for data collection) or long computation times (for simulations). An alternative to full nested evaluation is to perform a random sample of evaluations and use statistical methods to perform inference about the entire tree. We show that the most natural estimate is biased and consider three alternatives: normal-theory bias correction, the jackknife, and hierarchical Bayes inference. We explore the properties of these inferences through a simulation study and discuss general approaches to the problem.

Wavelet-based Functional Mixed Models

◆ Jeffrey S. Morris, University of Texas M.D. Anderson Cancer Center

University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd. Box 447, Houston, TX 77030-4009

jeffmo@odin.mdacc.tmc.edu

Key Words: *functional data analysis, wavelets, Bayesian methods, shrinkage estimators, mixed models*

An ever-increasing number of studies yield functional data, where the ideal observational units are curves and the observed data consist of sets of curves sampled on a fine grid. We develop new methodology generalizing the linear mixed model to the functional mixed model framework, with model-fitting done using a Bayesian wavelet-based approach. This method is very flexible, accommodating functions of arbitrary form and the full range of fixed effects between-curve covariance structures of the mixed models framework. It yields nonparametric estimates of the fixed and random effects functions that are adaptively regularized via a nonlinear shrinkage prior on the fixed effects' wavelet coefficients, plus estimates of any between-curve and within-curve correlation surfaces. Posterior samples allow us to perform Bayesian inference on all model quantities. This method is appropriate for functional data characterized by numerous local features like peaks, since our adaptive regularization procedure regularizes the functions with minimal attenuation of dominant local features. We apply this method to complex and irregular functional data from two biological studies.

Bayesian Modeling of Hydrograph

◆ James Merleau, Universite de Montreal

Universite de Montreal, CP 6128, succursale "centre-ville," Montreal, PQ H3C 3J7

merleau@dms.umontreal.ca

Key Words: *Bayesian nonparametric modeling, registration, regression splines*

The generation of synthetic hydrographs which reproduce accurately the typical characteristics of observed ones is a standing problem in hydrology. The difficulty of building a hydrograph generator stems from the varying periodicity of salient features from one year to another; this changing periodicity makes averaging of observed hydrographs useless in order to get a standard hydrograph. Registration of observed hydrographs, according to certain important characteristics such as the spring flood, can be used to obtain this profile. A Bayesian nonparametric model based on spline functions can then be used to generate hydrographs. From the posterior density on the spline coefficients, it is possible to generate sets of coefficients that define synthetic hydrographs which exhibit the typical features encountered in observed hydrographs.

387 Sampling and Other Aspects of Survey Design ▲

Social Statistics Section, Section on Survey Research Methods

Wednesday, August 11, 2:00 pm–3:50 pm

Responsive Design for Household Surveys

◆ Steven G. Heeringa, University of Michigan; Robert Groves, University of Michigan

University of Michigan, Survey Research Center, 426 Thompson St., Ann Arbor, MI 48104

sheering@isr.umich.edu

Key Words: *survey error, survey cost, process data, sample replication, response propensity, two-phase sampling*

For more than 70 years, the development of sampling techniques and data collection methods for household surveys has focused on reducing costs and errors. Historically, survey designers have relied heavily on experiences in prior surveys to model the error and cost structures for a new design. Despite the advantage that existing information brings to the survey planning process, many survey designs do not incorporate features that anticipate the uncertainty inherent in the essential survey conditions. Very few are designed to respond to real-time data with the possible exception of a crisis response to major failures in estimating survey design parameters or data collection costs. The transition to computerized data collection and survey management systems has presented survey designers and practitioners with a wealth of new real-time data on the survey process. These data include not only traditional sample design and production indicators but detailed data on the work progress and associated costs for distinct sampling units, individual interviewers and even for individual sample lines. This paper presents a new model for survey design and management.

Triad Sampling in Household Surveys

◆ Jeremy Aldworth, RTI International; James R. Chromy, RTI International

RTI International, 3040 Cornwallis Rd., RTP, NC 27709

jaldworth@rti.org

Key Words: *Brewer's method, Sampford's method, household rosters, person selection, unequal probability sampling*

The motivation for selecting three persons from a dwelling unit (DU) derives from an interest in studying the behavioral relationships among certain triads, (e.g., two parents and a child, a parent and two children, etc). Computer-assisted DU screening provides the mechanism for targeted sampling of individuals, pairs, and even triads within a DU. Chromy and Penne (2002) showed how a modification of Brewer's (1963, 1974) method for samples of size two was used to select samples of 0, 1, or 2 persons from eligible Dus. They also developed a second adaptation to control the number of pairs selected. Chromy and Penne's modification of Brewer's method and their adaptation to control the number of persons selected from a DU is extended to the case of sampling triads within DUs. Some empirical data on household roster composition, and response rates based on the number of persons selected, are presented from the National Survey on Drug Use and Health (NSDUH). Finally, the results of some simulation of the sample selection and response process are presented as a means of evaluating alternatives.

Design Effects of Linked Population/Establishment Surveys

◆ Monroe G. Sirken, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd., Room 5212,
Hyattsville, MD 20902

MGS2@CDC.gov

Key Words: *establishment surveys, population surveys, network sampling, two-phase sample surveys*

The linked/population/establishment survey (LPES) is a two-phase sample survey that estimates the volume of transactions between populations and establishments. Phase I is a population sample survey in which households identify the establishments with whom they had transactions during a specified calendar period and report the number of their transactions with each establishment. Phase II is an establishment sample survey in which the establishments identified in phase I report the variables of interest for samples of their transactions with all households. This paper compares the sampling errors of (1) the LPES and the conventional two-stage pps establishment survey of equivalent household and transaction sample sizes, and (2) the LPES and the conventional household survey of equivalent household sample size.

An Example of Using Two-phase Sampling Techniques to Integrate Disjoint Surveys and Data Sources into One Survey

◆ Jack R. Lothian, Statistics Canada; Tracy Tabuchi, Statistics Canada

Statistics Canada, R.H. Coats Bldg. 11-O, Business Survey Methods
Division, Ottawa, ON K1A 0T6 Canada

lothian@statcan.ca

Key Words: *two-phase sampling, survey redesign, administrative data*

In recent years a number of coverage, lack of consistency and data gap issues has led Statistics Canada to consider re-engineering the survey program of motor freight carriers. First, deregulation of the industry has increasingly blurred the lines between what were formerly considered to be two disjoint industries and surveys. Second, electronic reporting of administrative data is now the norm in the Canadian business community. Statistics Canada is increasingly using this data to reduce the burden on respondents, decrease survey costs, and improve the quality of small-area estimation. The re-engineering will integrate the two surveys and incorporate administrative taxation data into the survey. These changes present unique challenges because the original two surveys measured different characteristics in the two populations and the administrative data does not provide sufficient detailed data to estimate all the required characteristics. The paper presents a review of the problems encountered and their solutions including a two-phase sampling procedure that permits estimation of the original characteristics covered by the two surveys.

Empirical Study on the Second-stage Sample Size

◆ Ryan Petska, Ernst & Young LLP; Yan Liu, Ernst & Young LLP; Mary K. Batcher, Ernst & Young LLP; Amy Luo, Ernst & Young LLP

Ernst & Young LLP, 1225 Connecticut Ave. NW, Washington, DC
20036

Ryan.Petska@ey.com

Key Words: *second-stage sample, model-based approach, design-based approach*

In the two-stage sampling design, how large should the second-stage sample size be? In the design-based approach, the second-stage sample size is incorporated in the variance formula. In the model-based approach, the second-stage sample size is not explicitly expressed but only factored in the variance formula. In the resampling method, the second-stage sample size is not even counted in the variance formula. It is known that as the second-stage sample size increases, the overall variance decreases. But the variance is not quantified by the second-stage sample size. We will investigate the impact of the second-stage sample size to the overall variance in different estimation approaches.

The NASS Question Repository System

◆ Daniel Beckler, National Agricultural Statistics Service

National Agricultural Statistics Service, USDA, Room 305, 3251
Old Lee Hwy., Fairfax, VA 22030

dan_beckler@nass.usda.gov

Key Words: *web data collection, questionnaire design, form generation*

The National Agricultural Statistics Service (NASS) conducts hundreds of surveys annually on the nation's farmers and agribusinesses. For most surveys, multiple questionnaire versions are needed to address differences in agriculture between states. The questionnaires also need to be developed for multiple data collection modes: mail, telephone/CATI, and most recently the World Wide Web. In order to efficiently create the numerous questionnaires needed for all of the survey, state, and mode combinations, NASS developed a client-server based Question Repository System (QRS). The QRS includes a user interface to build properly formatted questions for the various modes; these questions are then stored in a central database. The stored questions may be retrieved and used to build questionnaires, which can be saved, printed or ported to a Web server. This paper describes the capabilities and some technical details of the QRS.

Quality Control of Data Entry for the American Community Survey and the Impact of Errors on Data Quality

◆ Andre L. Williams, U.S. Census Bureau; Rita J. Petroni, U.S. Census Bureau; Bryan D. Garrett, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd. STOP 9200, Washington, DC 20233-9200

andre.l.williams@census.gov

Key Words: *mail-out, keying, data capture, consequential errors, quality control*

The American Community Survey (ACS) is being developed by the Census Bureau to replace the decennial census long form. ACS collects important socioeconomic data continuously in three-month cycles. ACS mails out questionnaires during month one, then follows nonrespondents with computer assisted telephone and personal interviewing in months two and three, respectively. The paper questionnaires from mail-out are keyed. ACS has a quality control program which samples keyed documents and estimates keying error rates. A sample of errors from the 2002 ACS was examined to measure the impact of keying errors on data quality. Examination of these errors led to the classification of some errors as inconsequential since they were not expected to affect data quality. This paper estimates ACS keying error rates, provides a preliminary estimate of the proportion of errors that affect data quality, compares ACS error rates of related data capture methods used by Census 2000, and describes the distribution of the main types of keying errors.

388 Record Linkage and Administrative Data ▲ ⌘

Section on Survey Research Methods, Section on Government Statistics

Wednesday, August 11, 2:00 pm–3:50 pm

Regression-based Statistical Matching: Recent Developments

◆ Chris Moriarity, U.S. General Accounting Office; Fritz J. Scheuren, NORC, University of Chicago

U.S. General Accounting Office, 200 Spring Ave., Takoma Park, MD 20912

chrismor@cpcug.org

Key Words: *data fusion*

We have described a method in several articles (2001, 2003) for statistically matching two samples. One sample is assumed to contain (X,Z) and the other is assumed to contain (X,Y), both drawn from a common nonsingular normal (X,Y,Z) distribution. Following Kadane (1978) and Rubin (1986), we employ regression in our approach. We assess the uncertainty introduced during the match that is due to the unobserved (Y,Z) relationship by repetition over a range of (Y,Z) values that are consistent with the observed data. In the final step of our algorithm, we replace predicted values with observed data by a match of the two samples to obtain final files consisting only of observed data, consistent with traditional statistical matching procedures. Prior to matching, we add random residuals to our regression-based estimates, an essential step in our method. Our approach for estimation of the amount of residual to add, using subtraction and estimates from both files, can be negative. Rassler (2002) suggests a different approach for residual estimation, which is always non-negative. We compare the two methods and discuss other recent developments.

Evaluating Coverage on the Census Bureau's Master Address File Using Independent Housing Unit Estimates

◆ Chris Johns, U.S. Census Bureau; Joseph A. Burcham, U.S. Census Bureau

U.S. Census Bureau, 4301 Suitland Rd., Rm 1107/2, Suitland, MD 20746

Christopher.M.Johns@census.gov

Key Words: *dual system estimate, longitudinal analysis, Delivery Sequence File, Address Characteristic Type, housing unit coverage, administrative records*

In order to conduct Decennial Censuses as well as various sample surveys throughout the decade, the Census Bureau requires a Master Address File that is as complete and accurate as possible at all times. This maintained Master Address File calls for periodic evaluation activities to be used in planning operations to improve coverage and other aspects of quality on the Master Address File. This paper discusses our plans for using housing unit estimates from different sources to produce national measures of coverage on the Master Address File throughout the decade. We also will compare the different housing unit estimates at the county level to identify potential Master Address File problem counties that may need review, and examine various characteristics of these counties.

Statistical Use of Goods and Services Tax Data

◆ Louis Pierre, Statistics Canada; Marie Brodeur, Statistics Canada

Statistics Canada, Jean-Talon Bldg., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

pierlou@statcan.ca

Key Words: *administrative data, calendarization, model, edit and imputation, Goods and Services Tax (GST),*

In a rapidly changing world, decision-makers on the economic scene need quality information if they are to understand economic activity and react quickly. The monthly economic statistics program of Statistics Canada (STC) helps provide this information in a timely fashion. However, the collection of monthly data imposes a significant response burden on Canadian businesses. With the development of new methodological approaches, STC is increasingly using existing sources of administrative data such as the GST (Goods and services Tax) files to produce quality economic estimates while reducing the response burden. The GST is a tax levied on the consumption of goods and services and is accounted for on the value added at each stage of production. However, the use of administrative data poses major challenges. For example, sophisticated E&I programs must be developed in order to detect and correct various problems that might exist in the raw data. Also, administrative data are not all available in time on a monthly basis. It has therefore been necessary to develop methodologies for calendarization and techniques that combine survey data with GST data.

Using GIS-based Property Tax Records as a Cost-effective Alternative to Traditional Household Listing in Area Samples

◆ William D. Kalsbeek, University of North Carolina; Sarah T. Kavanagh, Quintiles Transnational Corporation

University of North Carolina, Survey Research Unit, 730 Airport Rd., Suite 100, Chapel Hill, NC 27599-2400

bill_kalsbeek@unc.edu

Key Words: *area sampling, GIS, household sampling, frame construction, sample coverage*

Household frames in area probability samples have traditionally been constructed by sending trained field staff to construct these lists on location. The cost of doing so often constitutes a sizable portion of the field operation budgets. This paper investigates an alternative solution to the household frame construction task using GIS and GPS technology applied to property tax files, which are increasing found to be available in GIS format through local tax offices. We present findings from an field comparison study in Central North Carolina involving the GIS-based approach and the traditional household listing approach, using both cost and frame coverage evaluation criteria. We find that the GIS-based approach is less costly, has comparable sample coverage outcomes, and provide locator information that is easy and effective to use.

389 Variance Estimation III △

Section on Survey Research Methods

Wednesday, August 11, 2:00 pm–3:50 pm

Estimation of Stratum Variances for Sample Allocation in NASS Agricultural Surveys

◆ Charles R. Perry Jr., National Agricultural Statistics Service; Raj S. Chhikara, University of Houston, Clear Lake; Marty Spears, University of Houston, Clear Lake

National Agricultural Statistics Service, U.S. Dept. of Agriculture, 3251 Old Lee Hwy., Room 305, Fairfax, VA 22030-1504

charles_perry@nass.usda.gov

NASS updates its area sampling frame for its agricultural surveys on a regular basis, but its update often has a lag period, sometimes of many years. Since the use of agricultural land in an area may change during two consecutive update periods, the previous sample design stratum variance estimates may not reflect the actual variance for one or more strata. This has led the NASS to consider development of an alternative method of estimating stratum variances to be used in sample allocation. A nonlinear model based on a variance function originally proposed by Mahalanobis (1939) for crop acreage and by Smith (1936) for crop yield, can be a basis for the development of a stratum variance model. In this study, the stratum variance is modeled for each of the eight agricultural items used in the multivariate sample allocation by the NASS for its agricultural surveys. Survey data from a previous year are used in the stratum variance model development. The modeling involves determination of a dominant region for each of the eight items, the level for data-fitting, and evaluation of model fits. The presentation will include a discussion of the modeling methodology and more.

Using the Peters-Belson Method to Measure Disparity in Complex Surveys

◆ Sowmya R. Rao, National Cancer Institute; Barry I. Graubard, National Cancer Institute; Joseph L. Gastwirth, George Washington University

National Cancer Institute, EPS 8047, 6120 Executive Blvd., MSC 7244, NIH, DHHS, Bethesda, MD 20892

raos@mail.nih.gov

Key Words: *disparity, survey methods, Taylor linearization variance, delete-one-group jackknife variance, logistic regression, multiple linear regression*

The Peters-Belson (PB) approach used to assess wage discrimination can also be used to analyze disparities in large surveys for a variety of health and social behaviors including cancer-screening. This approach fits a regression model to the majority group and estimates the expected value a minority individual would have from this equation. The average difference between observed and expected values for the minority members estimates the disparity not explained by the covariates. Because of the complex sample designs of surveys, we show how to apply the Taylor linearization and delete-one-group jackknife methods to estimate standard errors for the disparity. The 1998 National Health interview Survey on cancer-screening among women is used for illustration. We found that the Taylor and jackknife methods gave similar standard errors. Although the jackknife method is easy to apply, it is more computer intensive than the Taylor method. The Taylor linearization method can be easily implemented using the existing software packages with the appropriate commands to account for the complex survey design.

Inclusion Probability Proportional to Size Sampling: A Nonlinear Programming Approach to Ensure a Non-negative and Stable Variance Estimator

◆ Sun Woong Kim, Dongguk University; Steven G. Heeringa, University of Michigan; Peter W. Solenberger, University of Michigan

Dongguk University, Dept. of Statistics, Jung Gu Pil Dong 3 Ga 26, Seoul, 100-715 South Korea

sunwk@dongguk.edu

Key Words: *Sen-Yates-Grundy variance estimator, joint probabilities, binary block designs, nonlinear programming*

Sample designs that use inclusion probability proportional to size sampling (IPPS) should have certain desirable properties including non-negativity and stability of the Sen-Yates-Grundy (1953) variance estimator. Jessen (1969) proposes several sampling schemes that partially achieve these desired properties by exerting controls over the joint probabilities that influence the variance estimator. Following his work, Nigam, Kumar, and Gupta (1984) present a selection method that satisfactorily provides these properties, but their method involves considerable trial and error in the construction of binary incomplete block designs. We introduce several new sampling schemes based on nonlinear programming. These methods are easily implemented and not only assure non-negativity of the variance estimate but they are also flexible with respect to stability of the variance estimator. Some of them are closely related to the approach developed by Kim, Heeringa, and Solenberger (2003). We demonstrate the usefulness and practicability of the new methods by applying them to example problems from the literature.

Jackknife Variance Estimation for Two Samples after Imputation under Two-phase Sampling

◆ Jong-Min Kim, University of Minnesota, Morris; Jon E. Anderson, University of Minnesota, Morris

University of Minnesota, Morris, 2380 Science, UMM, Morris, MN 56267

jongmink@mrs.umn.edu

Key Words: *double sampling, ratio estimation, jackknife, imputation, stratified sampling*

We propose a jackknife variance estimator for the population average from two, two-phase samples after imputation. The jackknife method has long been used to estimate and reduce bias, but has now become a valuable tool for variance estimation. We apply two different sampling methods, (simple random sampling and stratified random sampling) to derive jackknife variance estimators for the two-sample case after imputation under two-phase sampling. We also apply calibration estimation to ratio imputation in stratified random sampling.

Double Sampling in a Multistage Design

◆ David R. Judkins, Westat; Mike Hidiroglou, Office of National Statistics

Westat, 1650 Research Blvd., Rockville, MD 20854

davidjudkins@westat.com

Key Words: *variance estimation, PPS selection, survey design*

It often happens in a multistage design that useful auxiliary information is available on the final stage units. A common practice is to use this information in some sort of systematic PPS selection within the penultimate stage units, either drawing a new random start for each penultimate-stage unit or carrying the leftover to determine the starting point in the next unit. It is known that carry-over introduces some biases in the variance estimates, but many place more importance on variance minimization than on unbiased or consistent variance estimation. This emphasis may be most common when the final-stage strata correspond to analytic domains and there are required precision levels for each domain. In such cases, it is not uncommon to use systematic PPS selection where the primary sort key is the final stage stratum and the earlier stages of sampling are completely ignored. In this case, both the impact on true variances and on the bias of estimated variances can be quite large. We propose a sampling method that satisfies constraints on sample sizes by final-stage stratum, keeps variances small, and admits unbiased variance estimates for linear statistics and consistent variance estimates for nonlinear statistics. The variance estimator proposed for use with this sampling method is new.

The Variance of the Variance Estimator When the Underlying Population Has a Known Distribution

◆ Eungchun Cho, Kentucky State University

Kentucky State University, 400 E. Main St., Frankfort, KY 40601

ecpcho@alumni.rutgers.edu

Key Words: *variance estimator, complex design, sample design, Bernoulli numbers*

It is often important to evaluate the variance of the variance estimator used with complex sample designs. For example, one may seek to determine before data collection whether a given variance estimator will have an unreasonably large variance, thus may be unsatisfactory for use in inference. The standard whole sample approach and Bernoulli numbers are used to directly derive the formulas for the variance of the variance estimators when the distribution of the population is known.

Covariance Estimates in Stratified and Multistage Clustered Sampling

Andrew Vogt, Georgetown University; ◆ Dhirendra Ghosh, Synectics for Management Decisions, Inc.

Synectics for Management Decisions, Inc., 1901 North Moore St., Suite 900, Arlington, VA 22209

dhiren@smdi.com

Key Words: *covariance estimates from clusters, partial regression coefficients, intracluster correlation*

Important parameters, such as regression and correlation coefficients and partial regression and correlation coefficients, are mathematical functions of covariances. Estimates for these are derived in the case of a sample survey of a stratified clustered population where a sample of clusters is chosen from each stratum by simple random sampling or with probability proportional to size, and a simple random sample of elements is chosen from these clusters. Each estimate of covariance has an intracluster component and an intercluster component and is unbiased. The associated mathematical functions involve operations of multiplication, division, and square roots, and their estimators are in general biased but consistent. The effect of alternative sampling designs on the estimates is studied.

390 COPSS Awards and Fisher Lecture Session

Cmte of Presidents of Statistics Societies (COPSS), ASA, ENAR, IMS, SSC, WNAR

Wednesday, August 11, 4:00 pm–5:50 pm

Announcement of Awards and Introduction of Fisher Lecture Speaker

◆ Linda J. Young, University of Florida

University of Florida, Statistics Dept., College of Medicine, Gainesville, FL 32610-0212

LYoung@biostat.ufl.edu

4:00 pm Announcement of Awards

- Florence Nightingale David Award
- Elizabeth Scott Award
- COPSS Presidents' Award

4:15 pm Introduction of Fisher Lecture Speaker

Causal Inference Using Potential Outcomes: Design, Modeling, Decisions

◆ Donald B. Rubin, Harvard University

rubin@stat.harvard.edu

Causal effects are defined as the comparison of potential outcomes under different treatments on a common set of units. Observed values of the potential outcomes are revealed by the assignment mechanism—a probabilistic model for the treatment each unit receives as a function of covariates and potential outcomes. Fisher made tremendous contributions to causal inference through his work on the design of randomized field experiments, but the potential outcomes perspective applies to other complex experiments and nonrandomized studies as well. As noted by Kempthorne in his discussion of Savage's Fisher lecture, Fisher never bridged his work on experimental design and his work on parametric modeling, a bridge that appears nearly automatic with an appropriate view of the potential outcomes framework, where the potential outcomes and covariates are given a Bayesian distribution to complete the model specification. Also, this framework crisply separates scientific inference for causal effects and decisions based on such inference, a distinction evident in Fisher's discussion of tests of significance versus tests in an accept/reject context.

391 The Future of Methods and Software for Analyzing Complex Samples ▲

Social Statistics Section, Business and Economics Statistics Section, Section on Survey Research Methods, Section on Statistical Graphics, SSC, Section on Government Statistics, Section on Statistical Computing

Thursday, August 12, 8:30 am–10:20 am

Development of Statistical Methods to Analyze Complex Health Surveys for Epidemiologic Studies: Some New Methods and Applications

◆ Barry I. Graubard, National Cancer Institute; Thomas R. Fears, National Cancer Institute; Steven B. Cohen, Agency for Healthcare Research and Quality; Lester R. Curtin, Centers for Disease Control and Prevention

National Cancer Institute, Biostatistics Branch, 6120 Executive Blvd., Rm. 8024, Rockville, MD 20892

graubarb@mail.nih.gov

Key Words: *complex survey, epidemiology, population attributable risk, absolute risk*

Over the past 20 years there has been great interest in using surveys to study health and behavioral risk factors for disease. To accomplish this research, surveys have been used to select population controls and cases for case control studies, for constructing cohorts by following up survey participants, and to conduct cross-sectional studies. With this expanded use of surveys in medical research, statistical methods used in epidemiology have been adapted to account for the complexities of the sample weighting and sample selection of demographic and institutional surveys. We briefly trace the development of statistical methods for analyzing survey data for epidemiological research and discuss recent work of using surveys to estimate absolute risk and population attributable risk and their standard errors. Methods based on influence function theory will used to show how to easily derive Taylor linearized standard errors for these risk estimates. The methods will be illustrated using data from the third National Health and Nutritional Examination Survey and the National Health Interview Survey.

Design-based Methods for Survey Data: Alternative Uses of Estimating Functions

◆ David A. Binder, Statistics Canada; Milorad S. Kovacevic, Statistics Canada; Georgia R. Roberts, Statistics Canada

Statistics Canada, Methodology Branch, RH Coats Bldg., Ottawa, ON K1A 0T6 Canada

binddav@statcan.ca

Key Words: *design-based variances, confidence intervals, linearization methods*

The appropriateness of using design-based methods to analyze survey data is now well recognized. Recent research has shown that design-based methods offer some protection against model misspecification and informative sample designs. The popularity of design-based methods has grown, now that software has been developed to make such methods more accessible to data analysts. Choices for estimating the design-based variances of estimated model parameters include linearization, balanced repeated replication, jackknifing, and bootstrapping. However, the use of survey bootstraps suffers from some of the same deficiencies as the standard bootstrap; in particular, the estimated variances can be unstable in certain circumstances. Recently, methods have been developed for making inferences using the estimating function bootstrap in a model-based setting. This approach seems provide more stable results. We adapt these methods to create a design-based estimating function survey bootstrap (EFSB). We study the EFSB and a linearized version of the EFSB. Results using real survey data collected by Statistics Canada are given.

Sample Survey Methods and Software for Non-survey Data Applications

◆ Gary G. Koch, University of North Carolina, Chapel Hill; Lisa M. LaVange, Inspire Pharmaceuticals, Inc.

University of North Carolina, Chapel Hill, Dept. of Biostatistics, CB 7420, Chapel Hill, NC 27599-7420

bcl@bios.unc.edu

Key Words: *sample survey design, clustering, estimating equations, randomization, intraclass correlation, multiple regression*

Survey data analysis approaches developed to accommodate clustering and unequal weighting in the sample design have more general application to clinical studies in which repeated measures structures are encountered, including multivisit studies, multiperiod cross-over studies, multicenter studies, and epidemiological studies of repeated illnesses. Multiple regression methods that allow for the specification of a primary sampling unit within which the repeated measures have intraclass correlation are available for logistic models, proportional odds models, and linear models. Survey methods for the analysis of clustered time-to-event measurements are applicable to clinical trials of survival data. More recently developed methods for nonparametric ANCOVA, motivated by sample survey techniques, have utility in randomized clinical trials, particularly in the presence of clustered randomization. A brief review of methods and software is provided, followed by examples from clinical studies. Areas of future research as well as the need for further software development in order to make these procedures more widely available to the nonsurvey audience will also be discussed.

392 Aberration Detection Methods for Public Health Surveillance Systems ▲

Section on Statistical Graphics, Section on Government Statistics, Social Statistics Section

Thursday, August 12, 8:30 am–10:20 am

Syndromic Surveillance for Post-vaccination Adverse Events Using SPRT: A Retrospective Analysis

◆ Margarette S. Kolczak, Centers for Disease Control and Prevention; Robert Davis, University of Washington; Edwin Lewis, Kaiser Permanente of Northern California; Michael J. Goodman, HealthPartners Research Foundation; Richard Platt, Harvard Medical School and Harvard Pilgrim Healthcare; Robert T. Chen, Centers for Disease Control and Prevention; David Shay, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 1600 Clifton Rd., Mailstop E-61, Atlanta, GA 30333

mfk9@cdc.gov

Key Words: *sequential probability ratio test, risk-adjustment, vaccination*

Methods are needed to quickly detect adverse events following the introduction of new vaccines. Here we propose a method that uses sequential probability ratio test (SPRT) charts. To evaluate this method, we use retrospective data from the Vaccine Safety Datalink (VSD) to look for expected changes in adverse events following the introduction of the rotavirus and diphtheria-tetanus-acellular pertussis (DTaP) vaccines. To construct a SPRT chart, a test must be formulated with specific values for the probability of a post-vaccination event under a null hypothesis (p0) and under an alternative hypothesis (p1). To test these hypotheses, we created weekly cohorts of vaccinated children. The test compares the hypothesized probability of an adverse event in the surveillance period which follows the introduction of a new vaccine to the probability obtained using a comparison vaccination in the baseline period before the introduction of the new vaccine. To control for potential confounders such as HMO, age, calendar time, season, and gender, we extended the SPRT by using the risk-adjustment methods described by Steiner et. al.

Using Simulation Methods to Assess the Validity of Aberration Detection Methods for Public Health Surveillance of Bioterrorism-associated Events

◆ Lori Hutwagner, Centers for Disease Control and Prevention; Timothy Browne, Centers for Disease Control and Prevention; Howard Burkom, Johns Hopkins University; Matthew Seeman, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 1600 Clifton Rd., Mail Stop C-18, Atlanta, GA 30333

Lhutwagner@cdc.gov

Key Words: *simulations, aberration detection, biosurveillance*

Empirical comparisons of aberration detection methods have been done. Simulated data are need for comparison of aberration detection methods based on actual sensitivity, specificity and time to detection. Datasets were generated using a negative binomial distribution based on the means and standard deviations from national, local, and syndromic surveillance. A given dataset includes daily data for six years with 1,000 iterations. A total of 10 types of outbreaks were randomly inserted though out the data. Eight of the 10 types of outbreaks are based on two types of outbreaks using two different means and two and three standard deviations. The other two types of outbreaks are spikes with the counts inflated by two and three standard deviations. A total of 55 datasets were generated. The simulated datasets are representative of actual data with varying seasonality, trends and frequencies. The outbreaks are smaller than and not as dramatic as some natural outbreaks appear. The outbreaks were left small so that they posed a challenge to all algorithms. Early detection of the outbreaks is of more interest.

Influenza Aberration Detection Models for Eight Large Cities in the United States

◆ William W. Thompson, Centers for Disease Control and Prevention; Eric Weintraub, Centers for Disease Control and Prevention; David Shay, Centers for Disease Control and Prevention; Lynnette Brammer, Centers for Disease Control and Prevention; Nancy Cox, Centers for Disease Control and Prevention; Keiji Fukuda, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 1600 Clifton Rd., NE, MS-E61, Atlanta, GA 30333

wct2@cdc.gov

Key Words: *aberration detection, Poisson regression*

We applied Poisson regression models to data obtained from several cities in the U.S. to identify daily aberrations in deaths that were associated with influenza. From the 1990 through 2000, we obtained mortality data from the National Center for Health Statistics, influenza data from the World Health Organization, and temperature data from the National Oceanic and Atmosphere Association. Poisson regression models were used to identify influenza-associated aberrations in deaths by city. After controlling for temperature, there was still significant variation in influenza parameter estimates between cities. Influenza was associated with 4-14% of pneumonia and influenza deaths, 3-5% of respiratory and circulatory deaths, and 1-3% of all-cause deaths. In addition to temperature, other city-level confounders were examined to assess the sensitivity and specificity of influenza aberration detection models by city. Future models will also assess interactions between climate factors and influenza.

Performance Assessment for Biosurveillance Alerting Algorithms

◆ Howard Burkom, Johns Hopkins University

Johns Hopkins University, 11100 Johns Hopkins Rd., Applied Physics Laboratory, Laurel, MD 20723-6099

Howard.Burkom@jhuapl.edu

Key Words: *alerting algorithms, biosurveillance, receiver operating characteristic, activity monitor operating characteristic, outbreak detection*

The objective of alerting algorithms in biosurveillance is to assist in the timely recognition of disease outbreaks by detecting their early effects on monitored data streams. These algorithms are applied both to clinical data such as physician office visits and to nonclinical data such as over-the-counter sales. The uncertainty in the effects of outbreaks on these datasets makes it difficult to evaluate detection methods. This talk presents a methodology to assess the utility of alerting algorithms using authentic background data and simulated outbreak effects based on theoretical epicurves of primary cases. For practical expected false alert rates, receiver operating characteristic (ROC) methods are used to measure the sensitivity of these algorithms, and activity monitor operating characteristic (AMOC) methods are used to measure their timeliness. The approach will be illustrated with methods of the Early Aberration Reporting System (EARS) developed at the U.S. Centers for Disease Control and Prevention. The detection performance of these algorithms will be analyzed for various data backgrounds and plausible signal effects. Optimal threshold selection will be discussed.

393 Dynamic Survival Models ⌘

ENAR, Section on Statistics in Epidemiology, WNAR

Thursday, August 12, 8:30 am–10:20 am

Model Selection for Dynamic Survival Analysis

◆ Ian W. McKeague, Florida State University

Florida State University, Dept. of Statistics, Tallahassee, FL 32306-4330

mckeague@stat.fsu.edu

Key Words: *hazard function regression, censored data, additive risk model, covariate selection, nonproportional hazards*

Model selection methods are well developed in parametric settings, and in recent years they have been extended to wide classes of nonparametric models. For dynamic survival models which involve complex time-dependent covariates and coefficients, however, generally applicable and fully validated procedures are not yet available. This talk discusses a new approach that applies to a flexible class of nonproportional hazard function regression models in which the influence of the covariates splits into the sum of a parametric part and a time-dependent nonparametric part. The approach allows covariate selection for the parametric part by adjusting for the implicit fitting of the nonparametric part. Asymptotic consistency is established, leading to asymptotically normal estimators of both parametric and nonparametric parts of the model in the presence of covariate selection. The approach is illustrated using real and simulated data.

Survival Models with Multiple Time Scales

◆ Thomas Scheike, University of Copenhagen

ts@biostat.ku.dk

Key Words: *survival models, counting process, additive risk, multiple time scales*

I present a nonparametric survival model with two time scales. The time scales are equivalent up to a constant that varies over the subjects. Covariate effects are modeled linearly on each time scale by additive Aalen models. Estimators of the cumulative intensities on the two time scales are suggested by solving approximate local maximum likelihood estimating equations. The local estimating equations necessitates only the choice of one bandwidth. The estimators are provided with large sample properties. The model is applied to data on patients with myocardial infarction, and used to describe the prognostic effect of covariates on the two time scales.

Functional Association Models for Multivariate Survival Processes

◆ Jun Yan, University of Iowa; Jason Fine, University of Wisconsin, Madison

University of Iowa, 241 Schaeffer Hall, Iowa City, IA 52242

jyan@stat.uiowa.edu

Key Words: *empirical process, functional estimating equation, local dependence, partially observed, uniform convergence, varying-coefficient*

We consider multivariate temporal processes that are continuously observed within overlapping time windows. The intended application is censored multistate and multivariate survival settings, where point processes are continuously observed. Functional mean and association regression models are studied for the point processes, with completely unspecified time-varying coefficients. The continuous observation scheme is exploited: the coefficients may be estimated nonparametrically by extending GEE to continuously observed data. The estimators automatically converge at the parametric rate, without smoothing, unlike with discretely observed data. Uniform consistency and weak convergence is established with empirical process techniques. Existing functional approaches to survival processes utilize intensity models, which require smoothing and depend critically on the choice of smoothing parameters, similarly to discretely observed data. The nonparametric estimators yield new tests for covariate effects, parametric submodeling of these effects, and goodness-of-fit testing. Simulation studies and an analysis of familial aggregation of alcoholism illustrate the methodology's utility.

394 Approaches in the Design and Analysis of Pharmacogenomics and Pharmacogenetics Studies ▲ ▨

Biopharmaceutical Section, Section on Physical and Engineering Sciences

Thursday, August 12, 8:30 am–10:20 am

Effectively Developing and Evaluating Medical Diagnostics Based on Whole Genome Technologies

◆ Richard Simon, National Institutes of Health

National Institutes of Health, National Cancer Institute, 9000 Rockville Pike, MSC #7434, Bethesda, MD 20892-7434

rsimon@nih.gov

Key Words: *microarrays, pharmacogenomics, diagnostic, classification*

Whole genome technologies, such as RNA transcript profiling, SNP genotyping, and proteomics spectral characterization, are providing new tools which will have major influence on the future of medicine. The effective utilization of these technologies requires inter-disciplinary collaboration among clinical investigators, basic scientists, and statisticians. Today there are many growing pains in the use of these technologies and substantial misinformation about the design and analysis of related studies.These problems threaten to seriously delay the utilization of whole genome technologies for improving public health. Misunderstanding of the principles for development of medical diagnostics based on whole genome technologies can result in inadequate studies and inappropriate regulatory policies. I will present an overview of the use of DNA microarray technology for predicting response to treatment. I will indicate how microarray technology has the potential to rapidly expedite the development of such diagnostic tests and identify the essential components in the design and analysis of studies for development and evaluation of microarray-based diagnostics.

Design and Analysis Issues in Pharmaco-genetics/-genomics Studies

◆ John S. Witte, University of California, San Francisco

University of California, San Francisco, Dept. of Epidemiology and Biostatistics, 500 Parnasus, MU - 420 West, San Francisco, CA 94143-0560

jwitte@itsa.ucsf.edu

Key Words: *clinical trials, pharmacogeneitcs, pharmacogenomics, sample size, power*

Incorporating information on genetic variants that impact disease susceptibility or drug efficacy may reduce clinical trial size and duration. Moreover, knowledge about genes that result in adverse drug reactions can help focus drug development and prescription. We first discuss the different study designs one can use to incorporate genetic information into trials. In particular, whether genotypes are screened prospectively or retrospectively, and whether one takes a candidate gene versus a genome-wide approach. Then we quantify the effects of including genetic information on the sample sizes and time required to detect differences between trial arms. Our calculations indicate that, depending on allele frequency and gene action, pharmacogenetic clinical trials can yield significant sampling/cost savings over traditional trials. Finally, we show how these benefits will depend on numerous factors, including ease of patient recruitment, cost of genotyping, long-term costs of study, and the magnitude of the effect.

A Multiple-testing Strategy for Analyzing cDNA Array Data on Gene Expression

◆ Ralph L. Kodell, U.S. Food and Drug Administration; Robert R. Delongchamp, U.S. Food and Drug Administration; John F. Bowyer, U.S. Food and Drug Administration; James J. Chen, National Center for Toxicological Research

U.S. Food and Drug Administration, NCTR, Division of Biometry and Risk Assessment, 3900 NCTR Rd., Jefferson, AR 72079

rkodell@nctr.fda.gov

Key Words: *subset selection, p value plot, false discovery rate, false nondiscovery rate, ROC curve, optimal cutoff*

CDNA arrays interrogate tissue samples for the levels of mRNA for hundreds to tens of thousands of genes to measure treatment-induced changes in gene expression. Selecting a significance level for by-gene hypothesis tests requires dealing with the multitude of treatment contrasts. The p values from these tests order the genes such that a p value cutoff divides the genes into two sets. The set of genes selected as affected will have false positives while the set selected as unaffected will contain false negatives. A p value plot allows one to estimate the number of true null hypotheses (truly unaffected genes). With this estimate, the false-positive and false-negative rates associated with any p value cutoff can be estimated. An optimal cutoff depends upon the relative cost of falsely classifying a gene as affected versus the cost of falsely classifying a gene as unaffected. Here, a method analogous to methods developed for ROC curves is proposed for selecting the cutoff. The false discovery rate (FDR) and false nondiscovery rate (FNR) associated with the cutoff are then estimated. Two functional genomics studies are used for illustration.

395 Recent Statistical Issues in Cancer Research ◬ ⌘

International Indian Statistical Association

Thursday, August 12, 8:30 am–10:20 am

The Value of Long-term Follow-up: Lessons from the National Wilms Tumor Study

◆ Norman E. Breslow, University of Washington

University of Washington, Dept. of Biostatistics, Box 357232, Seattle, WA 98195-7232

norm@u.washington.edu

Key Words: *molecular genetics, late effects, childhood cancer, continuity in data collection*

Wilms tumor is an embryonal tumor of the kidney that affects approximately one child in every 10,000. During the 20th century, cure rates increased from 10% to 90% as first radiation and then chemotherapy joined surgical removal of the diseased kidney as standard treatment. The National Wilms Tumor Study Group (NWTSG), etablished in 1969, conducted five protocol studies and registered nearly 10,000 patients before terminating patient accrual in 2002. During its last 15 years of operation, the study enrolled 70-80% of the 550 cases estimated to occur annually in North America. Its focus has been the identification of patient subgroups at high or low risk of relapse, and the substitution of combination chemotherapy for radiation therapy, with a primary goal to reduce long-term complications while producing the maximum number of cures. The NWTSG Data and Statistical Center, located in Seattle since the start of the study, has played a major role in this effort. This talk will describe how statistical analyses of the collected data have helped to answer several questions related to the molecular genetics of Wilms tumor.

On the Use of Biomarkers for Carcinogen Risk Assessment

◆ Louise M. Ryan, Harvard School of Public Health; Sally W. Thurston, University of Rochester; Wenzheng Huang, Millennium Operations

Harvard School of Public Health, Dept. Biostatistics, 655 Huntington Ave., Boston, MA 02115

lryan@hsph.harvard.edu

Key Words: *surrogate marker, benchmark dose, dose response*

There is growing interest in the use of biomarker data in environmental health research, and in particular for helping to quantify dose response relationships. We will discuss some of the challenging statistical issues that arise in this context. We specify a modeling framework that links environmental exposure, biomarkers, and outcome, and discuss in conceptual terms how such a formulation could be used to inform dose response modeling for the purpose of quantitative risk assessment. We analyze data from a case control study designed to elucidate the mechanisms of smoking-induced lung cancer. Because of sample size limitations, we use a likelihood-based analysis which subsumes both cohort and case-control designs as special cases. Our analysis allows us to (1) investigate the extent to which the markers explain the pathway from exposure to outcome; (2) quantify the degree to which biomarker data can improve on predicting outcome over and above exposure; and (3) estimate the association among multiple markers.

Models and Uses of Absolute Risk in Guiding Clinical Management

◆ Mitchell H. Gail, National Cancer Institute

National Cancer Institute, 6120 Executive Blvd., Room 8032, Bethesda, MD 20892-7244

gailm@mail.nih.gov

Key Words: *absolute risk, crude risk, clinical risk and benefit, calibration, concordance statistic, discriminatory power*

Absolute risk is the probability that a subject of age "a" with particular risk factors will develop the disease of interest in a defined age interval, say (a,a+b). Absolute risk is a "crude" risk of disease in the presence of competing risks. We review study designs and analytical methods used to estimate absolute risk, methods to determine if absolute risk models are well-calibrated, and methods to assess how well absolute risk models can predict who will and who will not develop disease. We consider the decision of whether or not to take tamoxifen to prevent breast cancer as an example of the use of models of absolute risk to guide clinical decision-making, and we highlight some areas in need of further methodologic and clinical epidemiologic research.

Genetic Association Studies Using Haplotypes

◆ Glen Satten, Centers for Disease Control and Prevention; Michael P. Epstein, Emory University; Andrew S. Allen, Duke University

Centers for Disease Control and Prevention, 4770 Buford Hwy., Mail Stop F-24, Chamblee, GA 30341

gsatten@cdc.gov

Current practice in association studies of genetic causes of diseases like cancer is to genotype several single nucleotide polymorphisms (SNPs) in a candidate gene. Such data can either be analyzed considering each SNP as a separate predictor, or by considering the effect of haplotypes (the set of SNPs that occur together on the same chromosome). Analyzing haplotypes can have several advantages: they may reduce dimension if there are fewer haplotypes than are combinatorially possible, and because haplotypes reflect the sequence actually transcribed, a haplotype analysis may be more powerful. However, haplotypes are typically unmeasured, and must be inferred statistically. Thus, the problem of analyzing the association between haplotypes and disease outcomes is a type of missing data problem. We consider a variety of new methods for haplotype analysis in association studies, both in case control studies and family-based association studies.

396 CAUSEweb: A Digital Library for Statistics Instructors

Section on Statistical Education, Section on Quality and Productivity, Advisory Cmte on Teacher Enhancement

Thursday, August 12, 8:30 am–10:20 am

CAUSEweb: A Digital Library for Statistics Instructors

◆ Dennis K. Pearl, The Ohio State University; ◆ Roger Woodard, North Carolina State University; ◆ Gerard L. Hanley, California State University; ◆ Ginger H. Rowell, Middle Tennessee State University

Middle Tennessee State University, 223 Kirksey Old Main, Mathematical Sciences Dept., Murfreesboro, TN 37132

rowell@mtsu.edu

This session will highlight CAUSEweb, a new digital library for statistics instructors operated by the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) through funding from the National Science Foundation. Arising from a strategic initiative of the ASA, CAUSE is a national organization whose mission is to support and advance undergraduate statistics education in four target areas: resources, professional development, outreach, and research. The director of CAUSE, Dennis Pearl, will provide an overview of the group's initiatives in each of these four areas. The CAUSEweb editor, Roger Woodard, will describe the nature of the collection and the editorial process for the reviewed portion of the library. Gerard Hanley, the executive director of the MERLOT project (see MERLOT.org) will describe how CAUSEweb fits into the broader context of MERLOT and the rapidly developing world of digital libraries. Ginger Holmes Rowell will then describe her experiences with using CAUSEweb materials in the classroom. The session will conclude with discussion from the floor.

397 Bayesian Methods in Health Service Research ▲

Section on Health Policy Statistics, Section on Bayesian Statistical Science

Thursday, August 12, 8:30 am–10:20 am

The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery

◆ Mary Beth Landrum, Harvard Medical School

Harvard Medical School, 180 Longwood, Boston, MA 02110

landrum@hcp.med.harvard.edu

Provider profiling has been urged by many as a fundamental step in medical care reform. Despite the building momentum for this strategy, the underlying premise of profiling—that disseminating information about provider quality will lead to improved quality of care—has yet to be established. We undertake a new approach aimed at resolving this fundamental controversy. Using a longitudinal data set containing all individuals who underwent CABG in a hospital in New York State over a nine-year period, we consider how report cards affect the behavior of individual providers, specifically providers who are publicly identified as being significantly better or worse than their peers. We use Bayesian longitudinal hierarchical regression models to examine patterns in volume, average severity, and mortality following surgery, adapting models previously used to profile providers using longitudinal data. Our analysis addresses three intended goals of provider profiling.

Analysis of Incomplete HRQoL Data in the REMATCH Trial

◆ Huiling Li, Columbia University; Daniel F. Heitjan, University of Pennsylvania

Columbia University, 100 Haven Ave., Apt 26G, New York, NY 10032

HL2034@columbia.edu

Key Words: *Bayesian inference, clinical trials, importance sampling, multiple imputation, pattern-mixture model*

A pattern-mixture model is applied for modeling the joint distribution of incomplete repeated measurements of quality of life and right-censored survival times in the evaluation of how a treatment affects an individual's well-being over time. The statistical model assumes that the survival times follow a multinomial distribution and that quality-of-life outcomes follow a multivariate normal distribution conditional on the survival time. The model is estimated using a Bayesian approach by importance sampling, and then used to create multiple imputation of the missing outcomes. The methods are illustrated through the analysis of the data from a randomized clinical trial of a treatment for cardiovascular disease.

Bayesian Factor Analysis for Multilevel Survey Data with Structured Nonresponse

◆ Alistair J. O'Malley, Harvard Medical School; Alan Zaslavsky, Harvard Medical School

Harvard Medical School, Dept. of Health Care Policy, 180 Longwood Ave., Boston, MA 02155-5899

omalley@hcp.med.harvard.edu

Key Words: *Bayesian, factor analysis, multilevel modeling, multivariate, nonresponse, survey data*

Health care quality surveys in the U.S. are administered to individual respondents (hospital patients, health plan members) to evaluate performance of health care units (hospitals, health plans). Due to both planned item nonresponse (caused by screener items and associated skip patterns) and unplanned nonresponse, quality measures, such as item means, are based on different subsets of the survey respondents. For better understanding and more parsimonious reporting of dimensions of quality, we analyze relationships between quality measures at the unit level, by applying techniques such as factor analysis to covariance structure estimated at the unit level in a hierarchical model. At the lower (patient) level we first fit generalized variance-covariance functions that take into account the nonresponse patterns in the survey responses. A between unit covariance matrix is then estimated using a hierarchical model, which evaluates the fitted generalized variance-covariance functions to account for sampling variation. Bayesian methods are used for model-fitting. An important advantage of this approach is that it allows inferences about the number of factors, the factor loadings, and more.

Bayesian Estimation of the Costs of Optimal Allocation of a Surgical Procedure

◆ Marshall M. Joffe, University of Pennsylvania; Michael R. Elliott, University of Pennsylvania School of Medicine

University of Pennsylvania, 602 Blockley Hall, 423 Guardian Dr., Philadelphia, PA 11367

mjoffe@cceb.upenn.edu

Key Words: *Bayesian statistics, causality, classification*

Subjects with end-stage renal disease requiring hemodialysis need a way to access their vascular system. When successful, the construction of a naïve arteriovenous fistula (AVF) is the best way to do this; unfortunately, the procedure is often unsuccessful. It is thus important to determine when the procedure is likely to be successful. One can use information on costs to choose an optimal rule for deciding when to perform the procedure. We consider estimation of the costs of different decision rules and contrast Bayesian with frequentist approaches when the same data are used for choosing the rule and estimating the costs. Naïve application of frequentist methods results in overoptimism about the costs of the selected rule; corrections are required. These corrections are not generally required in Bayesian methods. We illustrate these points using data from a study of surgery to create an AVF.

Modeling Count Data in Health Care Utilization with Endogeneity

◆ Zhen Chen, University of Pennsylvania

University of Pennsylvania, 423 Guardian Dr., 625 Blockley Hall, Philadelphia, PA 19104

zchen@cceb.upenn.edu

Key Words: *negative binomial, simultaneous equations, Poisson models, random effects, MCMC, moral hazads*

In modeling count data in healthcare utilization, it is usually the case that the regressor of primary interest is endogenously determined. Moreover, restrictions in data collection process may result in missing data problem. For instance, patients who seek no treatments have missingness in copayment variable if copayment information is obtained from pharmacy records. We propose a joint model that uses a truncated negative binomial regression for the count outcome given the endogenous regressor, and a proportional odds model for the categorized endogenous regressor given other exogenous variables. We implement an MCMC procedure to estimate model parameters and apply the method to a study that investigates copayment effects on number of drug prescriptions in patients with depression disorders.

398 Spatial Statistics for Geophysical and Environmental Process ▲

Section on Statistics and the Environment

Thursday, August 12, 8:30 am–10:20 am

Multiresponse Smoothing and Spatial Models

◆ Stephan R. Sain, University of Colorado, Denver

University of Colorado, Denver, Dept. of Mathematics, Campus Box 170, PO Box 173364, Denver, CO 80217-3364

ssain@math.cudenver.edu

Many investigations in the environmental and biological sciences encounter data that is inherently multivariate, exhibits complex structure, and has various forms of dependence, in particular spatial dependence. A strategy is presented for multiple-response regression problems that uses the connection between spatial models (universal kriging) and thin-plate splines to model nonlinear relationships between explanatory and response variables via a type of smoothing. Furthermore, this strategy is reminiscent of semiparametric mixed models. In addition, spatial dependence can also be easily be incorporated. The methodology will be demonstrated using several examples, including the modeling of soil water profiles (pedotransfer functions) for use with crop models to examine sources of variation in crop yields and the modeling of skin cancer rates and climate variables.

Winds from Parallel Bayesian Hierarchical Models: Introducing the Gibbs Coupler

◆ Timothy Hoar, National Center for Atmospheric Research; Doug Nychka, National Center for Atmospheric Research; Ralph F. Milliff, Colorado Research Associates

thoar@ucar.edu

Key Words: *parallel computing, spatio-temporal modelling, Gibbs Sampler*

Advances in computing power are allowing researchers to use Bayesian Hierarchical Models (BHMs) on problems previously considered computationally infeasible. The BHM in question is estimated in a massive Gibbs Sampler and combines the information from a scatterometer on board a polar-orbiting satellite and the result of a numerical weather prediction model to produce an ensemble of high-resolution tropical surface wind fields with physically realistic variability at all spatial scales. Since the model is only valid for a time period (epoch) much shorter than the data record, a separate model for each epoch is estimated. However, each model has an autoregressive component which connects the current time with the previous time. If the calculations are kept separate, the full conditional distributions for the states at the endpoints are not conditioned on states outside the epoch. This presentation discusses the procedure of simultaneously estimating multiple BHMs on a massively parallel platform while still providing a mechanism for maintaining proper distributions for the states at the endpoints.

Spatial Hierarchical Bayes Model for AOGCM Climate Projections

◆ Reinhard Furrer, University Corporation for Atmospheric Research; Stephan R. Sain, University of Colorado, Denver; Tom M.L. Wigley, University Corporation for Atmospheric Research; Doug Nychka, National Center for Atmospheric Research

University Corporation for Atmospheric Research, GSP, Boulder, CO 80307-3000

furrer@ucar.edu

Key Words: *hierarchical Bayes, climate change, Gibbs sampler, spatial processes, general circulation models*

Numerical experiments based on atmospheric-ocean general circulation models (AOGCMs) are one of the primary tools in deriving projections for future climate change. However, each model has its strengths and weaknesses within local and global scales. This motivates climate projections synthesized from results of several AOGCMs' output weighted according to model bias and convergence. We combine present day observations, present day, and future climate projections in a single hierarchical Bayes model. The challenging aspect is the modeling of a meaningful covariance structure of the spatial processes. We propose several approaches. The posterior distributions are obtained with computer-intensive MCMC simulations. The novelty of our approach is that we use gridded, high-resolution data within a spatial framework. The primary data source is provided by the MAGICC/SCENGEN program and consists of 17 AOGCMs on a 5x5 degree grid under several different emission scenarios. We consider variables such as the precipitation, temperature, and min/max thereof. Extensions such as a multivariate approach and heavy-tailed error distributions are discussed.

Spatial Precipitation Extremes

◆ Ulrike Schneider, University Corporation for Atmospheric Research

University Corporation for Atmospheric Research, P.O. Box 3000, Geophysical Statistics Project, Boulder, CO 80307-3000

uli@ucar.edu

Key Words: *spatial statistics, extreme value theory, Bayesian statistics*

We consider (hourly) rain gauge records collected over various stations throughout Colorado and investigate this dataset with regard to its extreme behavior and spatial dependence due to the unique topography and meterological properties of this region. To build the statistical model, we use the Generalized Pareto Distribution (GPD) within a threshold model, along with a Bayesian approach to incorporate the spatial structure of the different stations. We employ Markov chain Monte Carlo techniques to obtain parameter and uncertainty estimates. We also discuss how to model the apparent seasonal effects of precipitation.

Robust Prediction for Contaminated Random Field

◆ Baptiste Fournier, Swiss Federal Institute of Technology

Swiss Federal Institute of Technology, Institute of Mathematics, Chair of Applied Statistics, Lausanne, 1012 Switzerland

baptiste.fournier@epfl.ch

Key Words: *spatial statistics, substitutive outliers, robustness, prediction*

The prediction problem in the presence of outliers is a general problem in spatial statistics. Methods to robustify the ordinary kriging exist, but they usually treat the case of additive outliers. Here, we deal with substitutive outliers. With a certain probability, each site independently exhibits an outlier with the expectation of the nominal process, but with larger variance. We derive the conditional expectation of the true process given the observed one. This turns out to be a sum of predictors, one for each scenario of outliers locations. The number of scenari is equal to two to the number of sites in the field. This is hopelessly complex for a grid of reasonable size. Therefore, the space of contamination scenari will be restricted. The way to proceed consists in two steps. First, detect the contaminated locations. Second, consider a neighborhood of the scenario found in the first step. The more we increase the size of the neighborhood, the better is the quality of the estimation with the drawback that the computational cost increases. We will present the theoretical background of the described method and also show some simulation results illustrating its performance.

399 Nonresponse Issues in Panel Surveys ▲

Section on Survey Research Methods, Social Statistics Section, Section on Government Statistics

Thursday, August 12, 8:30 am–10:20 am

Response Rates and Nonresponse in BLS and Census Bureau Establishment Surveys

◆ Rita J. Petroni, U.S. Census Bureau; Stephen Cohen, Bureau of Labor Statistics; Richard Sigman, U.S. Census Bureau; Clyde Tucker, Bureau of Labor Statistics; Diane Willimack, U.S. Census Bureau; Howard Hogan, U.S. Census Bureau

U.S. Census Bureau, Room 3009, Bldg. 4, Washington, DC 20233

rita.jo.petroni@census.gov

Key Words: *response rate definitions, response rate measurement, response rate trends, methods to encourage response*

The Bureau of Labor Statistics and the Census Bureau grapple with defining establishment survey response rates in meaningful ways that measure response levels at various stages of the survey process and that measure the degree to which the resulting estimates summarize data from respondents plus approved alternative sources such as administrative records. Within and across organizations, evaluations of trends in meaningfully and comparatively defined rates over time can assist in pinpointing where improvements have been made and where they need to be made. This paper examines the two organizations' definitions of establishment response rates, trends in response rates, possible explanations of differences between the agencies response rates, methods to encourage response, and research on nonresponse reduction.

A Pattern-mixture Model for Panel Nonresponse in the Current Employment Statistics Survey

◆ Kennon R. Copeland, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212

copeland.kennon@bls.gov

Key Words: *late reporting, nonreporting, nonignorable nonresponse, establishment survey, estimate revisions, panel survey*

The Bureau of Labor Statistics' Current Employment Statistics (CES) survey collects employment, hours, and earnings data monthly from a sample of over 300,000 U.S. establishments. To provide timely information, preliminary estimates are generated three to four weeks after the survey reference period, using a weighted link-relative estimator. Final estimates are released two months later, incorporating data from late reporters. CES survey nonresponse in the preliminary estimates consists of late reporting and nonreporting. While both affect the overall accuracy of the CES estimates to some unknown extent, the impact of late reporting is more directly assessed by examining revisions between preliminary and final estimates. Overall accuracy can be assessed by examining annual benchmark revisions. The performance of a nonignorable nonresponse model, assuming month-to-month change varies by reporting status, will be compared to that of the current proportional regression model. In addition, link-relative estimates based on a refined estimation cell definition will be evaluated. Minimizing the magnitude of both final and benchmark revisions is a key objective.

Modeling of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview

◆ Moon Jung Cho, Bureau of Labor Statistics; John L. Eltinge, Bureau of Labor Statistics

Bureau of Labor Statistics, Office of Survey Methods Research, Room 1950, PSB, 2 Massachusetts Ave. NE, Washington, DC 20212

Cho_M@bls.gov

Key Words: *incomplete data, logistic regression, mixture distribution, nonidentified item nonresponse, parameter identification*

In the U.S. Consumer Expenditure Interview Survey (CEQ), consumer units (roughly equivalent to households) are asked to provide month-by-month reports of the amounts of money spent on each of a large number of items. Reported expenditures are recorded at a relatively fine level of detail defined by the six-digit Universal Classification Code (UCC). For a given month, most consumer units report nonzero expenditures for a relatively small proportion of the possible UCC items. When no expenditure is reported, available data do not allow one to distinguish between cases of no expenditure for this item in the specified month ("true nonexpenditure") and cases of failure to report a true nonzero expenditure ("nonidentified item nonresponse"). However, under specific models for relationships among true nonexpenditure, nonidentified item nonresponse and observable auxiliary variables, some important model parameters are estimable. This paper reviews the relevant models and available auxiliary information, discusses identifying restrictions for specific parameters, and presents related point estimators and variance estimators.

Characteristics of Survey Attrition in the Household Component of the Medical Expenditure Panel Survey

◆ David Kashihara, Agency for Healthcare Research and Quality; Trena M. Ezzati-Rice, Agency for Healthcare Research and Quality

Agency for Healthcare Research and Quality, DSRM/CFACT, 540 Gaither Rd., Rockville, MD 20850

dkashiha@ahrq.gov

Key Words: *nonresponse bias, weight adjustments, panel survey*

The household component of the Medical Expenditure Panel Survey (MEPS-HC) is designed to provide annual, nationally representative estimates of health care use, health status, health conditions, medical expenditures, sources of payment, insurance coverage, and health care access for the U.S. civilian noninstitutionalized population. Both person-level and household-level data are collected. The MEPS-HC consists of an overlapping panel design in which persons in a given panel are interviewed five times over a 30-month period. Using the most recent MEPS data, this paper examines and compares the characteristics of participants

in the initial round of data collection with those who discontinue participation in subsequent rounds. The identified variables are compared with those currently used to adjust the MEPS sampling weights for survey attrition. This study is part of ongoing efforts to ensure the accuracy of the MEPS survey estimates.

Using Propensity Score to Adjust Weights to Compensate for Nonresponse in the Medical Expenditure Panel Survey

◆ Lap-Ming Wun, Agency for Healthcare Research and Quality; Trena M. Ezzati-Rice, Agency for Healthcare Research and Quality; Robert M. Baskin, Agency for Healthcare Research and Quality; Marc W. Zodet, Agency for Healthcare Research and Quality; Janet Greenblat, Agency for Healthcare Research and Quality; Frank Potter, Mathematica Policy Research, Inc.; Nuria Diaz-Tena, Mathematica Policy Research, Inc.; Mourad Touzani, Mathematica Policy Research, Inc.

Agency for Healthcare Research and Quality, 540 Gaither Rd., Rockville, MD 20850

lwun@ahrq.gov

Key Words: *nonresponse, logistic regression, interaction, weight adjustment, propensity score*

The Medical Expenditure Panel Survey (MEPS) is sponsored by the Agency for Healthcare Research and Quality (AHRQ). MEPS, a complex national probability sample survey, is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. It comprises three component surveys with the Household Component (HC) as the core survey. The MEPS-HC, like most sample surveys, experiences unit nonresponse despite efforts to maximize response rates. Survey nonresponse is usually compensated for by some form of weighting adjustment to reduce the bias in survey estimates. Currently, a weighting class nonresponse adjustment using socioeconomic and demographic variables to create the weighting classes is used in the MEPS to adjust for potential nonresponse bias at the dwelling-unit level. An alternative method for forming nonresponse adjustment cells is to use response propensities. This paper summarizes research undertaken to investigate various potential use of response propensities to adjust weights to compensate for nonresponse in the MEPS.

400 Issues in Nonsampling Error ▲ ⌘

Section on Survey Research Methods

Thursday, August 12, 8:30 am–10:20 am

Synchronizing Survey Edit and Imputation Systems: A Look at the American Housing Survey's Utility System

◆ Mark Gorsak, U.S. Census Bureau/University of Maryland JPSM; Dennis Schwanz, U.S. Census Bureau

U.S. Census Bureau/University of Maryland JPSM, 12718 Mescalaro Lane, Lusby, MD 20657

mgorsak@chesapeake.net

Key Words: *edit, imputation, American Housing Survey*

Surveys are continuously modifying edit and imputation systems by using data from other sources. To create commonality within a survey, the application of data from one source to another needs to comprehensive. The utility system for the American Housing Surveys (AHS) uses data from the Residential Energy Consumption survey (RECs) and applies the findings from the RECs to the components of the AHS edit and imputation system. The AHS captures utility data from the respondent either by recording the monthly bills from specific months or asking for a monthly average over the past year. The AHS edit system determines an annual cost from the respondents' answers by regressing on the RECs findings. If the respondent does not supply the information, then the AHS imputation system determines an annual utility cost by matching to primary variables taken from the RECs findings. This paper looks at the development of the electric and natural gas costs in the AHS.

Comparison of Two Imputation Methods in the Survey of Doctorate Recipients

◆ Jeffrey H. Stratton, U.S. Census Bureau; John M. Finamore, U.S. Census Bureau; Todd R. Williams, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Rm. 3726-3, Washington, DC 20233-8700

jeffrey.h.stratton@census.gov

Key Words: *hot-deck imputation, Flexible Matching Imputation*

Nonresponse is a major concern in most sample surveys. Missing data are often imputed to provide users with a complete dataset. Many different methods of imputation have been proposed, and the methods used vary widely among different surveys. One of the most common techniques used to impute missing data is hot-deck imputation. Even having chosen this technique, there are many different ways to implement it. This paper examines the results of two different hot-deck imputation techniques applied to the 2001 Survey of Doctorate Recipients (SDR). The SDR is a longitudinal biennial survey sponsored by the National Science Foundation to measure the demographic characteristics of individuals who received a PhD in a science or engineering field from a U.S. institution. (Doctoral-level professional degrees such as those awarded in medicine, law, or education are not included.) The 2001 SDR included doctorate degrees earned between January 1, 1942, and June 30, 2000. We present results comparing the current hot-deck imputation method used in the SDR with a Flexible Matching Imputation (FMI) method proposed by Todd Williams of the U.S. Census Bureau. Both methods use model-based methodology to select important matching variables, but the FMI method is able to consider more variables. We use both methods to impute missing data in the 2001 SDR, and discuss their performance.

Latent Class Models for Analysis of Response Error and Rotation Group Bias in the Current Population Survey

◆ Khandaker Mansur, U.S. Census Bureau; Bac Tran, U.S. Census Bureau

U.S. Census Bureau, 4132 Henhawk Ct., Ellicott City, MD 21042-5925

khandaker.a.mansur@census.gov

Key Words: *latent class models, panel design, unemployment, rotation group bias, misclassification errors, response errors*

The Current Population Survey (CPS) uses a 4-8-4 rotating panel design. For any given month, the CPS sample can be grouped into eight subsamples corresponding to the eight rotation groups. It is well known that in a rotating panel design survey estimates for the same characteristics from different rotation groups relating to the same time period have different expected values, depending on the length of the time they have been included in the sample. This phenomenon is called rotation group bias. Latent class models are used here for the analysis of response error in the CPS. Latent class analysis (LCA) is used here to estimate the misclassification rates and to assess their impact on labor force estimates. The paper investigates the validity and efficiency of using LCA to estimate rotation group bias, response error and response probabilities. This paper studies response errors in the CPS and assesses their impact on the rotation group bias. The paper also discusses Shockey's results on rotation group bias and attempts to determine why the sizes of his rotation group bias are much larger than those reported by other authors, specifically Bailar (1975) and Mansur (1999) estimates. The paper also compares the response probabilities to those obtained from the Markov Latent Class Analysis presented in Tran (2003).

Resources vs. Coverage—How to Minimize One and Maximize the Other: A Case Study from an Education Survey

◆ Nancy R. Johnson, U.S. Census Bureau; Meagan Wilson, U.S. Census Bureau; Karen D. Deaver, U.S. Census Bureau; Stephen P. Broughman, National Center for Education Statistics

U.S. Census Bureau, 4700 Silver Hill Rd., Washington, DC 20233

nancy.r.johnson@census.gov

Key Words: *measuring quality-coverage, targeting, Private School Survey, quality reporting and quality indicators, improving process quality*

We've all been there... You design the optimal approach for survey coverage. Then the final budget comes through, or the schedule is tight, or other work demands your time, and you just can't do what you want. How do you proceed? What do you sacrifice? Can you actually plan for this? This paper will articulate the assumptions and describe the methodology and criteria for targeting specific state sources to maintain or improve the frame coverage for the Private School Survey (PSS). These criteria take into consideration the quality and timing of the state information as well as the net coverage improvement to the overall PSS at the national and subnational levels. As part of its frame development, the PSS collects data from Early Childhood Centers (ECCs). However, due to budget, scheduling, or other work competing for time or money it's not always feasible to update the frame from every state source. When this occurs, the challenge is to prioritize the processing of the state sources that take into consideration costs, schedule, and frame coverage requirements. This initiative is essential to ensuring that the quality of the estimates and total variance meets both the Census Bureau and customer expectations.

Converting from Paper to an Automated Mode for Collecting Group Quarters Data for Demographic Survey

◆ Sherry E. Thorpe, U.S. Census Bureau

U.S. Census Bureau, Demographic Statistical Methods Division, 4700 Silver Hill Rd., Washington, DC 20233

sherry.e.thorpe@census.gov

Key Words: *frame development, group quarters, nonsampling error, automated instrument, demographic surveys, sample redesign*

Group Quarters is a type of living accommodations contained in the sampling frame that supports the demographic surveys that are included in the post-2000 sample redesign. One of the critical activities that historically influenced the processing operations in the creating the GQ frame was the use of paper by the field enumerators. As part of the Census Bureau's continuing initiative to improve the frame development process, the Bureau developed an automated instrument (Group Quarters Automated Instrument for Listing, GAIL) to assist enumerators in collecting group quarters information for the demographic surveys. The GAIL was placed into production in September 2003. This paper will focus on three aspects of that conversion: the implications on the GQ frame development system with respect to types of subsystems, defining requirements, testing and system integration; historical and new sources of nonsampling error; and the implication on field enumerators with respect to training and skill requirements.

401 Special Editing and Imputation Problems for Economic Data ▲ ▦

Business and Economics Statistics Section, Section on Survey Research Methods

Thursday, August 12, 8:30 am–10:20 am

Searching for Donors: Finding an Imputation Strategy

◆ Michael Hogye, U.S. Dept. of Agriculture

U.S. Dept. of Agriculture, NASS, R&D Division, 3251 Old Lee Hwy., Room 305, Fairfax, VA 22030-1504

mike_hogye@nass.usda.gov

Key Words: *imputation, donor, editing*

For the 2002 U.S. Census of Agriculture, USDA/NASS employed an array of new tools developed for this purpose. This included the use of donor imputation to adjust records that could not be corrected through other editing strategies. While the goal of incorporating Fellegi-Holt principles proved to be overly ambitious, significant progress was made in broadening the scope of available donors. Large investments in automation of data-editing provided complementary benefits to the imputation process. On one hand, efficient data retrieval and powerful processing made possible the creation of large donor pools and extensive donor searches for nearest neighbors. On the other hand, interactive data-editing capabilities allowed a nationwide network of analysts to correct records for

which donor imputation proved problematic. This paper reviews the imputation techniques used, lessons learned, and improvements being considered for imputation in the 2007 Census. Specific issues include the construction of initial donor pools; selection of appropriate similarity measures; scaling of continuous imputed values; and balancing of computational demands with improvements to data quality.

Using Shortest-path Algorithms to Derive Implied Ratio Edits

◆ Brian V. Greenberg, Social Security Administration; James T. Fagan, U.S. Census Bureau

Social Security Administration, 6401 Security Blvd., 4-C-15 Operations Bldg., Office of Research, Evaluation, and Statistics, Baltimore, MD 21235

Brian.V.Greenberg@SSA.GOV

Key Words: *economic editing, SPEER edit system, ratio edits, implied edits*

The SPEER (Structured Program for Economic Editing and Referrals) edit system was developed at the Census Bureau in the early 1980s for editing and imputation of economic data collected in the economic censuses and other major economic surveys. This edit system subsequently has been incorporated into the Census Bureau's standard system for processing economic survey and census data. For each application for which it has been adapted, the SPEER system has been customized to meet specialized data needs, as planned in the original design. However, there have been no changes to the underlying edit methodology or the strategy for generating implied ratio edits. We present the theory for generating implied ratio edits as a shortest-path problem and introduce simple yet rigorous methods to derive implied edits. The methods discussed here have already been incorporated and successfully used in Census Bureau programs for processing economic surveys.

Use of GST Data by the Monthly Survey of Manufacturing

◆ Wesley Yung, Statistics Canada; Steven Thomas, Statistics Canada; Krista Cook, Statistics Canada

Statistics Canada, 11-F R.H. Coats Bldg., Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

Wesley.Yung@statcan.ca

Key Words: *administrative data, response burden, imputation*

Statistics Canada's Monthly Survey of Manufacturing (MSM) provides information on shipments, inventories, and orders for all manufacturing establishments in Canada. One of the principle users of MSM data is the System of National Accounts in calculating the Gross Domestic Product, an important economic indicator. Statistics Canada recognizes that monthly surveys, such as MSM, place a large burden on sampled units and is considering methods to use administrative sources of data to reduce this burden. One source of administrative data is the Goods and Services Tax (GST), which was introduced in 1991 as a tax on goods and services provided by businesses in Canada. This data can be used to reduce data-collection costs, as well as reduce response burden. However, many challenges exist with the use of the GST data such as timeliness, data comparability, and lack of remittance by establishments. This presentation will cover the challenges faced and solutions implemented by Statistics Canada to successfully integrate the GST data into the MSM for a large number of sampled units.

Evaluating Regression Imputation Models: An Example from the Services Sectors Portion of the Economic Census

◆ Quatracia Williams, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Suitland, MD 20746

quatracia.williams@census.gov

Key Words: *imputation, multicollinearity, multiple regression, mean absolute error, mean absolute deviation*

The Economic Census uses a variety of statistical models for item imputation. Currently, the services sectors portion of the Economic Census relies on industry-average-ratio imputation as its primary statistical imputation model, developing parameters from no-intercept simple linear regression models using weighted least squares estimation to compensate for heteroscedasticity. Our prior research (using 1997 data) showed improved predictions for these Economic Census sectors using multiple regression models. The high correlation between candidate covariates introduces the model-fitting issue of multicollinearity. The presence of multicollinearity and heteroscedasticity renders the "traditional" regression evaluation diagnostics inappropriate. We present a model selection procedure that uses alternative robust statistics and relies primarily on cross-validation. In theory, multiple regression models constructed from multicollinear data (with unequal error variances) can be used with other datasets for prediction, provided that the prediction region does not change. This paper evaluates our methods using data from the 1997 and 2002 Economic Censuses.

Automated Production of Foreign Trade Data Edit Parameters Using Resistant Fences

◆ Ryan M. Fescina, U.S. Census Bureau

U.S. Census Bureau, Washington, DC 20233-0001

Ryan.M.Fescina@census.gov

Key Words: *outlier detection, resistant fences, data editing, economic data*

The Foreign Trade Division (FTD) of the U.S. Census Bureau is responsible for publishing the official international trade statistics for the United States. This includes data for both imports and exports. Each month, FTD collects and edits approximately 3.4 million import and 1.8 million export records, covering nearly 17,000 import and 10,000 export commodities. These records pass through many edits using multiple sets of parameters. Due to resource constraints and the potential for an increase in the number of parameters, we are investigating ways to automatically create tolerances based on a Symmetrized Resistant Fences method. This paper presents our findings on the implementation of Symmetrized Resistant Fences applied to foreign trade ratio parameters, and the use of historical data to flag questionable parameters or automatically update the parameters as needed.

402 Data Monitoring Committees - Challenges ▲

Biopharmaceutical Section

Thursday, August 12, 8:30 am–10:20 am

What is Enough? Preparation of DMC Reports on Interim Data

◆ Marian R. Fisher, University of Wisconsin, Madison

University of Wisconsin, Madison, 610 Walnut St., Room 215, Dept. of Biostat & Med. Informatics, Madison, WI 53726

fisher@biostat.wisc.edu

Key Words: *DMC, interim analysis, sample report*

The DMC's primary responsibility is to ensure, throughout the trial duration, risks to patients are reasonable in relation to anticipated benefit. DMC reports, based on interim accumulating data, need to be thorough, timely, and objective. Study performance is included as well as treatment efficiacy and occurrence of adverse events. Interim data from different sources (e.g., accrual information, SAE reports, CRF data) should be analyzed. The specific structure or content need not be fixed and may change during the trial's course. Many factors (such as stage of the trial, nature of accumulating data, focus of a DMC meeting, and requests by DMC members) will influence report content or how data items are analyzed and presented. The report format should enable large amounts of information to be communicated in a manner that is both maximally informative and easy to review in a short period of time. A sample report produced from the Beta-Blocker Heart Attack Trial is available at *www.biostat.wisc.edu/clintrials/sdac/reportprod.html*.

Clinical Trial DMCs: Flexible Approaches for Optimal Monitoring

◆ Andreas Sashegyi, Eli Lilly and Company

Eli Lilly and Company, Lilly Corporate Center, Faris Building 1, DC 6072, Indianapolis, IN 46285

aisasheg@lilly.com

Key Words: *data-monitoring committee, interim analysis, data analysis group*

The regulatory environment of pharmaceutical development is one of ever-increasing focus on patient safety and minimization of sponsor involvement in any aspect of clinical trial conduct that could bias the final results. Consequently, the role of interim analyses of accumulating data in clinical trials and the processes that govern such analyses have recently received a great deal of attention. The FDA draft guidance on DMCs, issued in late 2001, was the first document to formally suggest principles for the establishment and operation of DMCs. In particular, it focuses on registration trials, recommending sponsor-independent DMCs, and also suggesting independent data analysis groups. Using the draft guidance as a starting point, we discuss the various components of the interim analysis/DMC process and their broader application to any trial which might include interim analyses. We suggest appropriate adaptations of the DMC process to a given study, depending on its purpose, phase, and endpoint. Guidelines are presented to aid in deciding on the optimal monitoring strategy.

Clinical Trials Data-monitoring Committees: The Who, What, When, and Why of Unmasking

◆ Anne S. Lindblad, EMMES Corporation

EMMES Corporation, 401 N. Washington St., Suite 700, Rockville, MD 20850

alindblad@emmes.com

Key Words: *DMC membership, unmasking, interim data, clinical trial*

Data Monitoring Committees (DMCs), composed of independent experts, review accumulating data within randomized controlled trials and provide recommendations to study sponsors (NIH and/or industry) for study continuation without change, modification, or termination. The recommendation is based on a review of the totality of the information collected within the trial as well as information that may be available publicly. Data from within the trial typically reviewed include, data quality and currency, adherence to study design, enrollment, and safety and efficacy outcome measures. DMCs may meet on a calendar basis or on a study-completion-level basis depending on monitoring guidelines adopted at study initiation. DMC meetings may include nonvoting members. The composition of nonvoting members with respect to their relative independence from study funding and study operation and the level of participation of these nonvoting members in DMC data review, both masked and unmasked, is under debate within the clinical trials community.

How to Organize a Data Monitoring Committee and Prepare an Optimal Data Package?

◆ Vipin K. Arora, Eli Lilly and Company; David Manner, Southwestern Medical School

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

arora_v@lilly.com

Key Words: *data monitoring committee, data analysis group, best practices*

The FDA draft guidance document, "On the Establishment and Operation of Clinical Trial Data Monitoring Committees," provides insight on organizing a data monitoring committee (DMC); however, the actual implementation of a DMC can become a challenging experience for statisticians. Some of challenges may include: identifying therapeutic area (TA) experts, selecting a chairperson, logistics of DMC meetings, protecting bias by arranging an independent data analysis group (DAG), unblinding process, and finalizing DMC charter. The statistician has key responsibility of collaborating with the DMC chairperson and finalize list of reports that are shared at each DMC meeting. An optimal data package is key to enable DMC to make an accurate assessment. Thorough discussions on TA, Protocol, and SAP must occur between the sponsor statistician and DAG to enable DAG to prepare responses to questions at the DMC meeting. In theory, SAP can cover all analyses that are executed by DAG. In practice, however, this arrangement can result in a situation wherein few key questions raised by DMC go unanswered due to lack of all background information within DAG.

403 Adapting to Changes in the Privacy and Confidentiality Landscape

Section on Government Statistics, Cmte on Privacy and Confidentiality, Section on Survey Research Methods, Social Statistics Section

Thursday, August 12, 8:30 am–10:20 am

Implementing the Confidential Information Protection and Statistical Efficiency Act of 2002

◆ Katherine K. Wallman, U.S. Office of Management and Budget; Brian A. Harris-Kojetin, U.S. Office of Management and Budget

U.S. Office of Management and Budget, 1101 Waverly Way, McLean, VA 22101-2239

kwallman@omb.eop.gov

Key Words: *confidentiality, data access, data-sharing*

The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) provides a set of uniform minimum safeguards for the protection of data gathered for exclusively statistical purposes under a pledge of confidentiality. Federal agencies have long sought the kind of strong confidentiality protection that CIPSEA offers. The Office of Management and Budget (OMB) chartered an interagency team to draft implementation guidance for this law. We will provide an overview of the proposed implementation guidance on CIPSEA and discuss some of the requirements for agencies using CIPSEA for their collections. Specifically, the proposed guidance includes requirements for agencies collecting information protected under CIPSEA, minimum standards for safeguarding confidential information, and provisions related to the designation of agents. We will also discuss some of the ways that this law has begun to affect practices and procedures at federal agencies.

How a U.S. Federal Statistical Agency Responded to CIPSEA

◆ William Parks, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 4040, Washington, DC 20212

parks.william@bls.gov

With the passage of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) in December 2002, the Bureau of Labor Statistics (BLS) began the complex task of implementing CIPSEA across its statistical programs. Many issues face a statistical agency immediately following the passage of a sweeping piece of legislation like the CIPSEA. BLS has worked to incorporate the new law into every aspect of its operations, ranging from data collection to employee training to reworking some aspects of its contractual relations with state data collection agencies. Managing all of the necessary changes and communicating them to BLS employees and agents has been challenging. This presentation will highlight what changes needed to occur at BLS as a result of CIPSEA and how the agency has handled such changes.

Incorporating HIPAA Privacy Rule into Medical Record Surveys

◆ Catharine W. Burt, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, NCHS, 3311 Toledo Rd., Hyattsville, MD 20782

cwb2@cdc.gov

Provider-based surveys that collect information from medical records were challenged by new requirements in the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) which became effective April 14, 2003. While CDC surveys are authorized under the Public Health Service Act, which assures confidentiality provided by law, the new regulation sets further standards for providers when disclosing protected health information for research or public health purposes. This paper describes the ways survey procedures were modified to assist providers in participating in the surveys under the new guidelines including obtaining or modifying IRB approval, creating data use agreements, completing accounting documents for disclosures made to assist the proivders' record keeping requirements, creating new training materials for field staff and new web page materials for providers, etc. Implementation of HIPAA's Privacy Rule led to increased survey costs but had less of an effect on survey response than was anticipated. In some respects, the additional assurances of confidentiality made some providers more comfortable with providing patient data. However, the full impact on response rates will need to be measured over time as providers and survey organizations become more confident about the provisions in the Privacy Rule allowing continued survey participation.

Federal Data Sharing Requirements and Issues: Contributions to be Made by Statistics, Survey Research, and Related Disciplines

◆ Virginia A. de Wolf, Consultant

Consultant, 522 Oakland Hills Lane, Kerrville, TX 78028

dewolf@ktc.com

Key Words: *NIH, NSF, HIPAA, statistical disclosure limitation methods, accessing confidential data, data protection*

Beginning Fall 2003 the National Institutes of Health (NIH) required that all funding proposals requesting over $500,000 in direct costs in a year include specific plans for data-sharing. These plans must safeguard the privacy of participants and protect confidential and proprietary data. The National Science Foundation (NSF), another major federal source of research funding, has a policy which strongly encourages the sharing of data collected by NSF grantees. And the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) permits the sharing of "protected health information" with or without patients' authorization. Researchers applying for NIH and NSF grants, as well as those requesting data under HIPAA, need to know about methods used to protect confidential data (termed "restricted access" and "restricted data"). The purposes of this paper are to briefly describe these methods, suggest academic disciplines that have knowledge of such methods, and strongly encourage faculty with knowledge of these methods to pool their expertise and serve as resources for their colleagues. Ideas for such collaborations are given.

Motivating Standards for Disclosure Avoidance

◆ Philip M. Steel, U.S. Census Bureau

U.S. Census Bureau, Rm 3209 FB4, 4700 Silver Hill Rd., Washington, DC 20233-9100

philip.m.steel@census.gov

Key Words: *confidentiality, HIPAA, record linkage*

Advances in computing and methods of access to data have made statistics an integral part of modern life. From marketing to evaluating public policy to clinical studies, solid data and good statistics are no longer the earmark of good, cutting-edge science, but rather an underpinning of everyday life. Many of these advances also raise privacy concerns and can create a trade-off between inadvertent disclosure of personal data and the accuracy of the inference from that data. Often data can be protected without harming the utility of the data and at very little cost. In some contexts protection and utility collide head on. Disclosure avoidance seeks to establish a best practices consistent with a statistical use of person level data and offer alternatives when utility cannot be maintained. This paper is intended as a general introduction to disclosure avoidance. We demonstrate some simple data attacks and techniques to block them. We give several examples of more complicated situations arising from the nature of the data and describe the outcome of those situations. We look at regulation in the federal data system, in HIPAA and in NIH funding requirements.

404 **Statistical Methods**

Section on Statistical Computing

Thursday, August 12, 8:30 am–10:20 am

Inference on General Linear Regression Models with Unequal Error Variances

◆ Shun-Yi Chen, Tamkang University; Miin-jye Wen, National Cheng Kung University

Tamkang University, Dept. of Mathematics, Tamsui, 251 Taiwan

sychen@math.tku.edu.tw

Key Words: *one-sample procedure, heteroscedasticity, t distribution, hypothesis testing*

A one-sample procedure is proposed for statistical inference on a general linear regression model with unknown and possibly unequal normal error variances. The goal is (1) to test the null hypothesis that the response variable is free of all on-hand predictor variables, and (2) to test the null hypothesis that a subset of predictor variables are unimportant. Two new test statistics based on weighted sample means at each predictor's data points are proposed. It turns out that these test statistics have sampling distributions completely independent of the unknown variances, and the p values of these test statistics can be calculated by simulation. The item (1) can used to test the treatment effects for one-way layout ANOVA model and the test (2) can be applied to regression model selection problem when one predictor variable is considered at a time as well to test the row and column effects, respectively, in a two-way layout ANOVA model.

Dealing with Model Bias in a Nonlinear Mixed Model

◆ Karen E.A. Chiswell, North Carolina State University; John Monahan, North Carolina State University

North Carolina State University, 2204 Hwy.158 W, Oxford, NC 27565

kechiswe@unity.ncsu.edu

Key Words: *first-order linearization, nonlinear mixed model, model misspecification, PBPK model, sensitivity analysis*

A common assumption when fitting a nonlinear model is that the nonlinear function correctly specifies the mean trajectory being described. However, it is quite common in particular applications (e.g., toxicology, pharmacokinetics) to work with mechanistic models that suffer from some degree of misspecification, or model bias. An important implication of this bias, especially in the case where data have a longitudinal structure, is the resulting difficulty in modeling within-individual covariance structure. We examine various approaches to handling the apparent model bias in a physiologically based pharmacokinetic model (PBPK), used to analyze data from a closed chamber study of CCl4 metabolism in rats. These approaches include sensitivity analysis, first-order linearization (FOL) of the likelihood function (based on normality assumptions), and smoothing methods.

Bootstrap Confidence Intervals for Classification Error Rate in Circular Models

◆ Hie-Choon Chung, Gwangju University; Chien-Pai Han, University of Texas, Arlington

Gwangju University, 592-1 Jinwoldong, Namgu, Gwangju, 502-703 South Korea

hcc@hosim.gwangju.ac.kr

Key Words: *bootstrap confidence interval, error rate, block of missing observations, circular models, intraclass correlation, Monte Carlo study*

In circular models, we consider bootstrap confidence intervals for classification error rate when the training samples contain a block of missing observations, and when observations are randomly missing. We also consider bootstrap confidence intervals for classification error rate with intraclass correlation when observations are randomly missing. A Monte Carlo study is conducted to evaluate the coverage probability and the average length of the confidence interval.

Statistical Analysis Software for Multiplicative Interaction Models

◆ Eun-Joo Lee, Kansas State University; Dallas E. Johnson, Kansas State University

dejohnsn@ksu.edu

Key Words: *AMMI, genotype/environmental interaction, two-way experiments*

In a two-way cross-classified experiment, one is almost always interested in whether the two factors interact or not. When there are no independent replications there are no traditional tests for interaction. This research considers the problem of analyzing a two-way cross-classified experiment using multiplicative interaction models when there are no independent replications and interaction between the two factors may exist. There is no generalized software readily available for the analysis of multiplicative interaction models. SAS macros have been developed to provide user-friendly statistical software for the analysis of the interaction in two-way experiments. The software also provides many useful graphical displays to help determine the pattern of interaction and to help one interpret the results of the analyses. The presentation will describe the macros and illustrate them with examples.

Maximum Likelihood Estimation from Local Moments

◆ Masahiko Sagae, Gifu University; Atsuyuki Kogure, Keio University

Gifu University, 1-1 Yanagido, Gifu-shi, Gifu-ken, 501-1193 Japan

sagae@info.gifu-u.ac.jp

Key Words: *local moment, likelihood, Sheppard's correction, maximum entropy density, polynomial histogram*

Observations should cause somewhat loss of information under data aggregation such as grouping. For example, the maximum likelihood (m.l.) estimate under grouping has bias depending on binwidth, which corresponds to Sheppard's correction for a specific case. We consider the m.l. estimates of parameters of continuous distribution from local moments. The local moments extended grouping was introduced in our previous works. The theoretical property of m.l. estimate under local moments is investigated. It is shown that the m.l. estimate from local moments does not cause loss of information in a specific case. As this result, we constitute specific local moments so that the m.l. estimates have no loss of information or so that local moments diminish loss of information with respect to m.l. estimation, compared with grouping.

The Cutting Plane Algorithm in the Error Localization Problem

◆ Bor-Chung Chen, U.S. Census Bureau; William E. Winkler, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop 9100, Washington, DC 20233-9100

bchen@info.census.gov

Key Words: *redundant covers, subcovers, integer programming, optimization*

The edit generation in the Fellegi-Holt model (1976) is a NP-hard problem. It is very difficult to generally generate a complete set of edits for the error localization problem in the model. A cutting plane algorithm is desired for the error localization problem if an edit-failing record is not able to be filled in with the observed explicit and implicit edits as described in Garfinkel, Kunnathur, and Liepins (GKL). We will describe a modified cutting plane algorithm of GKL Algorithm 2 to speed up the error localization process for the practical applications of the Fellegi-Holt editing model. We will show some numerical examples for the alternative approaches with and without generating the unobserved implicit edits.

University-wide Statistical Computing Support: A Practical Approach

◆ John V. Samuel, Indiana University; Takuya Noguchi, Indiana University

Indiana University, 410 N. Park Ave., Bloomington, IN 47408

samuel@indiana.edu

Key Words: *statistical software, computing, consulting, training, distribution*

A system-wide statistical and mathematical computing support program has been in place at Indiana University for approximately 10 years. Consultants with expertise in a variety of statistical and mathematical software packages work with faculty, staff, and students from diverse disciplines, answering questions ranging from very simple to those requiring an extended period of consultation. Regardless of whether the method of contact is phone, electronic, or walk-in, consultants work with the users to help them learn the process as they deal with a particular question, rather than spoon-feeding them the answer. The Stat/Math Center supports over 20 different statistical and mathematical software packages for Windows, Macintosh, and Unix environments for instruction and research. In addition to providing consulting support, the Center distributes statistical and mathematical software packages to the University community at a cost below the discounted educational price through a cost-sharing process. User feedback has consistenly been excellent over the years. Reasons for the program's success and cost savings to the University will be presented.

405 Clinical Trials

Biopharmaceutical Section, Biometrics Section

Thursday, August 12, 8:30 am–10:20 am

Analysis of Clinical Trial Data When Treatments Favor Different Endpoints

◆ Leslie A. McClure, University of Michigan; Morton B. Brown, University of Michigan

University of Michigan, School of Public Health, Bldg. II, Dept. of Biostatistics, 1420 Washington Heights, Room 4331, Ann Arbor, MI 48105-2029

lmcclure@umich.edu

Key Words: *clinical trials, multiple endpoints*

Most clinical trials are designed to have one primary endpoint. However, it is becoming increasingly common to design studies for which more than one primary outcome exists. In these situations, it is important to include all of the primary outcomes in the analyses. Methods that exist for more than one outcome are generally concerned with identifying one of two situations: (1) improvement in all outcomes from one treatment, as compared to the other treatment(s), or (2) improvement in a single efficacy endpoint, without an excess of toxicity. In contrast, the Colpopexy And urinary Reduction Effort (CARE) Study is a randomized clinical trial comparing two treatments, which presents a slightly different scenario. In this study, there are two outcomes, where each treatment may favor a different outcome. Few statistical methods exist to model bivariate data to test joint hypotheses that may be inconsistent in their inference. We will discuss both a likelihood-ratio-based approach and a Bayesian approach to test whether or not a treatment difference exists at the end of the study.

Multiplicity in Clinical Trials Involving Multiple Treatment Groups

◆ Kim Hung Lo, Centocor, Inc.; Mani Lakshminarayanan, Centocor, Inc.

Centocor, Inc., 200 Great Valley Pkwy., Malvern, PA 19355

KLo2@cntus.jnj.com

Key Words: *multiple comparison, Type I error rate, dose response, power, Dunnett's Test, pairwise comparison*

Clinical trials involving multiple treatment groups are very common in most of the late-stage drug development programs. For example, multiple treatment groups may be included in a Phase 2 trial designed to investigate the dose response of an experimental drug or in a Phase 3 trial with a primary objective of confirming one or more effective doses to be marketed. When there are multiple treatment groups, the question of multiplicity is of serious concern from the point of view of controlling the Type I error rate. Techniques that are typically used to control the Type I error rate includes an overall test (say, F-test in the case of linear models) followed by pairwise comparisons, Bonferroni adjustment, and closed test procedures. In Chapter 5 of "Multiple Comparisons: Theory and Methods," Jason Hsu provides arguments and other details of choosing one or more methods under certain conditions. One such method that is commonly used in practice involves performing an overall test comparing a combined group of dose levels of the test drug versus the control, followed by pairwise comparisons.

Multiplicity Adjustments for Clinical Trials with Multiple Doses in Two-dose Regimens

◆ Peng-Liang Zhao, Merck & Co., Inc.; Hui Quan, Merck & Co., Inc.

Merck & Co., Inc., RY34-A304, 126 E. Lincoln Ave., Rahway, NJ 07065

peng-liang_zhao@merck.com

Key Words: *familywise error rate, closed procedure, strong control, weak control*

Typically there is no prior information regarding which dose regimen of an active compound might be more effective or better, therefore two regimens (e.g., once-daily and twice-daily doses) can be considered in a phase II dose-range finding study. After the phase II study, if one regimen (e.g., once-daily dose) is numerically less efficacious than the other regimen, but is more convenient for patients than the other regimen and is also likely to be registrable, multiple doses in both regimens can be included in phase III confirmatory trials and be compared to the control. We discuss multiplicity adjustment procedures for two regimens and multiple doses in these trials. We consider procedures which use the dose-response relationship of each regimen to form the closed procedures and control the familywise Type I error rate in a strong sense. These procedures generally have a higher power than the traditional procedures such as the Bonferroni and the Hochberg procedures for the high doses. We also discuss procedures which control the familywise Type I error rate in a weak sense. Numerical examples will be used to illustrate the applications of our procedures.

A Comparison of Methodologies for Assessing Onset of Action in Neuropsychiatric Clinical Trials

◆ Craig H. Mallinckrodt, Eli Lilly and Company; Michael Detke, Eli Lilly and Company; Kaiser Christopher, Eli Lilly and Company; John Watkin, Eli Lilly and Company; Geert Molenberghs, Limburgs Universitair Centrum; Raymond J. Carroll, Texas A&M University

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

mallinckrodt_craig@lilly.com

Key Words: *mixed-effects models, survival analyses, antidepressants, missing data*

Antidepressant medications have a delay in onset of effect. Given the importance of onset of action, this is an active area of investigation. This study compared Type I error rates and power of the Kaplan-Meier Product Limit method (KM) to a categorical mixed-effects model repeated measures approach (MMRM-CAT). Four scenarios were simulated by varying mean improvements over time. Within each scenario, three sets of data were created by either leaving the data complete or deleting observations to generate MNAR data according to one of two dropout patterns. All 16 (twice weekly for eight weeks) post-baseline visits were included (frequent assessment), or only weeks 1, 2, 4, 6, and 8 (traditional assessments). Onset was defined as a sustained 20% improvement from baseline. Frequent assessments improved precision compared with traditional assessments in KM more than MMRM-CAT, but the reductions were small in all cases. Both KM and MMRM-CAT provided reasonable control of Type I error. The MMRM-CAT analysis using a traditional visit schedule yielded more power and was biased less by subject dropout than KM with either a traditional or a frequent assessment schedule.

Stochastic Models for Compliance Analysis and Applications

◆ Junfeng Sun, The Ohio State University; Haikady N. Nagaraja, The Ohio State University

The Ohio State University, 650 Trumbull Ct., Columbus, OH 43210

sun@stat.ohio-state.edu

Key Words: *electronic event-monitoring, compliance, stochastic modeling, Markov-dependent mixture models, empirical Bayes*

Compliance is the extent to which a patient follows the prescribed regimen. It is crucial to accurately measure compliance in both clinical trials and medical practice. It has received a lot of attention in the medical literature. However, little effort has been made to study the statistical properties of various compliance metrics. This study utilizes the information-rich electronic event monitoring (EEM) data and studies the statistical properties of several clinically meaningful compliance metrics. Markov-dependent mixture models are applied for the metrics based on continuous data. To pool the data from different patients, we use empirical Bayes approaches. As an illustration, we apply these metrics to an AIDS clinical trial. Practical applications of these compliance metrics will also be discussed.

Statistical Models for the Analysis of Carotid Intima-media Thickness Data in Clinical Trials

◆ Chunchung G. Law, Pfizer Global Research and Development; Mingxiu Hu, Pfizer Global Research & Development

Pfizer Global Research and Development, 50 Prequot Ave., New London, CT 06320

lawcg@groton.pfizer.com

Key Words: *multiple imputation*

Carotid intima-media thickness (CIMT) is widely used as a surrogate marker of atherosclerosis. By ultrasonographic imaging, it was measured on 12 segments of the carotid artery walls at k (multiple) visits in some clinical trials. The data may be incomplete due to nonvisualization of some segments and missed visits. The "working" correlation structures in the longitudinal models are complicated as, within each subject, the correlation between different segments and different visits need to be accounted. We considered two types of mixed-effects models: longitudinal models with (k x 12) repeated measurements and those with k repeated measurements by using the average over 12 segments. Three methods are evaluated for handling missing values. We assessed these approaches in different settings through extensive simulation studies. Under the scenarios we considered, the two mixed effects models showed no advantage over each other, while the data augmentation method demonstrated some benefits for imputing missing values. This presentation will also introduce the structure of our complex simulation program.

Statistical Validation of Surrgoate Endpoints in HCV Clinical Studies

◆ James Jiao, Roche Laboratories, Inc.; Sylvia Hu, Roche Laboratories, Inc.; Guoqin Su, Novartis

Roche Laboratories, Inc., 340 Kingsland Rd., Nutley, NJ 07110

james.jiao@roche.com

Key Words: *clinical trial, HCV, surrogate endpoint*

A statistical definition of surrogate endpoints as well as validation criteria was first presented by Prentice. Freedman et al. supplemented these criteria with the so-called proportion explained (PE). Buyse and Molenberghs proposed a new definition of surrogacy in terms of relative effect (RE) and adjusted association. In clinical studies on HCV disease, there are several potential surrogate endpoints whose surrogacy has not been thoroughly studied yet. The surrogacy of these endpoints will be studied based on (1) proportion explained (PE), and (2) the relative effect (RE) and the treatment-adjusted association between the surrogate endpoint and the true endpoint. The adjusted associations are examined both on individual patient-level and the trial level. Different strategies for the jointly statistical modeling with mixed binary-continuous endpoints are also presented.

406 Statistical Applications in Business and Economics ▲ ⌘

Business and Economics Statistics Section

Thursday, August 12, 8:30 am–10:20 am

A Modified Coefficient of Variation for Partially Grouped Data

◆ Keith Ord, Georgetown University; Harvey Iglarsh, Georgetown University

Georgetown University, McDonough School of Business, 320 Old North Bldg., Washington, DC 20057

ordk@georgetown.edu

Key Words: *coefficient of variation, income distributions, inequality measures, variable group width*

Observations from a bivariate distribution (X, Y) are grouped according to their X-values. The coefficient of variation (CV) of Y within each subgroup is often used to compare the variability in Y across those subgroups. Existing methods restrict ge to subgroups of equal width. We develop an extension of the CV that allows comparison of subgroups with unequal ranges. We use this measure to examine horizontal equity (the proposition that taxpayers with equal ability to pay [X] should have the same tax liability [Y]) in different income tax systems using grouped data from public IRS data files.

A Further Result on the Calculation of the Total Claims Distribution under Different Conditions

◆ Min Deng, Maryville University, St. Louis

Maryville University, St. Louis, 13550 Conway Rd., St. Louis, MO 63141

mdeng@maryville.edu

Key Words: *compound mixed distribution, compound distribution, characteristics of counting distribution, probability density function, recursive formula*

Various methods for calculating the total claim distribution have been discussed by Panjer, Shorter, Sundt, and Deng. We examine the distribution properties of those mixture classes under the different conditions. We also give the numerical examples to illustrate the results.

Pitfalls of Using Unit Values as a Price Measure or Price Index

◆ Ralph Bradley, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachuesetts Ave. NE, Room 3105, Washington, DC 20002

bradley.ralph@bls.gov

Key Words: *unit values, aggregation consistency*

With large datasets, it is tempting to use a unit value as a measure of price when constructing aggregate price indexes or modelling aggregate demand systems. This study shows that in most situations this will lead to inconsistent estimates. With superlative price

indexes, the use of unit values will misspecify substitution effects. The use of unit values as a price regressor in an aggregate demand model will misspecify the model even though the functional form of the demand model is correct. I present two empirical investigations using data from New York cereal sales. In the first, price indexes are constructed with and without the use of unit values. In the second, the model of Hausman (1997) is re-estimated with and without unit value. In both investigations, the use of unit values produces different results that use consistent aggregation methods.

Measuring the Preventive Value of Remote Maintenance

◆ Ping Zhang, Avaya Labs Research; James M. Landwehr, Avaya Labs Research; Mihaela Serban, Carnegie Mellon University

Avaya Labs Research, 233 Mt. Airy Rd., Rm. 2D-14, Basking Ridge, NJ 07920

pingzhang@avaya.com

Key Words: *Simpson's Paradox, case control study, network monitoring, return on investment, generalized linear models, cost-benefit analysis*

Modern businesses depend critically on the availability of communications networks. The maintenance and management of these networks is a valuable service often provided by service providers or equipment manufacturers. How do we measure the business value of a service agreement? The standard industry practice is to measure service quality through various service level agreements (SLA). We propose an alternative approach based on statistical comparison of customers with or without a service agreement. We analyze a set of real data with 450,000 trouble tickets from 68,000 products over a five-month period. Here service agreement includes remote monitoring. Our main result shows that, among customers that experienced major problems, having a service agreement reduces the chance of a network outage by 65%. In other words, remote monitoring has preventive value that is significant and quantifiable. We conclude by presenting a simple cost-benefit analysis showing that outage rate reduction can be translated to customer benefit in monetary terms.

Dynamic Profiling of Online Auctions Using Curve Clustering

◆ Wolfgang Jank, University of Maryland

University of Maryland, 4322 Van Munching Hall, Smith School of Business, College Park, MD 20742

wjank@rhsmith.umd.edu

Key Words: *functional data analysis, smoothing, clustering, bid sniping, electronic commerce, online auction*

Electronic commerce, and in particular online auctions, have received an extreme surge of popularity in recent years. While auction theory has been studied for a long time from a game-theory perspective, the electronic implementation of the auction mechanism poses new and challenging research questions. Although the body of empirical research on online auctions is growing, there is a lack of treatment of these data from a modern statistical point of view. We present a new source of rich auction data and introduce an innovative way of modeling and analyzing online bidding behavior. In particular, we use functional data analysis to investigate and scrutinize online auction dynamics. We describe the structure of such data and suggest suitable methods, including data smoothing and curve clustering, that allow one to profile online auctions and display different bidding behavior. We illustrate the methods on a set of eBay auction data and tie our results to observed phenomena like early bidding, bid sniping, and bid shilling in the existing literature on online auctions.

Shewhart's Dilemma

◆ Joyce N. Orsini, Fordham University

Fordham University, Graduate School of Business, Deming Scholars MBA Program, 33 West 60th St., New York, NY 10023

orsini@fordham.edu

Key Words: *Shewhart, Deming, control chart, enumerative vs. analytic studies, process control, probability*

Walter A. Shewhart tried to find ways to use probabilistic statistics in the creation of his control chart. He observed, however, that process data do not meet the criteria for use of probability. Process data come not from a fixed frame, nor are they randomly selected. They are not iid variables. And he made no such assumption. He realized he was in a different world than classic statistical theory. He concluded that his results were empirically determined. It wasn't until some 20 years later that W. Edwards Deming shed some light on the quandary by drawing a distinction between enumerative and analytic studies. Shewhart was clearly in the analytic realm. This paper will explore how Shewhart reached his conclusions, why he couldn't use probability, and why his analytic control chart bears such close resemblance to enumerative statistics, where probability might have been used.

Distributional Dynamics of Per Capita Income in Latin America: A Nonparametric Density Estimation Approach

◆ Stephen Pollard, California State University, Los Angeles; Giorgio Canarella, California State University, Los Angeles

California State University, Los Angeles, Simpson Tower 901, 5151 State University Dr., Los Angeles, CA 90032

spollar2@calstatela.edu

This paper explores the regional dynamics of per capita income among 17 Latin American countries during the period 1960-2000 using the distributional dynamics approach pioneered by Quah (1993). We find evidence of "twin-peakedness" in Latin America, with countries polarized into two groups: those relatively better off (upper-middle-income countries) and those relatively worse off (lower-middle-income countries). Thus, our results confirm the "convergence clubs" hypothesis put forward by Quah; however, unlike Quah's findings, the polarization identified in the unconditional dynamics cannot be explained by physical geography and spatial spillovers. Rather, our findings indicate that the "twin peaks" polarization of per capita income in Latin America is driven by macroeconomic factors such as trade openness, government spending, and investment. Our results indicate that Latin American countries will remain trapped in their state of underdevelopment if only market-oriented policies are followed.

407 Experimental and Study Design ⊞

General Methodology

Thursday, August 12, 8:30 am–10:20 am

An Investigation of Minimization Criteria

◆ Angie Wade, University College London

University College London, Institute of Child Health, 30 Guilford St., London, WC1N 1EH England

awade@ich.ucl.ac.uk

Key Words: *minimization, patient allocation, bias, randomized control trials, confounder*

The power of a RCT is maximized if patients are divided equally between treatments with respect to confounding factors. Minimization is a dynamic allocation procedure that has been used increasingly in recent years. Allocation of the next patient is biased, according to his or her characteristics, in favor of the treatment arm that maximizes similarity between the resultant treatment groups with respect to selected potential confounders (known as minimization criteria). Despite becoming more widely used, details of precisely how minimzation has been applied are rarely given and many questions arise for a researcher wishing to use it. For instance: How many and which confounders should be balanced? Should these confounders be prioritized in some way? How biased should the randomization be? This presentation investigates, via simulation of datasets, the relationships between the number, type and weighting of factors, degree of randomization bias, total sample size and the expected size of discrepancies obtained between treatment groups. Results are presented together with practical implications for clinical ge of minimization as a means of patient allocation.

Study Design for Evaluation of a Diagnostic Test with Partially Missing Gold Standard

◆ Ying Chen, Emory University; Andrzej Kosinski, Duke University

Emory University, 1404C Druid Valley Dr., Atlanta, GA 30329

ychen7@emory.edu

Key Words: *sensitivity, specificity, diagnostic test, study design, partially missing gold standard*

Sensitivity and specificity are often used to evaluate performance of a diagnostic test. Sufficiently large number of patients is required for the confidence interval for the true sensitivity or specificity to be narrow enough to provide a prespecified precision. In practice, it is common that not all study patients have gold standard ascertained. To achieve the same precision as for the complete gold standard verification situation one requires more patients. The required sample size increases when the proportion of unverified disease status patients increases. We first give the sample size computation approach under the MAR missing data mechanism. Second, the sample size computation with no knowledge of the missing data mechanism is proposed.

Comparisons of Random Nested Designs Using Quantile Dispersion Graphs

◆ Juneyoung Lee, Korea University; Byung-Chul Jung, Korea University

Korea University, College of Medicine, Dept. of Preventive Medicine, Seoul, 136-705 Korea

jyleeuf@korea.ac.kr

Key Words: *random model, nested model, variance component, quantiles, staggered design, inverted design*

To compare designs for a linear model, criteria based on single-valued functions are typically used. These functions are, however, often difficult to use not only because of their complex forms but also because of their dependency on the model's unknown variance components. To overcome these limitations, a use of exact quantile values of estimates of variance components have been suggested. The use of quantiles can provide an overall quality of an estimator. When quantiles of estimates of variance components can not be obtained exactly, empirical quantile values can be considered. We focus on a use of quantiles of estimated variance components for a nested random model. To do this, three competitive designs, namely, the two-fold balanced nested design, two-fold staggered nested design, and two-fold inverted nested design are considered. For methods of estimating variance components, ANOVA and ML are used. Graphical approaches known as quantile dispersion graphs and empirical quantile dispersion graphs are employed in order to provide an overall quality of an estimator and to evaluate the quality of estimation obtained with a given design.

Optimal Incomplete Designs with Two Blocks

◆ Bo Jin, Virginia Polytechnic Institute and State University; John P. Morgan, Virginia Polytechnic Institute and State University

Virginia Polytechnic Institute and State University, 1801 Grayland St., Apt. # 6, Blacksburg, VA 24060

bjin@vt.edu

Key Words: *incomplete block designs, A-optimality, D-optimality, E-optimality, binary designs, nonbinary designs*

Suppose there are v treatments to compare using n experimental units. The units are partitioned into b blocks of k units each. A block design is an assignment of the v treatments to the bk units. Denote the class of all possible such assignments as D(v,b,k). We investigate the optimal incomplete block designs when there are two blocks using A-, D- and E- criteria. Binary designs of a certain pattern are shown to be A- and D-optimal in D(v,2,k). However, binary designs are not always E-optimal in D(v,2,k). In fact, E-optimality of binary designs depends on the magnitude of the ratio k/v. Among other results, it is found that in a restricted class, binary designs are E-optimal if $1/2< k/v< 5/6$; certain nonbinary designs are E-optimal if $5/6< k/v< 1$; both the binary designs and the nonbinary designs are E-optiaml if $k/v=5/6$.

Optimal Time Spacing for Designing Panel Studies

◆ Chen-Pin Wang, University of Texas, San Antonio

University of Texas, San Antonio, Health Science Center, 11700 Wall St. #5303, San Antonio, TX 78230

cwang@verdict.uthscsa.edu

Key Words: *design, time-spacing, panel study, changepoint, missing at random, latent growth*

This paper presents optimal time spacing schemes for designing panel studies. These optimization schemes are sought to maximize statistical power and estimation efficiency under resource constraints. Various latent variable growth models arising from health science research are considered. The parameters of interest are the effects associated with growth factors (e.g., slope and changepoints). Under missing-at-random assumption, we show that the optimization rule can be characterized as a function of (a) Euclidean distance between changepoints and the center of the rest time points selected, (b) the “generalized” variance of the time points other than changepoints, and (c) the “generalized” coefficient of variation associated with the time points other than changepoints.

Optimal Compound Orthogonal Arrays and Single Arrays for Robust Parameter Design Experiments

◆ Yu Zhu, Purdue University; Peng Zeng, Purdue University

Purdue University, Statistics Dept., 150 N. University St., W. Lafayette, IN 47907

yuzhu@stat.purdue.edu

Key Words: *robust parameter design, strong compound orthogonal array, single array, fractional factorial design*

Although several experimental strategies such as cross array, compound orthogonal array and single array are available for robust parameter design, the issue of constructing optimal experimental plans has not been settled with satisfaction. Recently, several papers intended to use a single criterion to choose the overall best designs; unfortunately, the proposed criteria can fail in many cases. We discuss the connections between the existing strategies and argue that the one-single-criterion approach may not be appropriate. We introduce two concepts: namely, strong compound orthogonal array (SCOA) and efficient single array (ESA). Various aberration-based criteria are proposed for constructing optimal SCOAs and ESAs according to the available experiment capacities. The properties of optimal SCOAs and ESAs are also investigated. The obtained results provide a clear guidance for experimenters to select proper designs in practice. Useful optimal SCOAs and ESAs are tabulated.

Designs Accounting for Potential Missing Trials in Multiple-objective Dose-response Studies

◆ InYoung Baek, SUNY, Stony Brook; Wei Zhu, SUNY, Stony Brook

SUNY, Stony Brook, AMS Dept. Math Tower, Stony Brook, NY 11794-3600

iypaik@ams.sunysb.edu

Key Words: *Bayesian optimal design, dose-response, missing data, multiple-objective*

The toxicity and efficiency of a therapeutic drug are determined before it is approved for general use. Quantal dose-response experiments are routinely conducted to examine the response rates at various dose levels of interest. However, due to side effect or lack of efficacy or other problems, responses are not always observed at all selected dose levels. Furthermore, the missing data patterns are often dose-dependent. Missing trials would invariably lead to reduced precision in parameter estimation and power in hypothesis testing. Depending on different missing data mechanisms, one can protect against potential missing trials by incorporating prior information in the design stage of the experiment. We examine missing data patterns from previous dose-response studies and subsequently design future multiple-objective dose-response studies accounting for potential missing observations in a Bayesian approach. The performances of designs with and without accounting for potential missing trials are compared in terms of the precision of related parameter estimation and the power of related tests.

408 Distribution Theory and Analysis

General Methodology

Thursday, August 12, 8:30 am–10:20 am

Improved Interval Estimation for the Two-parameter Birnbaum-Saunders Distribution

◆ Jianrong Wu, St. Jude Children’s Research Hospital; Augustine Wong, York University

St. Jude Children’s Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105

jianrong.wu@stjude.org

Key Words: *Birnbaum-Saunders distribution, confidence interval, pivotal quantity, probability coverage, signed log likelihood ratio statistic*

An improved interval estimation for the two-parameter Birnbaum-Saunders distribution is discussed. The proposed method is based on the recently developed higher-order likelihood-based procedure. The probability coverages of confidence intervals based on the proposed method and those procedures discussed in Ng et al. (2003) are evaluated using Monte Carlo simulations for small, moderate, and large sample sizes.

Tests for Location-scale Families Based on the Empirical Characteristic Function

◆ Thomas W. Epps, University of Virginia

University of Virginia, Economics Dept., Rouss Hall, Charlottesville, VA 22903

twe@virginia.edu

Key Words: *Anderson-Darling test, goodness-of-fit test, BHEP test*

A unified procedure is proposed for testing the fit of location-scale families. The principle is analogous to that of Anderson-Darling (AD) but applies the empirical characteristic function (CF) rather than the empirical distribution function (DF). Whereas AD statistics are derived as weighted integrals of squared distances between DFs of standardized data and models, the proposed CF statistics are weighted integrals of squared distances between empirical and model CFs. As do AD statistics, these have simple closed forms when squared moduli of model CFs are used as weight functions. While applications of the principle to normal and Cauchy laws are already well established, the paper extends the procedure to other location-scale models: uniform, exponential, Laplace, logistic, and extreme-value. Small-sample powers compare favorably with those of A-D and have particular advantage for uniform, exponential, Cauchy, and Laplace models.

Estimation of Hazard Rate Using Length-biased Sample

◆ Olcay Akman, College of Charleston

College of Charleston, Dept. of Mathematics, 66 George St., Charleston, SC 29424

akmano@cofc.edu

Key Words: *length-biased, inverse Gaussian, hazard rate, lognormal, reliability*

In the presence of length-biasedness, a lifetime measure of interest may be estimated in two ways: (i) by modeling the data correctly using a length-biased distribution and using the resulting estimators in the original model as an adjustment, or (ii) by modeling the data correctly using a length-biased distribution, and obtaining the original lifetime measure of interest via a transformation, if one exists. We examine sufficiency in information context under transformations.

A Modeling Approach for Distributions with Parametric Tails

◆ Jin Cao, Bell Labs Lucent Technologies; Yongzhao Shao, New York University

Bell Labs Lucent Technologies, 700 Mountain Ave., Room 2C-260, Murray Hill, NJ 07974

jincao@lucent.com

Key Words: *parametric tail, Pareto, Moran's statistic, semiparametric, communication networks, sample spacings*

Many datasets can be modeled by a distribution with a known parametric tail. For example, the web file sizes in communication networks have heavy tails, and thus the tail is often modelled by a Pareto distribution. To make inference on such a distribution (e.g., estimating its mean), we propose to use a semiparametric model where the distribution is modeled nonparametrically with a tail from a known parametric family $f(x;\theta)$. We propose an estimator that are used to determine both the starting point of the parametric tail and the possibly unknown parameter value θ simultaneously based on the method of spacings and establish its consistency property under general conditions. We illustrate our modeling approach using a real dataset of web file sizes.

Properties of Beta-Gamma Distribution

◆ Lingji Kong, Central Michigan University; Carl Lee, Central Michigan University; Jungsywan Sepanski, Central Michigan University

Central Michigan University, Dept. of Math., Mt. Pleasant, MI 48859

kong1l@cmich.edu

Key Words: *limit, mode, moments, reliability functions, hazard functions*

A class of generalized gamma distribution is proposed, the properties of this distribution are addressed, namely Beta-Gamma distribution in this paper. The properties of this distribution are addressed, including moments, limiting cases and percentiles. Graphs of the density functions are used to illustrate the shape properties for different combinations of parameters. Reliability and hazard functions are also derived.

409 Missing Data and Models for Health Policy Decisions ▲ ▦

Section on Health Policy Statistics, Section on Survey Research Methods

Thursday, August 12, 8:30 am–10:20 am

Imputation of Financial Fields in Integrated Anonymous Patient-level Databases

◆ Evguenia I. Jilinskaia, Pharmetrics Inc.; Teri Condon Goodwin, Pharmetrics Inc.; Cathy Johnson, Pharmetrics Inc.; Stanley Norton, Pharmetrics Inc

Pharmetrics Inc., 311 Arsenal St., Watertown, MA 02472

ev@pharmetrics.com

Key Words: *imputation, factor analysis, cluster analysis, generalized linear models*

Anonymous patient-level databases of integrated medical and pharmacy claims, provide the most complete representation of a patient experience with the health care system, including diagnoses, retail and mail order pharmacy prescriptions, medical procedures, laboratory tests, and corresponding financial fields: paid, allowed, and charged amounts. In some cases, the process of aggregating data from different sources to create Regional and National Normative studies and reports involves adjustments for differences of participating health plans ("Plan Effect") together with imputation of missing financial fields (due to capitation, etc.). To impute financial fields, claim records are classified by record type, type of product, place of service, provider type, and inpatient/outpatient settings. Further stratification of claims data is performed, based on factor scores and cluster analysis, to form homogeneous subsets of data. Mean values for dependent cost variables are estimated using generalized loglinear regression modeling (SAS Proc GenMod); based on the exponential family of distributions, which constitute a direct extension of traditional regression and analysis of variance. Predictor variables involve discrete categories, as well as continuous variables such as year, region, size of health plan, type of product.

Evaluating Factors of a Hot-deck Imputation Procedure in a Longitudinal Depression Treatment Study

◆ Juned Siddique, University of California, Los Angeles; Thomas R. Belin, University of California, Los Angeles

University of California, Los Angeles, Dept. of Biostatistics, Los Angeles, CA 90095-1772

siddique@ucla.edu

Key Words: *imputation, hot-deck, longitudinal study*

We discuss a predictive-mean-matching hot-deck approach for imputing missing data from a large longitudinal treatment study of low income minority women with depression. The hot-deck strategy investigated here involves imputing variables one at a time by using regression equations to match prospective donors and donees with similar predicted values. Important practical questions revolve around the extent to which investigator choices in the hot-deck procedure might affect subsequent inferences. We contrast results from two alternative approaches for variables in the predictive-mean-matching model: one where covariates in the imputation model were selected by the investigator on a priori grounds, and one where covariates are selected using a stepwise-regression approach. We also discuss the use of multiple imputation to accurately reflect imputation uncertainty and outline additional factors that we plan to study in terms of their impact on hot-deck results.

Physiological Reactivity and Post-traumatic Stress Disorder Diagnosis When Some Data are Missing

◆ Haekyung Jeon-Slaughter, University of Oklahoma Health Sciences Center; Phebe Tucker, University of Oklahoma Health Sciences Center; Carol S. North, Washington University School of Medicine; Betty Pfefferbaum, University of Oklahoma of Health Sciences Center

University of Oklahoma Health Sciences Center, Dept. of Psychiatry & Behavioral Sci., 920 Stanton L. Young Blvd., WP3212, Oklahoma City, OK 73104

hattie-jeon-slaughter@ouhsc.edu

Key Words: *incomplete data, missing values, PTSD, interview data, physiological reactivity, imputation*

This paper examines the effects of incomplete participation in research on the relationship between chronic post-traumatic stress disorder (PTSD) and physiological reactivity. Seven years after the Oklahoma City bombing, survivors were compared to community controls. All participants completed a structured diagnostic assessment, but 30% declined the physiological assessment (pulse and blood pressure measurements in response to a trauma cue interview). Missing values in physiological assessment may be random and unrelated to PTSD (ignorable), or they may not be random because participants with PTSD may be uncomfortable with the assessment (nonignorable results that may relate to study variables). Thus, excluding incomplete datasets may bias the results. We utilize graphic, maximum likelihood estimation, and simple mean substitute analyses to determine if missing values in this study are ignorable or nonignorable. Logistic regression and discriminant analyses are employed to investigate how well physiological reactivity predicts the research-generated PTSD diagnosis.

Latent Classes within the Military: Who May Need "Zero Tolerance" Interventions?

◆ Theresa J. Gromala, University of Minnesota

University of Minnesota, Health Services Research, Policy, and Administration, 2200 West 89th St., Bloomington, MN 55431-2007

groma001@umn.edu

Key Words: *latent class analysis, sexual harassment, sexual assault and rape, military, intervention, zero tolerance*

Latent classes within the military may indicate groups of individauls who are more likely to be sexually harassed or sexually assulted or raped. To explore this possibility, we examine responses from the Department of Defense (DOD) 1995 survey of sexual harassment. We answer four questions. Are there any meaningful classes underlying the observed variables selected from the DOD Survey? If classes are present, we plan to learn which classes are most at risk for sexual harassment and assault/rape. Are there any differences among the five military branches based on the classes we identify? The branches of the military include the Army, Air Force, Coast Guard, Marines, and Navy. Is there a class that represents the sexually harassed class? Finally, who is being hurt? If we learn who is most at risk for sexual harassment and assault/rape injury, we could develop interventions targeted toward the high-risk group. We determine cut-points between the classes we identify to choose the high-risk group. Through latent class analysis, we identify two classes of response patterns. We fitted a two class model. The first class represented the sexually harassed and assaulted/raped group.

A Prospective Approach to Decision-making for Conducting Public Health Interventions

◆ Thomas Wasser, Lehigh Valley Hospital

Lehigh Valley Hospital, 17th & Chew Sts., PO Box 7017, Allentown, PA 18011

Thomas.Wasser@LVH.com

Key Words: *interventions, public health, health policy, theoretical model, biostatistics, model-building*

A model is presented for describing the prospective feasibility of public health programs. This is a theoretical set theory solution containing three elements: (1) proportion of population that will receive the intervention divided into the proportion that possess the behavior to be changed and those that don't possess the trait; (2) effect size (ES), an index of potential success of the intervention; and (3) the proportion of population eligible for but not receiving the intervention. The model yields a Utility score (U), which is compared to the distribution of all possible values of U computed by simulation algorithm. U distribution was divided into fifths and given feasibility ratings in terms of expected impact: none, poor, moderate, good, and excellent. Distributions are presented in two ways: (1) Distribution cutpoints of U for 10 ES from 2% to 20% (the most common effect sizes in the public health setting), and (2) ES = 0.99 (representing a successful immunization program). While this data simulation is still theoretical, the U distribution is put forth as a first step to prospectively quantify public health interventions.

A Comparison of Methods Used for Constructing Confidence Intervals for Cost-effectiveness Ratio

◆ Ming-Yu Fan, University of Washington; Xiao-Hua A. Zhou, University of Washington

University of Washington, 1959 NE Pacific St., Box 356490, Seattle, WA 98195-6490

myfan@u.washington.edu

Key Words: *cost-effectiveness, simulation, bootstrap*

The incremental cost-effectiveness ratio (ICER) has been well accepted as a measure for cost-effectiveness in health sciences. Due to the skewed distribution nature for costs and estimated ICER, however, constructing an appropriate confidence interval for ICER is a challenge. Many methods have been proposed, yet systematic comparisons are rarely found in the literature. Also, there has been a conflict recommendation on which method is better. Polsky et al. (1997) and Briggs et al. (1999) have performed simulation studies to compare four and eight methods, respectively. However, their simulations only considered a selected number of available approaches and a very limited type of distributions for cost and effectiveness outcomes. We have conducted an extensive simulation to compare most commonly used as well as some new methods by generating different types of skewed samples. The methods are also compared across wide range of parameters to incorporate the variation observed in real data. The result suggests that bootstrap-t method is consistently better than other bootstrap or normal-theory-based methods for non-normal data. For normal data, all methods yield desirable result.

Type of Health Insurance, Race, and the Rate of Prostate Cancer

◆ Negasi Beyene, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, 3311 Toledo Rd. Suite #4121, CDC/NCHS/Research Data Center, Hyattsville, MD 20782

nbeyene@cdc.gov

Key Words: *cancer, prostate, screening*

Prostate cancer occurs when some of the cells that make up the prostate gland escape from the normal control on their growth and start to divide, grow and spread in an uncontrolled manner. At first the growth of cancer occurs very slowly and is usually limited within the prostate gland. Later on in the course of the illness, the prostate cancer cells can spread around the body, particularly to the bones where they can cause pain and disability. Estimates show that the cancer may have been growing in some men for up to 10 years before it causes symptoms and is diagnosed. Thus, early detection and treatment could prolong the lives of patients. Using the National Health Interview Survey of 1995-2002, I am trying to find out if there is a difference in the rate of prostate cancer by type of health insurance people have and their race. This study considers men of 40 years of age and older.

410 Topics in Sports Statistics ▲

Section on Statistics in Sports

Thursday, August 12, 8:30 am–10:20 am

Models for Third-down Efficiency in the NFL

◆ Christopher J. Rigdon, Southern Illinois University, Edwardsville; Steven E. Rigdon, Southern Illinois University, Edwardsville

Southern Illinois University, Edwardsville, Dept. of Mathematics & Statistics, Edwardsville, IL 62026

crigdon@siue.edu

Key Words: *logistic regression, third-down conversion*

We consider a logistic regression model for the probability of making a first down on a third-down play. The predictor variable is the yards to go for the first down. A regression analysis is performed for each NFL team's offense. We also consider the same model for each team's defense; that is, we model the probability that an opponent converts a third-down play. Finally, we consider a combined model for all 32 NFL teams where the probability of third down conversion depends on both the offense and the defense.

Football Performance Plots

◆ James M. Lucas, J.M. Lucas and Associates

J.M. Lucas and Associates, 5120 New Kent Rd., Wilmington, DE 19808

JamesM.Lucas@worldnet.att.net

Key Words: *winning percentage, CUSUM*

We evaluate the performance of college football programs and coaches using football performance plots. These plots show the strong effect of coaches on a teams winning percentage. In this they clearly show an effect that is known to exist. The sensitivity of these plots also allows an assistant coach effect to be discovered. The winning percentage of football teams can be strongly influenced by a superior assistant coach. Formally, a football performance plot is a CUSUM that plots: Cumulative Wins - (k) Cumulative Games versus Cumulative Games. The empirically determined reference value (k) is team- and coach-dependent. For strong programs a reference value of 0.75 works well as few football programs have had a winning percentage this high. For outstanding coaches a reference value of 0.8 often works well. Two examples of an assistant coach effect: Bobby Bowden had a 70.6% winning record at Samford, West Virginia, for his first 11 years at Florida State. While Mark Richt was an assistant coach, his winning percentage was 88.9%. Since Richt has left his winning percentage has reverted to its lower level. Joe Paterno's career winning percentage is above 75%.

Hybrid Paired Comparison Analysis

◆ David H. Annis, Naval Postgraduate School; Bruce A. Craig, Purdue University

Naval Postgraduate School, 1221 Cambia Dr., Apt. 1208, Schaumburg, IL 60193

annis@nps.edu

Key Words: *paired comparisons, ranking, football*

Existing college football ranking procedures focus on either binary responses (wins and losses) or points (either via each team's point total or a margin of victory). While each is reasonable, each fails to produce satisfactory rankings in frequently arising situations due to its ignorance of additional data. A new, hybrid model incorporating both wins and constituent scores is proposed. It is seen to outperform its competitors based on a series of simulation studies. This new approach is illustrated using the results of the 2003 NCAA Division-IA college football season.

Probability Models for 64-Team NCAA Tournaments

◆ Scott Stewart, University of Arkansas for Medical Sciences; Christopher R. Bilder, University of Nebraska, Lincoln

stewartscotta@uams.edu

Key Words: *logistic regression, NCAA tournament, Sagarin*

The Division I men's NCAA basketball tournament has been studied and analyzed extensively in recent years. There are several other tournaments in college athletics that warrant investigation. Probability models are built here as a function of team strength information, such as team seeds. Four different tournaments are used to build models: men's basketball, women's basketball, women's volleyball, and men's baseball. The fit and assumptions of the models suggest that the logistic regression model is preferred over other types of models. Findings suggest that there is more competitive balance, or parity, in some tournaments than others. In the women's volleyball tournament, the best teams advance further than the best teams in the other tournaments. In men's basketball, models based on Sagarin ratings are useful predictor models and findings suggest that they may be better than models based on seed strength information. Interestingly, and in contrast to what many have assumed true, there is an insignificant change over time across all four tournaments studied.

The Winning Probability and A Priori Relative Ranking in the Game of Soccer

◆ Alfred A. Akinsete, Marshall University; Nicholas J. Wintz, Marshall University

Marshall University, Dept. of Mathematics, One John Marshall Dr., Huntington, WV 25755-2560

akinsete@marshall.edu

Key Words: *soccer, probability, ranking, English FA Premier League, home advantage, game*

Soccer League started in England in 1888 with 12 teams, and has grown to different league divisions with the FA Premier League (FAPLE) occupying the highest echelon. FAPLE currently consists of 20 teams, forming the set of the strongest teams in the entire league formation. The teams play against one another, juxtaposing locations at home and away, resulting in 380 games throughout the season. There are immense contributions in literature on the statistical analyses of data arising from soccer league tournaments. Of significant contribution is the influence of home advantage as a contributing factor to the performance for teams when playing on their home grounds. We are interested in the statistical analysis of the 1993-2003 FAPLE results. We investigate the significance of the distance between any pair of teams, as ranked on the league table prior to their encounter. The estimated winning probabilities of the teams are obtained, and are used in the final ranking and predicting the outcomes of the league results in 2003/2004 season. The results are intended to form the basis for further investigation of some stochastic modeling.

Where Have You Gone, Carl Yastrzemski? Multivariate Analysis of Baseball and the Triple Crown

◆ John E. Daniels, Central Michigan University

Central Michigan University, Mt. Pleasant, MI 48859

john.daniels@cmich.edu

Key Words: *multivariate simulation, baseball, Triple Crown, SAS simulation*

In the sport of baseball, a batter who wins the league batting average, home run, and RBI (runs batted in) titles all in the same year is awarded the Triple Crown. This is a rare, but not impossible event. In 104 years of baseball (The Modern Era), there have been 14 baseball players (nine American League and five National League) who have won the Triple Crown. The last man to win the Triple Crown was Carl Yastrzemski (Boston Red Sox) in 1967. Why have 36 years passed by without another Triple Crown winner? Prior to 1967, the longest period without a Triple Crown winner was 10 years. Has the game of baseball been fundamentally changed, or is it simply that the right athlete hasn't come along? Is there perhaps another explanation for this drought of Triple Crown winners? Using multivariate statistics and a little common sense, we will explore these issues along with some interesting facts about the history of our national pastime.

U-Statistics in Olympics and Baseball: Objective Scores for Profiles of Medals and Hits

◆ Knut M. Wittkowski, Rockefeller University; Lubomir Merkov, Rockefeller University

Rockefeller University, General Clinical Research Center, 1230 New York Ave. Box 322, New York, NY 10021

kmw@rockefeller.edu

Key Words: *multivariate, rank test, ordinal, U-statistics, profile, sport*

Scoring countries by their profiles of gold, silver, and bronze medals and scoring batters by their profiles of singles, doubles, triples, and home runs is traditionally done by weighted averages based on arbitrary weights. While different media ranking countries differently has drawn little attention, Sabermetricians have long debated over batting average (all hits are given equal weights) and bases per plate appearance (all hits are weighted by the number of bases reached). The consensus is that the best weighting

scheme should be somewhere between these extremes, but no solution has been found. It is not clear, however, why the best scores should be based on a linear combination, i.e., why the same weights should apply across the whole scale and over time. When applying u-statistics to multivariate ordinal data, this unrealistic restriction can be overcome. For the first time, batters and countries can be scored objectively, i.e., without the need for arbitrary weights. The new approach for profiles of graded variables is applied to the data from the 2002 Winter Olympics and the 2003 MBL.

411 New Technologies in Surveys ◬

Section on Survey Research Methods, Social Statistics Section

Thursday, August 12, 8:30 am–10:20 am

The Online Survey: Its Contributions and Potential Problems

◆ Kevin McIntyre, Saint Louis University; Hisako Matsuo, Saint Louis University; Terry Tomazic, Saint Louis University; Barry Katz, Saint Louis University

Saint Louis University, Dept. of Research Methodology, St. Louis, MO 63108

mcintykp@slu.edu

The use of the internet as a research tool has increased substantially in the past decade. Online surveys have the potential to facilitate researchers in several ways, including subject recruitment, data collection, data tabulation, and data analysis. However, careful considerations must be taken when designing online surveys to ensure their validity and representativeness. To help researchers in this process, the current paper has four primary objectives: (1) To discuss the methodological issues involved in conducting online surveys (e.g., accessing special populations, controlling testing conditions); (2) To discuss the statistical considerations involved in using large samples for hypothesis testing (e.g., p values vs. effect sizes); (3) To review the literature on the use of internet-based surveys; and (4) To review the features and limitations of several software programs used in the creation of online surveys. Finally, the paper discusses the authors' experiences and lessons learned following the completion of two large-scale internet surveys.

The Net Effect: A Comparison of Internet vs. Mail Survey Respondents

◆ Esther M. Friedman, Mathematica Policy Research, Inc.; Nancy A. Clusen, Mathematica Policy Research, Inc.; Michael Hartzell, U.S. Dept. of Defense

Mathematica Policy Research, Inc., 600 Maryland Ave. SW, Suite 550, Washington, DC 20007

efriedman@mathematica-mpr.com

Key Words: *web survey, internet, Health Care Survey of Department of Defense Beneficiaries*

As computer access becomes increasingly prevalent, web surveys have become increasingly popular. The primary advantage to using the Internet over more traditional modes of data collection is that web surveys eliminate the need for an interviewer or data-enterer and, consequently, reduce costs and potential for error. However, not everyone has internet access and levels of computer literacy vary, so important subsets of the population may be excluded if an internet survey is the exclusive mode of data collection. The Health Care Survey of Department of Defense Beneficiaries, traditionally a mail survey, recently gave respondents the option of completing the survey over the Internet. This paper describes the characteristics of beneficiaries who chose to respond online and compares them to those of beneficiaries who responded by mail to determine if web respondents differ from mail respondents and whether particular subsets of the population prefer to respond online. Differences in key estimates are also examined as an indicator of whether response bias may be introduced in an estimate due to the mode of administration effect.

Data Collection Mode Effects Controlling for Sample Origins in an Internet Panel Survey

◆ J. Michael Dennis, Knowledge Networks, Inc.; Cindy Chatt, University of Nebraska

Knowledge Networks, Inc., 1350 Willow Rd., Suite 102, Menlo Park, CA 94404

mdennis@knowledgenetworks.com

Key Words: *mode effects, Knowledge Networks, internet panel, telephone*

We evaluate telephone- and internet-based modes of survey data collection by controlling for sample origin. Previous research has focused on sample effects only. Our main result is that substantive response differences are primarily associated with mode of data collection and not with sample origin. Sample origin is controlled by conducting both internet and telephone interviews with members of the Knowledge Networks (KN) web-enabled panel. The survey, which was sponsored by RTI International, measures policy and civic attitudes regarding 9/11 in early 2002, and was designed by RTI International and the Odum Institute at the University of North Carolina. The survey analysis is based on 2,979 web interviews with KN panelists, 300 telephone interviews with KN panelists, and 600 telephone interviews with persons that refused to join the KN panel or else take the web panel survey. The differences caused by mode in this internet vs. telephone study were strikingly similar to the telephone versus mail mode effects found in civic attitude studies by Tarnai and Dillman and in telephone versus face-to-face mode effects by Krysan. These studies found a tendency (which we confirm) for telephone respondents to answer on the extreme positive end of the scale. The internet respondents are more likely than both telephone sample groups to use the full range of scales.

Web, Mail, and Mixed-mode Data Collection in a Survey of Advanced Technology Program Applicants

◆ Jeffrey Kerwin, Westat; Pat D. Brick, Westat; Kerry Levin, Westat; David Cantor, Westat; Jennifer O'Brien, Westat; Andrew Wang, National Institute of Standards and Technology; Stephen Campbell, National Institute of Standards and Technology; Stephanie Shipp, National Institute of Standards and Technology

Westat, 1650 Research Blvd., RA 1250, Rockville, MD 20850

jeffreykerwin@westat.com

Key Words: *web surveys, internet surveys, mail surveys, survey mode*

Web surveys are becoming an increasingly popular method of data collection. In some cases, they are replacing surveys previously conducted by mail. For example, the NIST Advanced Technology Program's Survey of ATP Applicants 2002 is currently underway predominantly as a web survey, whereas a survey of year 2000 applicants was conducted by mail. But questions remain regarding the data quality of web surveys. Several published experiments have found that comparable surveys conducted by web and mail often yield different response rates. The two modes can also differ with respect to item nonresponse, and length of answers to open-ended questions. We present findings from an experiment comparing three modes of data collection in a survey of 778 companies that applied for R&D funding from the ATP program in 2002. The modes include web, mail, and web combined with mail follow-up. All three conditions include follow-up of nonrespondents by telephone. Outcomes of interest include response rates before and after telephone follow-up, item nonresponse, response distributions, length of answers to an open-ended question, and level of effort taken to complete the survey.

Protocol for Converting Respondents from Touchtone to the Internet in the Current Employment Statistics Program

◆ Richard Rosen, Bureau of Labor Statistics; Louis Harrell, Bureau of Labor Statistics; David Ballard, Bureau of Labor Statistics; Ric Wise, Bureau of Labor Statistics; Tony Gomes, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 4860, Washington, DC 20212

rosen.richard@bls.gov

Key Words: *touchtone, internet, contact method*

Many survey organizations are beginning to add internet as a mode of collection. For panel surveys, this often involves transitioning existing units from their current mode of reporting to internet. The Current Employment Statistics (CES) currently has over 75,000 sample units that self-report using their touchtone phone. These sample units would appear to be prime candidates for internet reporting since they are already self-reporting. The primary purpose of this study is to determine the most effective protocol for converting these touchtone respondents to the internet as well as to measure the pool of eligible and willing respondents. Several alternative conversion/contact protocols are tested including, telephone, mail, fax, and a special message on the touchtone collection system. A test sample will be selected from current touchtone respondents and divided among the four contact methods. We will evaluate the relative effectiveness of conversion for each mode in terms of initial conversion rates and ongoing response rates.

Computer Assisted Survey Data Collection for Geographic Features

◆ Sarah Nusser, Iowa State University

Iowa State University, Center for Survey Statistics and Methodology, Dept. of Statistics, Ames, IA

nusser@iastate.edu

Key Words: *CASIC, survey methods, data collection, geospatial data, GIS, GPS*

Survey data often include characteristics of geographic features recorded as codes (e.g., type of land, block ID), text (e.g., address), or quantitative measurements (e.g., surface area of a tract of land). It is now possible to digitally capture spatial attributes (e.g., coordinates, boundaries) in addition to alphanumeric attributes of geographic features during data collection. Geospatial software tools used to record spatial data objects are quite complex relative to computer-assisted survey instruments, and very little is known about how tools and interfaces should be configured to minimize survey error. This paper will describe findings from recent studies in developing computer-assisted data collection systems for recording boundaries of land types. In one study, agricultural interviewers, who had not previously been exposed to computer-assisted data collection, used pen-based tablet computers to delineate field boundaries on digital photographs. The second investigation considered interface designs for collecting boundaries of land cover/use types within an area segment, as well as how changes over time should be recorded when the segment is revisited.

412 Surveys of Special Population ⟁

Section on Survey Research Methods, Social Statistics Section

Thursday, August 12, 8:30 am–10:20 am

Survey Errors and Survey Costs: Experience from Surveys of Arrestees

◆ Yonghe M. Yang, NORC, University of Chicago

NORC, University of Chicago, 1350 Connecticut Ave., NW, Suite 500, Washington, DC 20036

yang-michael@norc.net

Key Words: *design effects, survey errors, survey costs, cost models*

The Arrestee Drug Abuse Monitoring program (ADAM) collects substance abuse data from samples of newly arrested criminals in 35 counties across the United States. The general survey literature contains little discussion of design and sampling issues involving arrestees. The purpose of this paper is to describe the sample designs adopted by ADAM. In particular, it discusses ADAM design options and decisions as they are related to survey errors and survey costs. Standard survey methods designed for the general population are not feasible for studying the highly clustered, highly skewed, and highly dynamic arrestee population. Common designs (e.g., sampling in time and space) developed for surveys of various mobile populations are not readily applicable either due to

operational and budgetary constraints. Survey errors are measured by design effects arising from unequal weighting and clustering, and survey costs are estimated through explicit cost models. Nonsampling errors and administrative measures of survey quality are also considered.

A Three-stage Cluster-sampling Approach for Clinical Surveillance of Persons in Care for HIV Infection in the United States

◆ Maxine M. Denniston, Centers for Disease Control and Prevention; Mitchell L. Wolfe, Centers for Disease Control and Prevention; Amy J. Drake, Centers for Disease Control and Prevention; Patrick Sullivan, Centers for Disease Control and Prevention; Sandra Berry, RAND Corporation; Sam Bozzette, RAND Corporation; Sally Morton, RAND Corporation; Martin Frankel, Baruch College, CUNY

Centers for Disease Control and Prevention, NCHSTP/DHAP/SE/SDMB, 1600 Clifton Rd. NE, Mail Stop E48, Atlanta, GA 30333

mmd1@cdc.gov

Key Words: *study design, surveillance, cluster sampling*

Historically, the need for national data on morbidity and care of HIV-infected persons has been met by a combination of HIV/AIDS case surveillance and supplemental research projects. Although nationwide AIDS surveillance is a nearly complete census of persons with AIDS, surveillance reports contain limited behavioral risk and clinical data and collect no data on access to/use of treatment, quality of care, impact of treatment, insurance status, comorbidities, immunizations, prophylaxis, or viral resistance. Problems with prior HIV care surveys include estimating service needs in the pre-HAART era, small sample size, designs which did not prioritize the ability to calculate state-level estimates, sampling biases and the high cost of a representative national survey. We propose a three-stage cluster-sampling approach that uses AIDS/HIV surveillance data to construct a frame of HIV care providers for second-stage sampling. This approach would provide a national population-based sample of HIV-infected persons in care while reducing the expense of constructing the provider frame, which was a major cost in a similar prior study. Additionally, the design would allow for state-level estimates.

Challenges in the Redesign of the Canadian Trucking Origin/Destination Survey

◆ François Gagnon, Statistics Canada

Statistics Canada, R.H. Coats Bldg. - 11th Floor, Tunney's Pasture, Ottawa, ON K1A0T6 Canada

gagnfra@statcan.ca

Key Words: *origin/destination survey, survey redesign, automated coding*

The quarterly Canadian For-Hire Trucking Origin/Destination Survey recently underwent a major redesign in order to better satisfy the clients' needs. The redesigned survey provides annual estimates of both local and long-distance commodity transportation by large (annual revenues above $1 million) trucking companies. It employs a four-stage design, where carriers are selected at the first stage and a period of time is selected at the second stage. At the third stage, a sample of shipping documents is selected from each sampled carrier via personal, onsite visits, while at the fourth stage, a sample of shipments is selected from each shipping document. The redesigned survey better meets the clients' needs by improving the survey coverage and increasing the quality of the estimates through a larger sample size and a better methodology (more efficient sample design, automation of the coding, edit and imputation processes, use of more cost-efficient collection methods). The emphasis of the presentation will be put on the methodological challenges that were met in the development of this redesigned survey.

Sample Design for Monitoring the Data Collection Process of Large-scale Natural Resource Surveys

◆ Cristiano Ferraz, Iowa State University; Jean Opsomer, Iowa State University; Sarah Nusser, Iowa State University

Iowa State University, Dept. of Statistics, Center for Survey Statistics and Methodology, Ames, IA 50011

cferraz@iastate.edu

Key Words: *quality inspection, probability sampling*

A data quality monitoring plan is an important part of any large scale survey. The implementation of such plans requires selection of samples for inspection during the data collection process. The goal of this paper is to discuss guidelines for the design of a probabilistic sampling procedure for quality inspection purposes. In particular, the feasibility of some sample designs will be investigated as applied to the USDA's National Resources Inventory.

Sampling Issues in Transportation

◆ Promod Chandhok, Bureau of Transportation Statistics

Bureau of Transportation Statistics, 400 Seventh St. SW, Washington, DC 20590

promod.chandhok@bts.gov

Key Words: *post-stratification, travel*

Post-stratification for the National Household Travel Survey and calculating estimates of travel behavior of older Americans from the National Household Travel Survey are discussed.

Sample Frame Deduplication in the World Trade Center Health Registry: Minimizing Overcoverage and Cost

◆ Joseph Murphy, RTI International; Paul Pulliam, RTI International; Randolph Lucas, RTI International; Deborah Walker, NYC Dept. of Health and Mental Hygiene

jmurphy@rti.org

Key Words: *registry, deduplication, coverage, WTC*

The World Trade Center (WTC) Health Registry is designed to assess the health effects of the WTC disaster of September 11, 2001. It will follow those exposed to dust and fumes on 9/11 and in the ensuing weeks as the fires burned. Persons who may enroll in the Registry include those who were in lower Manhattan on 9/11; residents and school children south of Canal St.; and persons involved in rescue, recovery, or clean-up at the WTC site or Staten Island Recovery Operations between September 11, 2001 and June 30, 2002. The sample frame will include people from more than 1,000 potential overlapping list sources of individuals. To avoid overcoverage, list entries are systematically deduplicated using an algorithm to identify likely duplicates. Indeterminates are manually reviewed to assure that the same individual was not included in the sample more than once. This paper describes the process of deduplication and assesses the resulting increase in quality and reduction in cost. Respondent demographics and list sources are evaluated to determine where overcoverage may have been most problematic, had deduplication not taken place.

Sample Design for the Terrorism Risk Insurance Program Survey

◆ G. Hussain Choudhry, Westat; Mats G. Nyfjall, Westat; Marianne Winglee, Westat

Westat, 1650 Research Blvd., Rockville, MD 20850

choudhh@westat.org

Key Words: *stratified sample design, systematic sampling, optimum sample allocation, nonlinear programming, composite selection probabilities*

We describe the sample design for the Terrorism Risk Insurance Program (TRIP) survey that was conducted for the U.S. department of Treasury to estimate at the national level and for a number of domains the uptake rate and the average premium paid for the terrorism risk insurance. The domains of interest were the industry groups, geographic domains, and size categories. The sampling frame for the private sector was constructed from the Dun and Bradstreet listings of businesses and that for the state and local governments and special districts was compiled from the 2002 Census of Governments conducted by the U.S. Bureau of the Census. The sample design was a stratified design with systematic sampling of business entities from within strata. We used a nonlinear programming technique to determine the optimum sample allocation to minimize the total sample while achieving the required precision levels for the survey estimates. We determined the composite selection probabilities for the systematic sampling procedure that at least one member (headquarters or a subsidiary) of the business was selected, and constructed the sampling weights based on the composite selection probabilities.

413 New Methods in Survival Analysis

Biometrics Section, Section on Statistics in Epidemiology

Thursday, August 12, 8:30 am–10:20 am

The Development of a Frailty Model for Unexplained Heterogeneity in the Framingham Heart Study

◆ Usha S. Govindarajulu, Harvard School of Public Health; Mark E. Glickman, Boston University; Ralph B. D'Agostino, Sr., Boston University

Harvard School of Public Health, 111 Cummington St., Rm. 142, Dept. of Math and Statistics, Boston, MA 02215

ugovinda@hsph.harvard.edu

Key Words: *frailty, Framingham Heart Study, univariate, accelerated failure time*

If differential survival patterns exist among members of a population or rather, unexplained heterogeneity exists in a population, a model needs to be chosen to accurately account for this phenomenon. Frailty models are potential choices for modeling unexplained heterogeneity in a population. Due to the interest of modeling heterogeneity at the individual level, univariate frailty models are considered, namely for a set of persons in the Framingham Heart Study who had atrial fibrillation events and who were followed forward in time for the development of stroke. A new frailty model that models frailty as a function of covariates using accelerated failure time modeling is developed and evaluated on this data. The new frailty model performed considerably well, in comparison to other frailty and survival models, on the atrial fibrillation dataset. In addition, the new frailty model has an added advantage of being able to model the frailty as a function of covariates where the frailty effect is modulated by the effect of the covariates.

A Gap Time Approach for the Analysis of Interval-censored Recurrent Event Data

◆ Dan Sheng, New York University; Mimi Kim, Albert Einstein College of Medicine

New York University, 650 First Ave., 5th Floor (Biostatistics Program), New York, NY 10016

ds744@nyu.edu

Key Words: *interval censoring, recurrent events, gap time, robust inference, multivariate survival analysis*

Prospective biomedical studies frequently involve the monitoring of recurrent events. When the occurrences of the event can only be determined through periodic assessments or laboratory tests performed at pre-scheduled visits, the event times can sometimes be interval-censored. In evaluating the effect of covariates on the gap times between interval-censored recurrent events, the complicating issue is that the gap time in between two events may be doubly interval-censored. To address this problem, we propose a method based on the discrete analogue of the proportional hazards model for the marginal distribution of each event. A robust estimator for the variance-covariance matrix is utilized to account for the potential within-subject correlation in the gap times between multiple events. An extensive simulation study indicates that the proposed method yields estimates which are less biased than a midpoint imputation approach. The proposed method is applied to data from the Systemic Lupus Erythematosus National Assessment (SELENA) study.

Survival Likelihood Intervals with Censored Data

◆ Alain C. Vandal, McGill University/SMBD Jewish General Hospital; Robert Gentleman, Harvard School of Public Health; Xuecheng Liu, McGill University

McGill University/SMBD Jewish General Hospital, Dept. of Mathematics & Statistics, Burnside Hall - 805, rue Sherbrooke ouest, Montréal, PQ H3A 2K6 Canada

vandal@math.mcgill.ca

Key Words: *empirical likelihood, vertex-exchange method, interval-censored data, nonparametric estimation, EM algorithm, current status data*

The determination of a likelihood-ratio-based confidence interval for a survival function given any type of censored data can be boiled down to a constrained optimization problem. We characterize the constraint as it applies to the cumulative distribution function's (CDF) nonparametric maximum likelihood estimator (NPMLE); special care must be taken in doing so since the CDF NPMLE is defined only up to an equivalence class for every possible constraint. We represent the estimand as a discrete probability vector. Various algorithms can then be specialized to solve the constrained problem; we propose two. The constrained EM works by rescaling at every iteration. The constrained VEM applies the constraint once on subsimplices of the probability vector simplex and essentially runs an unconstrained VEM in each of them. The constrained VEM is efficient enough to compute empirical likelihood ratio curves rapidly and can be used in practice to produce likelihood intervals for the CDF NPMLE. Asymptotic coverage probabilities for these intervals is known for some cases including current status data. We illustrate the technique on a current status dataset of muscle necrosis after injury.

Flowgraph Models for Recurrent Events

◆ C. Lillian Yau, Tulane University; Aparna V. Huzurbazar, University of New Mexico

cyau@tulane.edu

Key Words: *survival analysis, saddlepoint approximations, counting process, aging, semi-Markov model, repeated falls*

Multistate models have gained popularity for analyzing event history data. Recurrent events are one example of such data. Recurrent events can be studied by observing the number of occurrences over a prescribed period of time, or by observing the waiting times between consecutive events. Semi-Markov models have played important roles in modeling the transition waiting times between events. The emphasis of a semi-Markov model is on specifying appropriate transition intensity functions. They are then converted to survival functions. However, intensity functions are never observed, and in practice, assumptions on these functions are simplified. Alternatively, flowgraph models relax these assumptions. They model the waiting time distributions directly. They are naturally apt for multivariate survival data. Flowgraphs operate on moment-generating functions. Saddlepoint approximations are employed to convert the MGFs to waiting time distributions, survival functions, and hazard funcitons. We illustrate the methods by applying a series flowgraph model to data collected from New Mexico Aging Process Study. The data consist of number of falls and survival time between falls.

Estimating Conditional Survival Probability with Cox's PH Model for Two-stage Design Studies

◆ Chi-hong Tseng, University of California, Los Angeles; Gang Li, University of California, Los Angeles

dinaemic@ucla.edu

Key Words: *survival data, two-stage design*

We present a method to estimate the conditional survival probability for a two-stage design study under the framework of Cox's Proportional Hazard Model. A typical setting involves a first-stage sample drawn from population of interest and a second-stage sample, which is a subsample of the first-stage sample. The first-stage sample usually contains cheaper, rough, or incomplete information, while the second-stage sample consists of accurate and complete information. Our proposed method combines both datasets to estimate the conditional probability and achieve higher asymptotic efficency than estimation based only on the second-stage sample. A simulation study is presented to illustrate the effectiveness of this method and applications to the SEER dataset on cancer incidence and survival are discussed.

Analysis of Case-cohort Data with Accelerated Failure-time Model

◆ Lan Kong, University of Pittsburgh

University of Pittsburgh, Dept. of Biostatistics, 313 Parron Hall, 130 DeSoto St., Pittsburgh, PA 15261

kong@pitt.edu

Key Words: *accelerated failure-time model, case-cohort design, survival data*

In a case-cohort design, covariates are assembled only for a subcohort that is randomly selected from the entire cohort, and any additional cases outside the subcohort. This design is appealing for large cohort studies of rare disease, especially when the exposures of interest are expensive to ascertain for all the subjects. We propose a statistical method for fitting the case-cohort data with a semiparametric accelerated failure time model that interprets the covariates effects as to accelerate or decelerate the time to failure. Asymptotic properties of the proposed estimators are developed and the finite sample properties are assessed via simulation studies. The efficiency of case-cohort estimators relative to the full cohort estimators is also investigated in the simulation studies.

Empirical Likelihood-based Semiparametric Inference of the Treatment Effect in the Two-sample Problem

◆ Hua Liang, St. Jude Children's Research Hospital

St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105

hua.liang@stjude.org

Key Words: *estimating equation, confidence interval, empirical likelihood ratio/function, Kaplan-Meier estimation*

To compare two samples of censored data, we propose a unified semiparametric inference for the parameter of interest when one sample is parametric and the other is nonparametric. The parameter of interest may represent, for example, a comparison of means, survival probabilities, survival competition probability. The confidence interval drawn from the semiparametric inference, which is based on the empirical likelihood principle, improves the counterpart constructed from the common estimating equation. The empirical likelihood ratio, which can be used to construct test and confidence intervals of the parameter of interest, is shown to be asymptotically chi-squared. Simulation experiments illustrate that our method based on the empirical likelihood substantially outperforms the method of the estimating equation. A real dataset is analyzed with the proposed methods.

414 Bayesian Theory

Section on Bayesian Statistical Science

Thursday, August 12, 8:30 am–10:20 am

Asymptotic Admissibility and Bayesian Estimation

◆ John W. Emerson, Yale University

Yale University, PO Box 208290, New Haven, CT 06520

john.emerson@yale.edu

Key Words: *Bayesian, asymptotic, admissibility*

Hartigan (1998) shows that when the maximum likelihood prior fails to exist, the maximum likelihood estimator is asymptotically inadmissible in Kullback-Leibler risk. However, no explicit asymptotic estimator beating maximum likelihood is exhibited. We propose a method of discovering a discrete least favorable prior leading to a continuous prior and corresponding Bayes estimator "near to" and yet improving on maximum likelihood with respect to Kullback-Leibler loss.

A Fundamental Lemma of Specifying Conditional Densities for the Existence of a Joint Density

◆ Lung-An Li, Academia Sinica

Academia Sinica, Institute of Statistical Science, Taipei, 11529 Taiwan

lali@stat.sinica.edu.tw

Key Words: *conditional density, compatibility, joint density*

Specifying conditional densities is a crucial step in Markov chain Monte Carlo methodology, which is an important tool in applied stochastically modeling and Bayesian inference, when the joint density has no closed analytical form or is hard to be implemented. However, some compatible requirements among these conditional densities are needed to guarantee the existence of a joint density of a random vector to have these specified conditional densities as its conditional densities. The lemma shown in this paper gives necessary and sufficient conditions for such compatibility requirement, and provides a solution to the uniqueness problem that the joint density is the only one for those given conditional densities. The conditions in this lemma also provide a closed form of the "u" and "v" functions in a necessary and sufficient condition of the existence problem given in an Arnold and Press (1989) paper.

Bayesian Alternatives to "Exact" P Values in Sparse Two-way Tables

◆ Murray Aitkin, University of Newcastle-upon-Tyne; Magreth Njau, University of Newcastle-upon-Tyne

University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, NE1 7RU UK

murray.aitkin@ncl.ac.uk

Key Words: *two-way tables, exact p value, Bayes analysis, Bayesian p value, likelihood ratio, likelihood ratio test*

"Exact" (conditional) p values are widely used to assess the null hypothesis of independence or treatment homogeneity in sparse two-way contingency tables, where the likelihood ratio test (LRT) may fail to have its asymptotic chi-squared distribution. The "exact" p values are based on the conditional likelihood, taking the second margin of the table as fixed. The second term in the likelihood is ignored. This paper proposes Bayesian alternatives that use the full likelihood and are therefor efficient, with Dirichlet priors on the group multinomial probabilities. A result of Dempster (1974, 1997) extended by Aitkin (1997) is used to provide Bayesian alternatives to the p value; these alternatives also provide an assessment of the validity of the LRT p value. The power of the Bayesian procedure compared to the conditional procedure is assessed by simulation for some simple cases.

Reducing Subjectivity in the Likelihood

◆ S. James Press, University of California, Riverside

University of California, Riverside, Dept. of Statistics, 2636 Statistics-Computer Building, Riverside, CA 92521

jpress@ucrac1.ucr.edu

Key Words: *subjectivity, likelihood*

This paper is about reducing subjectivity in the likelihood function for observed data. We note that the likelihood function is generally taken to be a summary of the distribution of objectively observed data. We argue that the likelihood function itself is often subjectively based, which may not be desired. Building on this notion, we model the subjectivity in the likelihood function by generalizing its definition and showing how to reduce the contribution of its subjective component. We distinguish several different types of subjectivity in the likelihood function and show with examples how the subjective components may be reduced.

Directional Approach to Assessing Likelihood Stability

◆ Michael B. Brimacombe, New Jersey Medical School - UMDNJ

New Jersey Medical School - UMDNJ, 185 S. Orange Ave. (MSB F-647), PO Box 1709, Newark, NJ 07101-1709

brimacmb@umdnj.edu

Key Words: *likelihood, Weibull, logistic regression, posterior, high dimensions*

In many modeling applications where the underlying statistical model is inherently nonlinear—for example, logistic regression and weibull based survival analysis—the stability of the likelihood function is an issue affecting the robustness of inferences in both

Bayesian and frequentist-likelihood settings. The use of directional assessment of the joint posterior density is shown to provide a useful global geometric setting for assessing these issues relevant to all likelihood-based approaches. The basic tool is a conditional-directional distribution that can be used to examine the shape of the likelihood-posterior in specific directions out from its mode. The obtained distribution can then be compared to that expected under typical asymptotic assumptions or stabilizing reparameterization. Examples drawn from logistic regression and survival analysis are discussed as are issues of graphical representation.

Bias Reduction in Stochastic Covering Problems via a Bayesian Approach with a Conjugate Prior

◆ James J. Cochran, Louisiana Tech University

Louisiana Tech University, PO Box 10318, Ruston, LA 71272

jcochran@cab.latech.edu

Key Words: *combinatorial optimization, stochastic programming, Bayesian statistics, Dirichlet distribution*

Finding a subset collection that provides optimal population coverage is a frequently encountered deterministic Integer Programming (IP) problem. A random sample is often used to formulate the IP model, which is then used to select the subsets that provide the estimated optimal population coverage. The result is a constrained combinatorial optimization problem with a fixed, known feasible space and a stochastic objective function; such problems are ubiquitous and occur in both the public and private sectors. Examples include media selection, placement of municipal services such as sirens and waste dumps, and reserve site selection. When the estimators used to derive objective function coefficients for this class of problems are not negatively biased, the associated estimate of the maximand (maximum value of the objective function) is positively biased. To demonstrate how well a Bayesian approach utilizing a conjugate prior mitigates this bias, we randomly partition large instances of this class of problem (with live data) the data into smaller, computationally manageable problems, calculate the exact bias for these smaller problems, and compare estimates generated by the Bayesian approach to these values.

The Use of Kullback-Leibler Divergences in Bayesian Sensitivity Analysis

◆ Daniel M. Farewell, Lancaster University

Lancaster University, Dept. of Mathematics and Statistics, Fylde College, Lancaster, LA1 4YF England

d.farewell@lancaster.ac.uk

Key Words: *Bayesian sensitivity analysis, Kullback-Leibler divergence*

Model checking or "criticism" is an important aspect of statistics, and for the Bayesian statistician it falls neatly into two categories: the criticism of the choice of likelihood and the criticism of the prior distribution(s). This talk examines a technique proposed by Anthony O'Hagan in Highly Structured Stochastic Systems (2003) for criticizing the choices of prior. In order to criticize the choice of prior for a parameter p, the Kullback-Leibler divergence between the (marginal) prior and posterior distributions for p is estimated, approximating the information gained about p from the data. The scale and interpretation of the K-L divergence will be discussed. Estimation is briefly considered, and an unbiased estimate (which could be routinely implemented in computer packages such as WinBUGS) is proposed. While simple to estimate, the Kullback-Leibler divergence is difficult to interpret and thus methodology for an automated "red flag" signaling posterior sensitivity to the choice of priors remains an area for future research.

415 Statistical Methods for Microarray Data

Biometrics Section, WNAR

Thursday, August 12, 8:30 am–10:20 am

The Impact of Loess Normalization on Intraslide Correlation in Microarray Data

◆ Eric R. Siegel, University of Arkansas for Medical Sciences; John Thaden, University of Arkansas for Medical Sciences; Pippa M. Simpson, Arkansas Children's Hospital

University of Arkansas for Medical Sciences, 4301 West Markham St, #781, Little Rock, AR 72205

siegelericr@uams.edu

Key Words: *microarray, normalization, correlation, variance components, loess, locally weighted regression*

Locally weighted regression (loess) is often used with demonstrated effectiveness to "normalize" or correct microarray-derived gene expression data for spot intensity differences, spatial effects, and dye bias. Less is known about its impact on systematic variation due to treatment. We used variance components analysis to study the impact of loess normalization on intraslide correlation. Our data come from SMD microarray slides gridded with DNA probing >17,000 C. elegans genes, with more than 1,000 probe species spotted at least twice per slide. After no adjustment or mixed-models adjustment by slide, we subjected the data to no normalization or to loess normalization by slide, by print-tip, and by print-tip with rescaling to a common Median Absolute Deviation. Then we estimated the intraslide correlation of each replicated probe as the ratio of its between-slide variance to the sum of its between-slide and within-slide variances. When we compared the resulting distributions, we found that all loess normalizations markedly reduced intraslide correlation. We investigate the conditions under which this would occur and how this influences treatment effect.

Clustering Gene Expression Data Based on P Values

◆ Jun Li, Rutgers University; Rebecka J. Jornsten, Rutgers University; Regina Y. Liu, Rutgers University

Rutgers University, 23689 Bpo Way, Piscataway, NJ 08854

junli@stat.rutgers.edu

Key Words: *clustering, measure of similarity, pairwise p values, combining p values*

Clustering is an important task in many statistical analyses, including the analysis of microarray gene expression data, machine learning,and information retrieval. We use the analysis of microarray gene expression data to motivate and develop a new test-based clustering methodology that can reflect the exact experimental setup under which gene expression data are collected. We group genes by testing the equality or similarity of the condition-mean vectors and condition-variances. We use the p value from this test as a measure of similarity between two (or two groups of) genes, in which a small p value indicates that the mean-vectors and/or variance-vectors differ significantly. We view this measure of similarity as less arbitrary than the existing choices such as Euclidean, correlation etc. We cluster genes using all pairwise p values, building clusters in a bottom-up or top-down manner. For validation we use Fisher's method for combining p values to assign final p values to candidate clusters. Only clusters that satisfy the standard testing criteria are retained. To illustrate our clustering methodology, we use simulated and publicly available gene expression datasets.

Assessing Statistical Significance in Microarray Experiments Using the Distance between Microarrays

◆ Douglas Hayden, Massachusetts General Hospital; David Schoenfeld, Massachusetts General Hospital

doug@gcrc.mgh.harvard.edu

Key Words: *microarray, distance matrix*

In microarray cluster analyses it is often considered evidence of a group effect when gene chips from subjects in the same group cluster together. However, formal methods of inference are needed. We develop a method using the distance matrix to test equality of the gene expression pattern and variability between factor levels in an experiment. The within-level mean distance measures the within-level variability. The between-level mean distance measures the difference in the gene expression pattern. Permutation tests are used to assess the statistical significance of these measures. Analysis can also be based on the distance between linear contrasts of chips rather than individual chips. We apply the methodology to the comparison of two protocols for whole blood leukocyte RNA isolation. This work was supported by The Inflammation and Host Response to Injury Large Scale Collaborative Research Grant, NIGMS U54 GM-62119-02.

Multiple Imputation in Microarray Data Analysis

◆ John Barnard, Cleveland Clinic Foundation; Xuejun Peng, Cleveland Clinic Foundation; Yaomin Xu, Cleveland Clinic Foundation

Cleveland Clinic Foundation, 9500 Euclid Ave./wb4, Cleveland, OH 44195

jbarnard@bio.ri.ccf.org

Key Words: *multiple inputation, microarray, missing data*

Various reasons may cause incomplete data in microarray analysis and these may lead to Missing Complete at Random, Missing at Random, or Missing not at Random. This makes it necessary to apply different missing data imputation methods accordingly. We discuss in this paper how we choose imputation methods and how we evaluate the "goodness" of specific methods. Simulations and real data application are used to demonstrate how well these multiple imputation methods are able to restore complete data and how they affect downstream statistical inference.

You Can Learn a Lot by Lookin': Insights into Affymetrix Data

◆ Terry M. Therneau, Mayo Clinic; Karla Ballman, Mayo Clinic; Ann Oberg, Mayo Clinic

Mayo Clinic, 200 First St. SW, Rochester, MN 55905

therneau.terry@mayo.edu

Key Words: *microarray, Affymetrix, backround binding*

The above advice from Yogi Berra has proven useful to our group in many a medical research study. Feller suggests that nearly all dose response curves are well fit by a logistic curve with x=log(dose) and y=log(response). Guided by this, we looked at the individual data for each probe from three publicly available spike-in experiments: (a) the Gene Logic U95A data containing 3 subexperiments with 200, 220, and 220 probes, respectively; (b) the Affymetrix U95A data with 256 probes; and (c) the Affymetrix U133A data with 498 total probes. This generated 1,394 plots on 90 pages (one page/gene), using the raw data from each experiment (no normalization or backround correction). To each plot we fit a simple model y= a + b*f(c(x-d)) where f is the logistic function. Separate curves for PM and MM are on each plot, but differ only in the location of their inflection point d. (1) The logistic fits were excellent, with an $R^2 > 0.95$. (2) Residuals are equivariant across the range of x. (3) Both PM and MM probes show strong specific binding. (4) 10% of the probes show evidence of possible cross-binding. (5) 5-10% of the probes show no activity above backround.

Estimating Misclassification Error with Small Samples

◆ Wenjiang J. Fu, Texas A&M University; Raymond J. Carroll, Texas A&M University

Texas A&M University, 447 Blocker Bldg., 3143 TAMU, College Station, TX 77843

wfu@stat.tamu.edu

Key Words: *bioinformatics, bootstrap, cross-validation, microarray data, misclassification error, small samples*

One important issue in bioinformatics is the estimation of misclassification error. Although a variety of methods have been proposed, such as cross-validation (CV), resubstitution, and bootstrap, estimation of misclassification error still remains a major challenge in small sample studies, especially in microarray studies. It is known that CV is unbiased but highly variable while resubstitution is biased. We thus propose a resampling approach—the bootstrap cross-validation (BCV) method—and assess its performance through simulations in comparison with other methods, i.e., the CV and resubstitution. The simulation results show that BCV performs consistently better than CV and the resubstitution. We will also apply the method to a microarray study to provide a comparison in real data analysis.

Semi-nonparametric Model for Cell Cycle Microarray Data

◆ Guei-Feng Tsai, Oregon State University; Annie Qu, Oregon State University

Oregon State University, Dept. of Statistics, 97331

tsai@stat.orst.edu

Key Words: *microarray, cluster, nonparametric, quadratic inference function*

A new approach is developed to deal with cell cycle microarray data. There are two kinds of correlations in cell cycle data. Measurements are certainly correlated within a gene where it is measured over cycles, and measurements could also be correlated between genes, since some genes are biologically related and regulate the same phenotypical characteristics. The proposed procedure combines cluster data analysis, the quadratic inference function method and nonparametric techniques for complex high-dimensional longitudinal data. We first perform clustering data analysis to classify genes with similar cell cycle patterns into the same class, or into a class with no cell cycle phenomena at all. We use genes within the same cluster as replicates to develop nonparametric models. To incorporate correlation of longitudinal measurements, the quadratic inference function method is applied. This also allows us to perform a goodness-of-fit test for testing whether the coefficients are time varying. This leads us to the determination of whether certain genes regulate cell cycles.

416 Beyond the Proportional Hazards Models: New Paradigms in Survival Analysis

Section on Statistics in Epidemiology, WNAR

Thursday, August 12, 10:30 am–12:20 pm

Semiparametric Box-Cox Power Transformation Models for Censored Survival Observations

◆ Tianxi Cai, Harvard University

Harvard University, Dept. of Biostatistics, 655 Huntington Ave., Boston, MA 02115

tcai@hsph.harvard.edu

Key Words: *accelerated failure-time model, prediction, resampling method, simultaneous confidence interval*

The accelerated failure time model specifies that the logarithm of the failure time is linearly related to the covariate vector without assuming a parametric error distribution. We consider the semiparametric Box-Cox transformation model, which includes the above regression model as a special case, to analyse possibly censored failure time observations. Inference procedures for the transformation and regression parameters are proposed via a resampling technique. Prediction of the survival function of future subjects with a specific covariate vector is also provided via pointwise and simultaneous interval estimates. All the proposals are illustrated with the datasets from two clinical studies.

On Collapsibility

◆ David Oakes, University of Rochester

University of Rochester, Dep.t Biostatistics and Computational Biology, 601 Elmwood Ave. Box 630, Rochester, NY 14642

oakes@bst.rochester.edu

Key Words: *accelerated life model, hazard function, multiple timescales, survival analysis*

The notion of collapsibility provides a simple approach to quantification of the contributions to the overall risk of failure due to multiple time scales; for example, the months and miles a car has been driven. An epidemiologic application could involve the influence on mortality of the age of a cigarette smoker and pack-years of cigarettes consumed. This talk will review some recent work in the area of collapsibility.

Semiparametric Regression Analysis of Mean Residual Life with Censored Survival Data

◆ Ying Qing Chen, University of California, Berkeley

University of California, Berkeley, 140 Warren Hall #7360, Berkeley, CA 94720

yqchen@stat.berkeley.edu

Key Words: *stochastic process, counting process, estimating equation, failure time, life expectancy, proportional model*

As a function of time t, mean residual life is the remaining life expectancy of a subject given survival up to t. The proportional mean residual life model, proposed by Oakes & Dasu (1990), provides an alternative to the Cox proportional hazards model to study the association between survival times and covariates. In the presence of censoring, we develop semiparametric inference procedures for the regression coefficients of the Oakes-Dasu model using martingale theory for counting processes. We also present simulation studies and an application to the Veterans' Administration lung cancer data.

417 Model-based Statistical Inference on Biological Processes

Biometrics Section, Section on Statistics in Epidemiology

Thursday, August 12, 10:30 am–12:20 pm

Stochastic Modeling of Hematopoiesis

◆ Sandra N. Catlin, University of Nevada, Las Vegas

University of Nevada, Las Vegas, Dept. of Mathematical Sciences, 4505 Maryland Pkwy., Box 454020, Las Vegas, NV 89154

catlins@unlv.nevada.edu

Key Words: *birth-death process, compartmental model, hidden Markov model, stem cell*

This talk describes the application of stochastic modeling to understanding stem cell behavior. Stem cells reside in the bone marrow and are the cells from which all the constituents of blood derive.

Hematopoiesis is the multistage process in which stem cells, through sequential division, differentiation, and maturation, give rise to all types of circulating blood cells. Because stem cells are difficult to identify, their behavior (e.g. rates of self-replication and differentiation) must be inferred from observations of partially differentiated progenitor cells. A two-compartment hidden Markov model has been proposed to describe samples of cells representative of this stage of development taken over time from female Safari cats. We review several methods of parameter estimation in the model, and the inherent difficulties involved. We show how the model can be used to gain insight into stem cell related diseases, such as chronic myelogenous leukemia, and possible therapeutic strategies.

Random Effects Models for the Transmission of Dynamic Mutations

◆ Richard Huggins, La Trobe University; Minh Bui, La Trobe University; Guoqi Qian, La Trobe University; Danuta Loesch, La Trobe University

La Trobe University, Dept. of Statistical Science, Bundoora, 3086 Australia

r.huggins@latrobe.edu.au

Key Words: *MCMC, model selection, hierarchical Bayes*

Hierarchical models for the transmission of the CGG repeat sequence associated with the fragile X dynamic mutation in the FMR1 gene are developed. The models incorporate both genetic and family effects on the expansion rate of the sequence as random effects. The emphasis is on comparing models for the distribution of the residuals.

Inference on Cell Proliferation from Clonal Data

◆ Ollivier Hyrien, University of Rochester Medical Center

University of Rochester Medical Center, Dept. of Biostatistics and Computational Biology, 601 Elmwood Ave., Box 630, Rochester, NY 14642

Ollivier_Hyrien@urmc.rochester.edu

Key Words: *branching process, cell proliferation, simulation-based inference*

We present a method for statistical inference on the proliferation of cell clones composed of two (or more) distinct types of cells. A stochastic model designed to analyze experimental data on the development of cell clones is presented. The proposed model is an extension of the traditional multitype Bellman-Harris branching process allowing for nonidentical time-to-transformation distributions defined for different cell types. A simulation-based procedure has been developed for parametric inference from experimental data on cell clones under the proposed model. The model and associated methods of parametric inference are applied to the analysis of proliferation and differentiation of cultured O-2A progenitor cells that play a key role in the development of the central nervous system.

418 Risk-utility Approaches to Statistical Disclosure Limitation ▲ ⌘

National Institute of Statistical Sciences, Social Statistics Section, Section on Government Statistics, Section on Survey Research Methods

Thursday, August 12, 10:30 am–12:20 pm

Inference-based Measures of Utility for Contingency Tables

◆ Ashish Sanil, National Institute of Statistical Sciences

ashish@niss.org

The process of disseminating cross-tabulated data derived from confidential microdata involves a trade-off between protecting confidentiality while releasing as much useful information as possible. Over the years, a number of procedures have been devised that allow disclosure-protected dissemination of tabular data. These methods employ various well-established criteria to ascertain risk, and also various ways of measuring utility of the release based on relatively less rigorous notions of information loss. We propose an approach to measure the utlity of a released table (or a set of tables) based on the notion of “statistical usefulness” of the data to the end user who is interested in statistical analyses of the data (i.e., understanding the structure of associations among attributes of the dataset). This family of utility measures is founded on the extent to which the statistical properties of the released data agree with those of the original dataset. We will illustrate how a risk-utility formulation based on the inference-based measures of utility with examples.

Disclosure Risk and Data Utility for Remote-access Regression Servers

◆ Jerome Reiter, Duke University; Ashish Sanil, National Institute of Statistical Sciences; Alan F. Karr, National Institute of Statistical Sciences; Shanti Gomatam, U.S. Food and Drug Administration

Duke University, ISDS, Box 90251, Durham, NC 27708

jerry@stat.duke.edu

Key Words: *confidentiality, diagnostic, disclosure, regression, remote access*

Given the public’s ever increasing concerns about data confidentiality, in the near future statistical agencies may not be willing, or may not be legally allowed, to release any genuine microdata (data on individual units). In such a world, one microdata dissemination strategy is remote-access computer servers, to which users submit requests for output from statistical models fit using the collected data, but they are not allowed access to the genuine data. Remote servers, however, are not free from the risks of unintentional data disclosures. This paper describes these risks, and it suggests quantifiable measures of risk and data utility that can be used to specify which queries can be answered with output. This risk-utility framework is illustrated for regression models using simulated data. Methods for releasing safe and useful model diagnostics are also discussed.

Data Swapping: Disclosure Risk vs. Data Utility

◆ S. Lynne Stokes, Southern Methodist University; George Duncan, Carnegie Mellon University; Sallie Keller-McNulty, Los Alamos National Laboratory

Southern Methodist University, Dept. of Statistics, PO Box 750332, Dallas, TX 75275

slstokes@mail.smu.edu

Key Words: *confidentiality, disclosure limitation, masking*

Data swapping is a useful method for statistical disclosure limitation because of its intuitive appeal and wide use in practice. Its implementation has been largely guided by intuition because of a lack of knowledge of its impact on disclosure risk and data utility. We characterize data swapping as a matrix mask, which allows development of an R-U confidentiality map for simple implementations. This produces a way to help the information organization compare swapping procedures and parameters, such as how swapping partners are chosen or at what level to set the probability of a swap.

Risk-utility Considerations in Secure Data Integration

◆ Xiaodong Lin, NISS and SAMSI

NISS and SAMSI, PO Box 14006, Research Triangle Park, NC 27709-4006

linxd@samsi.info

Key Words: *data integration, privacy, secure computation, disclosure risk, data utility*

We show how techniques from computer science, such as secure summation, and modifications, such as secure data pooling, allow informative analyses of multiple datasets without actually performing any data integration. "Horizontally partitioned regressions" are a prominent example. A risk-utility formulation is proposed for these kinds of problems, and we describe ways in which our procedures are optimal in the sense of maximizing utility, in this case of analyses of the data, subject to constraints on disclosure risk.

419 Recent Development in Proteomics Analysis ▲ ⌘

Biopharmaceutical Section, Section on Statistical Graphics

Thursday, August 12, 10:30 am–12:20 pm

Linking and Pattern-matching in Multiple Large Data Two-way Tables

◆ S. Stanley Young, NISS; Li Liu, Aventis Pharmaceuticals; Douglas Hawkins, University of Minnesota

NISS, RTP, NC 27709

genetree@bellsouth.net

Key Words: *microarray, proteomics, metabolomics, robust SVD*

Drug discovery is coming into multiple large data sets, microarrays, protein arrays, metabolic data chemical structural descriptors. All of these datasets have many more columns than rows, n < p. We will explore various methods of pattern matching between multiple two-way tables. The benefit of the methods will be the discovery and possible validation of biological targets for drug discovery.

Death by Data Analysis? The Demise of 2D-Gel Proteomics

◆ Bert Gunter, Merck & Co., Inc.

Merck & Co., Inc., PO Box 2000, RY 33-300, Rahway, NJ 07065

bert_gunter@merck.com

Key Words: *proteomics, quality control, biotechnology, image analysis*

Modern biotechnology has made fantastic advances that yield very rich data—witness gene chips, medical imaging, and HPLC-coupled mass spectrometry. Unfortunately, quality control to remove or compensate for the many confounding systematic sources of variability in these complex measurement processes has often been lacking. An inadequate understanding of variability has led to primitive or just plain wrong data analyses and poor (but often glitzy!) data analysis software. We discuss these issues in the context of 2-D gel proteomics, a once promising technology that seems to be dying due to, among other reasons, the inability to overcome difficult fundamental quality control and data analysis issues.

Computational Strategies for Analyzing Protein Interaction Networks

◆ George Michailidis, University of Michigan

University of Michigan, 439 West Hall, 550 East University, Ann Arbor, MI 48109-1092

gmichail@umich.edu

Key Words: *protein interactions, graph theory, data integration*

Advances in high-throughput proteomics techniques, such as yeast > two-hybrid system, have produced a rapidly expanding volume of protein interaction data. We use a graph theoretic approach to discuss models that integrate protein interaction information from different experimental sources in order to produce a complete interaction network. Furthermore, we develop algorithms that aid in the identification of protein complexes.

420 Empirical Likelihood: The State of the Art ▲ ⌘

IMS

Thursday, August 12, 10:30 am–12:20 pm

Weighted Empirical Likelihood

◆ Jian-Jian Ren, University of Central Florida

University of Central Florida, Dept. of Mathematics, Orlando, FL 32816

jren@mail.ucf.edu

Key Words: *empirical likelihood, bootstrap, right-censored data, doubly censored data, interval-censored data, quantiles*

Weighted empirical likelihood is a newly developed likelihood method which can be used to construct tests and confidence intervals for various types of incomplete data, such as right-censored data, doubly censored data, interval-censored data and partly interval-censored data.So far, there has not been any published work on likelihood-based interval estimate for quantiles with doubly censored data, partly interval-censored data, or interval-censored data. We give a rather general theorem for constructing Weighted Empirical Likelihood Ratio Confidence Intervals (WELRCI) which include the mean, survival probabilities, median, quantiles, M-statistic, and trimmed mean as special cases. Also, a general theoretical coverage accuracy equation for WELRCI is established under a unified framework for various types of censored data. In particular, it is shown that the coverage accuracy of WELRCI for survival probabilities and quantiles with right-censored data is at least $O(n^{-1/2})$. Theory and simulation show that WELRCI is accurate and efficient, and the use of the n out n bootstrap in its implementation appears to give satisfactory performance in all cases considered.

Comparison of Density Functions via Empirical Likelihood

Ricardo Cao, University of La Coruna; ◆ Ingrid Van Keilegom, Universite catholique de Louvain

Universite catholique de Louvain, Institut de Statistique, Voie du Roman Pays 20, Louvain-la-Neuve, 1348 Belgium

vankeile@stat.ucl.ac.be

Key Words: *empirical likelihood, smoothing, kernel estimation, hypothesis testing*

Consider p independent random variables X_1,...,X_p with density f_1,...,f_p respectively. We consider the problem of testing the equality of f_1,...,f_p. This task is accomplished by using a local empirical likelihood approach. The asymptotic behavior of the proposed test statistic is proved and its performance is illustrated by means of some finite sample simulations. A bootstrap appoximation is proposed to better estimate the critical value of the test. Finally, the choice of the bandwidth is discussed.

Estimation of Semiparametric Mixture Models Using an Empirical Likelihood-based Algorithm

◆ Yuichi Kitamura, University of Pennsylvania

University of Pennsylvania, Dept. of Economics, 3718 Locust Walk, Philadelphia, PA 19104

ykitamur@ssc.upenn.edu

Key Words: *empirical likelihood, mixture models, EM algorithm, semiparametric models, nonparametric models, unobserved heterogeneity*

(Finite) mixture models are useful in applied econometrics. They can be used to model unobserved heterogeneity, which plays major roles in labor economics, industrial organization, and other fields. Mixtures are also convenient in dealing with contaminated sampling models and models with multiple equilibria. Most of the currently available estimation methods for mixtures are entirely parametric, or at least they usually employ parametric component likelihood functions. This paper studies estimation of models with semiparametric component distributions. First, it presents some new nonparametric identification results. It is shown that mixture models are identified under very weak assumptions that are plausible in economic applications. Second, it proposes a method to estimate a mixture model treating its component distributions semiparametrically. It constructs an appropriate empirical likelihood function for the semiparametric model. While direct maximization of the empirical likelihood function for the model is impractical, an EM-type algorithm developed in the paper solves the difficulty. Some desirable properties of the new estimator are discussed.

421 Recent Advances in Fractional Factorial Designs

Business and Economics Statistics Section, Section on Quality and Productivity

Thursday, August 12, 10:30 am–12:20 pm

Doubling and Projection: A Method of Constructing Two-level Designs of Resolution IV

◆ Ching-Shui Cheng, Academia Sinica and University of California, Berkeley

Academia Sinica and University of California, Berkeley, Institute of Statistical Science, Taipei, Taiwan

cheng@stat.sinica.edu.tw

Key Words: *minimum aberration, wordlength pattern, maximal design*

Given a two-level regular fractional factorial design of resolution IV, the method of doubling produces another design of resolution IV which doubles both the run size and the number of factors of the initial design. On the other hand, the projection of a design of resolution IV onto a subset of factors is of resolution IV or higher. Recent work in the literature of projective geometry essentially determines the structures of all regular designs of resolution IV with $n \geq N/4+1$ in terms of doubling and projection, where N is the run size and n is the number of factors. These results imply that, for instance, all regular designs of resolution IV with N/2>n>5N/16 must be projections of the regular design of resolution IV with N/2 factors. We show that for 5N/16>n>=9N/32, all minimum aberration designs are projections of the minimum aberration design with 5N/16 factors which can be constructed by repeatedly doubling the $2^{(5-1)}$ design defined by I=ABCDE. To prove this result, we also derive some properties of doubling, including an identity that relates the wordlength pattern of a design to that of its double and a result that does the same for the alias patterns of two-factor interactions.

Optimal Foldover Plans for Two-level Regular and Nonregular Orthogonal Designs

◆ William Li, University of Minnesota; Dennis K.J. Lin, Pennsylvania State University; Kenny Q. Ye, SUNY, Stony Brook

University of Minnesota, Operations and Management Sci. Dept., 321 19th Ave. S, Minneapolis, MN 55455

wli@csom.umn.edu

Key Words: *fractional length word, full foldover, generalized word length pattern, optimal foldover*

We consider optimal foldover plans for both regular and nonregular designs. For regular designs, the minimum aberration criterion is used for selecting foldover designs. Most of the optimal foldover plans are different from the traditional plans that involve reversing the signs of one or all columns. For nonregular designs, the generalized aberration criterion is used. By using the indicator function, we define words with fractional lengths. The extended word length pattern is then used to select among nonregular foldover designs. We prove that the full-foldover plan that reverses the signs of all factors is optimal for all 12-run and 20-run two-level orthogonal designs. The optimal foldover plans for all 16-run (regular and nonregular) orthogonal designs are constructed and tabulated for practical use. We introduce a new class of designs called combined-optimal designs, which minimize the aberration criterion of the combined design of the initial and foldover designs.

Model Selection in the Presence of Multiple Dispersion Effects in Unreplicated Fractional Factorials

◆ Richard N. McGrath, Bowling Green State University

Bowling Green State University, Applied Statistics and Operations Research, Bowling Green, OH 43403

rnmcgra@cba.bgsu.edu

Key Words: *generalized linear models, variance, dispersion, unreplicated, fractional factorial*

Unreplicated fractional factorials can be efficient screening designs in that they can be useful for identifying the most important factors, and, possibly, interactions. Traditionally, these designs have been used to identify location effects, those that impact the mean response. More recently, research has shown that they may be useful for identifying dispersion (variance) effects. However, when trying to identify both location and dispersion effects with $n=2^{(k-p)}$ observations, $2n=2^{(k-p+1)}$ effects are being estimated. This implies confounding among location and dispersion effect estimates. I will discuss a synthesis of existing approaches including the use of generalized linear models. New insights and limitations will be provided.

Bayesian D-optimal Supersaturated Design

◆ Bradley Jones, SAS Institute Inc.

SAS Institute Inc., Cary, NC 27513

Bradley.Jones@jmp.com

Key Words: *optimal design, supersaturated, Bayesian*

This talk introduces a new class of supersaturated designs using Bayesian D-optimality. The design generated using this approach can have arbitrary sample sizes, can have any number of blocks of any size, and can incorporate categorical factors with more than two levels. In side-by-side diagnostic comparisons based on the E(s^2) criterion for two-level experiments having even sample size, these designs either match or out-perform the best designs published to date. To show the practical benefit of the new method, we present a case study design having 15 runs in 20 factors divided into three blocks of size five.

422 Outliers in Finite Population Sampling ▲

Business and Economics Statistics Section, Section on Survey Research Methods, Social Statistics Section, Section on Physical and Engineering Sciences, SSC

Thursday, August 12, 10:30 am–12:20 pm

Influence Functions and Robust Small-area Estimation

◆ Malay Ghosh, University of Florida

University of Florida, Dept. of Statistics, Box 118545, Gainesville, FL 32611-8545

ghoshm@stat.ufl.edu

Key Words: *predictive influence function, small-area estimation, robust estimation*

The paper introduces the notion of predictive influence functions, and their role in model-based finite population sampling. In particular, we consider robust small-area estimation based on predictive influence functions.

Outlier Robust Imputation of Survey Data

Raymond Chambers, University of Southampton; ◆ Ruilin Ren, ORC Macro International Inc.

ORC Macro International Inc., 11785 Beltsville Dr., Calverton, MD 20705

Ruilin.Ren@orcmacro.com

Key Words: *outliers, missing data, imputation, robust estimation, sample survey*

Outlier robust methods of survey estimation, e.g., trimming, winsorization, are well known. However, such methods do not address the important practical problem of creating an “outlier-free” dataset for general and public use. In particular, what is required in this situation is a dataset from which an outlier robust survey estimate can be recovered by the application of standard methods of survey estimation, such as weighting. We describe an imputation method, which we call reverse calibration, to achieve this aim. This method can be used to correct gross errors in survey data, as well as to impute missing values. We compare this method with two classical missing data imputation methods, regression imputation and nearest neighbor imputation, which we modify to accommodate outliers in the survey data. The paper concludes with an evaluation of these methods based on a realistic survey dataset.

Outlier Treatment for Disaggregated Estimates

◆ Louis-Paul Rivest, Université Laval; Mike Hidiroglou, Office of National Statistics

Université Laval, Dept. of Mathematics and Statistics, Ste-Foy, PQ G1K 7P4 Canada

lpr@mat.ulaval.ca

Key Words: *outliers, Winsorization, skew distributions, efficiency comparisons, bias, mean squared error*

In many surveys the total sample is large enough for outliers to have a negligible impact on aggregated estimates. Their contributions can still be substantial at disaggregated levels, for domain estimation. In surveys repeated over time, yearly or biennial estimates are typically outlier-resistant, while monthly or quarterly estimates can vary a lot when outliers occur. This presentation suggests methods to reduce the impact of outliers on disaggregated estimates while keeping aggregated estimates unchanged. The proposed method is akin to using a “Surprise Stratum” as proposed by Leslie Kish in his 1965 book, Survey Sampling. It is implemented for a stratified random sampling plan, where the objective is to reduce the impact of outliers on stratum estimates while keeping the population estimate unchanged. In each stratum a winsorization cut-off is set; the stratum estimate is given by the stratum winsorized estimate plus the stratum share of the data values exceeding their stratum cut-offs.

423 Multiple Testing ▲ ⊞

Biometrics Section, Section on Physical and Engineering Sciences

Thursday, August 12, 10:30 am–12:20 pm

Multiple Testing in 2^k Experiments

◆ Jane Y. Chang, Bowling Green State University; Jason Hsu, The Ohio State University

Bowling Green State University, Dept. of ASOR, Bowling Green, OH 43402

changj@bgnet.bgsu.edu

Key Words: *factorial experiments, Familywise Error Rate, False Discovery Rate*

In 2^k factorial experiments a total of 2^{k-1} hypotheses on the factor effects are simultaneously tested. Most of the literature in this area has ignored the multiplicity aspect. To address the multiplicity issue, various concepts of error rate, including Familywise Error Rate and False Discovery Rate, are reviewed. Also statistical test procedures that control these error rates for 2^k factorial experiments are described.

Sample Size Calculation for Simulation-based Multiple Testing Procedures

◆ Heejung Bang, University of North Carolina; Sin-Ho Jung, Duke University; Stephen George, Duke University

University of North Carolina, 137 E. Franklin St. Suite 400, CB# 8030, Chapel Hill, NC 27599

heejung_bang@unc.edu

Key Words: *Bonferroni test, multiple outcomes, power, quality of life, sample size, simulation*

We present a simple method to calculate sample size and power using a simulation-based multiple testing procedure which gives a sharper critical value than the standard Bonferroni method. The method is especially useful when several highly correlated test statistics are involved in a multiple testing procedure. The formula for sample size calculation will be useful in designing clinical trials with multiple endpoints or correlated outcomes. We illustrate our method with a quality-of-life study for patients with early stage prostate cancer. Our method can also be used for comparing multiple independent groups.

Maximum Likelihood Estimation for the Number of True Null Hypotheses in a Multiple Two-sided Hypotheses-testing Problem

◆ Huey-Miin Hsueh, National Cheng-Chi University

National Cheng-Chi University, 64, Sec. 2, Zhi-Nan Rd., Dept. of Statistics, Taipei, 116 Taiwan

hsueh@nccu.edu.tw

Key Words: *maximum likelihood estimation, multiple testing, power, Type I error rate*

When there are many hypotheses to be tested, the risk of false positive finding in the simultaneous inference severely increases. A multiple comparison procedure (MCP), which uses a conservative adjustment in significance level of each test, is suggested for well-controlled familywise Type I error rate (FEW). However, the conservativeness of a MCP becomes substantial as the number of hypotheses increases. If the number of true null hypotheses, m0, is known, it can be used to improve the power for a MCP. In a multiple two-sided hypotheses testing problem, two types of maximum likelihood estimator (MLE) for m0, which is based on the relationship between per-comparison-wise Type I error rate (CWE) and power of each hypothesis testing procedure, will be introduced in this talk. Through intensive simulation studies, the MLE is found to have satisfactory performance and is comparable to the existent preferred method. The analysis of a real dataset from a microarray experiment is given as an illustration.

Multiple Testing in a Hierarchical Framework

◆ Aurelie Labbe, University of Waterloo

University of Waterloo, Statistics and Actuarial Science, 200 University Ave. W, Waterloo, ON N2L3G1 Canada

alabbe@math.uwaterloo.ca

Key Words: *microarray, multiple testing*

We study the problem of testing the equality of two sample means under a Bayesian framework, when the number of tests performed is large. We propose to use the posterior probability of the one-sided alternative as a measure of significance for each test. We show that under some conditions on a model associated with Jeffreys prior, this posterior probability can be seen as a conditional p value, given one of the two observed samples. When a model with an informative prior is used, the conjugate gamma model for example, we also show that it is approximately uniformly distributed under the null hypothesis. In both cases, this type of posterior probability can then be used in the spirit of a p value in some standard multiple testing procedures developed under a frequentist framework, with the advantage of working with a probability combining the prior and the observed information. The proposed methodology shows all its potential in a practical problem like microarray data analysis, where the number of hypotheses tested is large.

A Two-stage Method for Multiple-hypothesis-testing

◆ Shuhui Wen, National Taiwan University and Yuanpei University of Science and Technology; Chu-Hsing Hsiao, National Taiwan University

National Taiwan University and Yuanpei University of Science and Technology, Taiwan, R.O.C.

shwen1976@yahoo.com.tw

Key Words: *association test, Bonferroni, false positive rate, SNP marker, multiple-hypothesis-testing, two-stage*

Recent association studies often involve a large number of markers under investigation and thus lead to the problem of multiple hypothesis testing. Traditional Bonferroni's procedure is too conservative in the sense that much attention has been focused on reducing the Type I error and a strict threshold is employed due to a large number of tests considered simultaneously. The current interest, however, lies in the identification of markers mostly of mild effect. In other words, the goal may be better achieved via reducing the probability of false negatives among those with association. The two types of errors cannot be minimized at the same time. We propose a two-stage method for multiple testing procedure. The main principle is to remove those obviously unassociated markers and retain those of mild effect in the first stage, followed by a test that adjusts for the possibly incurred error. We provide the implementation of the proposed procedure and derive its theoretical success and failure rates. We also illustrate the derivation of the sample size under this design and recommend the choice between two stages. The performance of the two-stage method is evaluated.

424 Disease Specific Methodologies

Biometrics Section

Thursday, August 12, 10:30 am–12:20 pm

A Nonlinear Mixed Effect Model for Hepatitis C Viral Dynamics

◆ KyungAh Im, University of Pittsburgh; Abdus S. Wahed, University of Pittsburgh; Thelma E. Wiley, Rush University; Steven H. Belle, University of Pittsburgh

University of Pittsburgh, Graduate School of Public Health, Epidemiology Data Center, 127 Parran Hall, 130 DeSoto St., Pittsburgh, PA 15261

im@edc.pitt.edu

Key Words: *viral dynamics, pharmacokinetics, population model, hepatitis C (HCV), combination theraphy, nonlinear mixed effect model*

Treatment of patients infected with Hepatitis C Virus (HCV) using antiviral therapy such as Peginterferon or Peginterferon in combination with Ribavirin has been proven to be effective. The efficacy of these antiviral agents is often assessed by the viral levels during the treatment period and by the post-treatment sustained virological response. Compartment models such as the one described by Neumann et al. are routinely used for characterizing the HCV dynamics. The Neumann et al. model fails to take into account the pharmacokinetic behavior of the drug (especially, the interferon level in the blood serum over time). Also, the model does not explicitly include other patient characteristics such as age, race, or body mass index that might explain the variation in the viral level. The Virahep-C study recruited over 400 patients for whom viral levels and Peginterferon alfa-2a levels were measured at several time points in the first four weeks following start of treatment. We propose to extend the Neumann et al. model to incorporate the pharmacokinetic behavior of Peginterferon and other patient characteristics on HCV dynamics.

A Queueing Model for Chronic Recurrent Conditions

◆ Catherine Crespi, University of California, Los Angeles; William G. Cumberland, University of California, Los Angeles; Sally Blower, University of California, Los Angeles

University of California, Los Angeles, School of Public Health, Dept. of Biostatistics, Los Angeles, CA 90095-1772

ccrespi@ucla.edu

Key Words: *stochastic model, queueing process, hidden Markov model, virus, Bayesian inference, Markov chain Monte Carlo*

In many chronic physiological conditions, subjects alternate between an active and inactive state, and sojourns into the active state may involve multiple lesions, infections, or other recurrences with different times of onset and resolution. We present a biologically interpretable model of such chronic, recurrent conditions based on a queueing process. The model has a birth-death process describing recurrences and a semi-Markov process describing the alternation between active and inactive states. We accommodate individual heterogeneity and covariates using a random effects model, and simulate the posterior distribution of unknowns using a Markov chain Monte Carlo algorithm. Application to herpes simplex virus shows how the method can characterize the biology of the condition and estimate treatment efficacy.

Hierarchical Modeling of Glaucomatous Visual Field Progression

◆ Luohua Jiang, University of California, Los Angeles; Gang Li, University of California, Los Angeles; Joseph Caprioli, Jules Stein Eye Institute

University of California, Los Angeles, 3172 Barrington Ave., Apt. I, Los Angeles, CA 90066

lhjiang@ucla.edu

Key Words: *glaucoma progression, visual field, image analysis, spatio-temporal data, hierarchical model, mixture models*

Reliable detection of glaucomatous deterioration remains one of the most difficult problems clinicians facing in glaucoma management. Visual field data, the data used to detect glaucoma progression, are hierarchical, with 15 or more longitudinal images available for most of the eyes, and within each image, 52 locations nested in 12 clusters. Most previous analyses for this type of data focused on simple linear regression at each location of an eye, which totally ignored the hierarchical structure and spatial-temporal correlation of the data. In addition, those methods did not incorporate the information of all the eyes when making diagnostic decisions. This work develops a Bayesian approach to model all the data simultaneously. The proposed method uses a multilevel random effects model with spatial mixture distributions to classify patients. In this framework one can estimate the mean profiles of stable eyes and progressive eyes and then calculate the posterior probability for an eye being progressive, which provides a more objective criterion than previous methods. The model also allows us to assess heterogeneity in profiles among clusters and among locations for a given cluster.

Multicriteria Inference for Process Models: Structural and Parametric Inference for a Stochastic Model of Feline Hematopoiesis

◆ Joel H. Reynolds, US Fish & Wildlife Service; Daniela Golinelli, RAND Corporation

US Fish & Wildlife Service, 1011 E. Tudor Rd., MS 221, Anchorage, AK 99503

joelr@alaskalife.net

Key Words: *Pareto optimization, goodness-of-fit, stem cells, hidden Markov chain, model structure*

Stochastic process models live at the intersection of statistical and mechanistic modeling. Their assessment and calibration raise important questions regarding the appropriateness of statistical inference methods when one cannot implicitly assume a correct model specification. Such methods make parameter inference the primary objective and structural inference secondary, exactly the opposite of the early stages of process modeling. A stochastic model of feline hematoepiesis is used to compare inferences from the method of moments to those from a goodness-of-fit-based structural inference method, the Pareto Optimal Model Assessment Cycle (POMAC). Traditional statistical methods conduct parameter inference and then informally assess model structure adequacy at the selected parameterization. POMAC uses multicriteria optimization to directly assess model structure adequacy, with parameter inference a byproduct upon achieving an adequate model structure. The example model appears adequate with regard to the selected assessment criteria. Simulations based on the POMAC-based parameter estimates more closely mimic the experimental observations.

Optimal Design for Simultaneous Inference Using Pooled Assessments

◆ Joshua M. Tebbs, Kansas State University; Melinda H. McCann, Oklahoma State University

Kansas State University, Dept. of Statistics, Manhattan, KS 66506

tebbs@ksu.edu

Key Words: *confidence intervals, group testing, pairwise comparisons, HIV seroprevalence estimation, proportion, multiple comparisons*

When estimating the prevalence of a rare trait such as HIV, the use of pooled testing can confer substantial benefits when compared to individual testing. In addition to screening experiments for infectious diseases, pooled testing has been used in other applications as well, including drug-testing, multiple-vector transfer designs in plant pathology, and epidemiological studies involving animal disease. Within a pooled-testing context, we consider situations wherein different strata or treatments are to be compared with the goals of assessing practical differences between strata and ranking strata in terms of prevalence. With two competing simultaneous pairwise interval procedures, we present a unified approach to determining pool sizes which deliver desired coverage properties while taking testing costs and interval precision into account. We illustrate our methods using data from an observational HIV study.

The Relationship of the Six-minute Test to Maximal Oxygen Consumption under Skew-normal Assumptions

◆ Corrado Crocetta, Università di Foggia; Nicola Loperfido, Università di Urbino

Università di Foggia, Via IV Novembre, 1, Foggia, 71100 Italy

c.crocetta@unifg.it

Key Words: *six-minute walk test, maximal oxygen consumption, skew-normal distribution, cardiopulmonary exercise testing*

Many authors showed that Peak Oxygen Uptake (Vo2 max) is the best predictor of long-term overall and event-free survival in patient with advanced heart failure. The Vo2 max is quite expensive and sometimes it cannot even be performed. Lipkin et al. (1988) found a curvilinear relationship between distances abulated during the Six Minutes Walk Test (6-WT) and Vo2 max. Some authors used linear regression to predict Vo2 max from the results of 6-WT, under the normality assumption. We show that the assumption of normality can be misleading and that a skew-normal distribution is more appropriate when predicting Vo2 max using 6-WT.

425 Data Mining and Scaling up Statistical Methods

Section on Statistical Computing

Thursday, August 12, 10:30 am–12:20 pm

Improving Likelihood-based Data-squashing

◆ Yuchung J. Wang, SUNY, Rutgers; Xiaodong Sun, SUNY, Rutgers

SUNY, Rutgers, 370 Amitage Hall, Camden, NJ 08102

yuwang@crab.rutgers.edu

Key Words: *data mining, imputation, uniform design, clustering, representative observation*

Data-squashing is reducing a very large dataset, called the mother data, into a much smaller dataset, called a squashed data, with the requirement that statistical models/decisions derived from a squashed data are identical/similar to those derived from the mother data. A squashed data may be viewed as a collection of imputations. Madigan et al. (2003) proposed a likelihood-based squashing, which assumed a certain likelihood is the data-generating mechanism of the mother data. They artificially crate a likelihood profile for every observation in the mother data and subsequently cluster the mother data according to their profiles. Out of each cluster, an imputed observation will be constructed along with a weight. We address the issue: how to construct better profile. We shall illustrate that (a) the more the likelihood computed, the better the imputation, and (b) uniform design can improve upon factorial design in sampling the likelihood. We also propose an accelerated algorithm to perform clustering, because the k-mean clustering converges too slowly to be computationally feasible when the number of profiles is large.

The Development and Exploration of Online Classifiers for Backscatter from Denial of Service Attacks

◆ Kendall Giles, Johns Hopkins University; David J. Marchette, Naval Surface Warfare Center; Carey Priebe, Johns Hopkins University

Johns Hopkins University, 7694 Willow Point Dr., Falls Church, VA 22042

kgiles@acm.org

Key Words: *classification, backscatter, denial of service, network security*

As a motivation for understanding network attacks and designing better network security systems, this paper details methods for the investigation of denial of service attacks using statistical classification techniques. In particular, we focus on classifying what appear to be different attack types based on analysis of sensed denial-of-service attack artifacts called backscatter. We also emphasize developing models and classifiers for the purpose of detecting the onset of distributed denial-of-service attacks using sensed packets. With these goals in mind, we develop and empirically evaluate classifiers and classification methodologies, with a concentration on online classifiers. These classifiers would necessarily be useful in real-time sensor network settings.

A Statistical Foundation for Association Rules Based on Clustering

◆ Carlos Ordonez, Teradata, NCR

Teradata, NCR, 17095 Via del Campo, San Diego, CA 92127

carlos.ordonez@ncr.com

Key Words: *association rules, binary data, clustering, support bounds*

Association rules are a data-mining technique for which there exist efficient algorithms, but that as of today lack an adequate statistical foundation. Therefore, we propose a statistical model for association rules. This approach is instrumental in gaining a deeper understanding of association rules based on a simple, yet enlightening, statistical model. A dataset with binary variables is modeled as a set of clusters. Such binary data clusters are used to get tight lower and upper bounds on association support. Clusters are also used to get lower and upper bounds on association rule confidence. Then bound averages are used to estimate support and confidence. The model estimates show asymptotic increasing accuracy as the number of clusters increases. We discuss our progress on proving theoretical guarantees provided by the model.

Fast Kernel Density Estimation and Its Applications

◆ Aiyou Chen, University of California, Berkeley; Peter Bickel, University of California, Berkeley; Haihong Sheng, University of California, Davis

University of California, Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860

aychen@stat.berkeley.edu

Key Words: *kernel density estimation, profile likelihood, independent component analysis*

We provide a fast algorithm for univariate kernel density estimation, which requires exact linear time given the order statistics of random samples. This nonparametric density estimator makes the (profile) likelihood principle very practical in many semiparametric settings. We provide application examples by using this new density estimator, such as independent component analysis. Statistical efficiency will be briefly commented on.

Efficient Clustering for Principal Components

◆ Bruce Barrett, University of Alabama; Wade Watkins, University of Alabama

University of Alabama, Information Systems, Statistics, and Mgt. Science, Box 870226, Tuscaloosa, AL 35487-0226

bbarrett@cba.ua.edu

Key Words: *clustering, branch and bound, principal components*

Traditional hierarchical methods for clustering variables are typically not aimed at optimizing any particular objective function. Rather, they provide for groupings tied to the more general "similarity" information contained in the correlation matrix. However, in some applications, clustering is a preliminary step for some secondary analysis with quite specific criteria. For example, if the number of variables, p, is large, clustering might be employed prior to a principal components analysis, or as a surrogate for factor analysis. In these instances, we may wish to achieve the cluster configuration that maximizes the sum of the largest k/P eigenvalues for the k correlation submatrices, indexed by the clusters. Optimization necessarily involves consideration of each possible partitioning. We present an algorithm for an efficient search, based on a branch-and-bound procedure and compare this result with that arrived at by various hierarchical methods.

Permutation Tests in Assessing Survival Forests for Prognosis Based on Gene Profiles

◆ Van L. Parsons, National Center for Health Statistics; Thu Hoang, Université René Descartes

National Center for Health Statistics, Office of Research and Methodology, 3311 Toledo Rd., Rm 3219, Hyattsville, MD 20782

vparsons@cdc.gov

Key Words: *bootstrap, resampling, tree, survival analysis*

Combinations of survival regression trees called survival forests (SF) applied to microarray data provide both prediction of individual survival functions and the corresponding ranking of variable importance although without assessment for the latter. A basic question is whether the structures of SF and resulting statistics can be attributed to chance alone. For small datasets, we propose the use of permutation tests as a way to determine significance of the SF performance, both for assessment of the fit and to test the significance of the genes identified as prognostic markers.

A Framework for Building R-based ActiveX Components

◆ David A. James, Bell Laboratories; Duncan T. Lang, Bell Laboratories

Bell Laboratories, 600 Mountain Ave., 2C-253, Murray Hill, NJ 07974

dj@bell-labs.com

Key Words: *R language, distributed computing, Microsoft's COM*

We describe a framework for building software components for statistical computing implemented in R (*www.r-project.org*). These software components provide a transparent ActiveX (COM) interface between traditional R objects and methods and Microsoft's ActiveX clients, such as Excel, Access, and corporate applications such as Business Objects (*www.businessObjects.com*), Oracle, etc.

426 Nonparametric Methods for Censored Data

Section on Nonparametric Statistics

Thursday, August 12, 10:30 am–12:20 pm

A Class of One-step Estimators in Interval Censoring

◆ Gheorghe Doros, Yale University; David Pollard, Yale University

Yale University, Dept. of Statistics, PO Box 208290, New Haven, CT 06520-8290

gheorghe.doros@yale.edu

Key Words: *asymptotics, constrained optimization, isotonic regression, current status data, infinite-dimensional one-step procedure*

We develop an infinite-dimensional analog of the classical one-step method for solving statistical optimization problems. Specifically, we establish some asymptotic theory for estimators defined by means of convex minorants of weighted cumulative sum (CUSUM) processes. We illustrate our results by giving a new derivation of the Central Limit Theorem for the Current Status model and by proving some asymptotic results uniformly in the weights for the Weighted Isotonic Regression estimator.

Nonparametric and Semiparametric Analysis of Failure-time Data with Time-varying Group Status: Applications to HIV Genomic Data

◆ Thomas LaFramboise, Harvard University

Harvard University, Dept. of Biostatistics, School of Public Health, Boston, MA 02134

tlaframb@hsph.harvard.edu

Key Words: *time-dependent covariates, HIV genomics, survival analysis, U-statistics, interval censoring*

Studies in which we wish to analyze the effect of group status on failure time may be complicated by the fact that the group status may vary over time. Moreover, failure for each subject may sometimes be observed only at certain time points—a potentially different set of time points for each patient. For example, in an HIV study a subject's virus(es) may be genotyped at varying time points during his or her treatment. If we define failure to be the occurrence of a detectable mutation at a specific site on the HIV genome, then the situation described above applies—we can only specify an interval during which the mutation occurred. We describe non- and semi-parametric approaches to the analysis of this type of data. Rather than consider one failure time for each subject, we record time to event from each patient observation, thereby accommodating the changing value of the covariate. Constructing one-sample U-statistics in this setting, we automatically account for the obvious correlations between the multiple failure times recorded for each individual. We apply our method to HIV viral genomic data from a collection of three recent clinical trials.

Estimation with Univariate "Mixed Case" Interval-censored Data

◆ Shuguang Song, Boeing Company

Boeing Company, Applied Statistics, Phantom Works, PO Box 3707, MS: 7L-22, Seattle, WA 98124

shuguang.song@boeing.com

Key Words: *consistency, empirical process, interval censoring, maximum likelihood, rate of convergence, asymptotic distribution*

We study the Nonparametric Maximum Likelihood Estimator (NPMLE) of univariate "Mixed Case" interval-censored data in which the number of observation times, and observation time themselves are random variables. We provide a characterization of the NPMLE, then use the ICM algorithm to compute the NPMLE. We also study the asymptotic properties of the NPMLE: consistency, global rates of convergence with and without a separation condition, an asymptotic minimax lower bound, and pointwise asymptotic distribution.

Comparison of Regression Curves for Censored Responses

◆ Juan C. Pardo-Fernandez, Universidade de Vigo; Ingrid van Keilegom, Université Catholique de Louvain; Wenceslao Gonzalez-Manteiga, Universidade de Santiago de Compostela

Universidade de Vigo, Departamento de Estìstica e I.O., E.U.E.T.I., Rua Torrecedeira, 86, Vigo (Pontevedra), 36208 Spain

juancp@uvigo.es

Key Words: *comparison of regression curves, censored data, goodness of fit, bootstrap*

Consider two heteroscedastic regression models for two independent populations. When working with lifetimes it is quite common that the response variables are censored, therefore we allow for right-censoring in the response variables of our models. We introduce a procedure to test the equality of the regression functions. The test is based on a comparison of the Kaplan-Meier estimator of the censored residuals. A Kolmogorov-Smirnov-type statistic and a Cramer-von Mises type statistic are considered. Some asymptotic results are proved: weak convergence of the process of interest, convergence of the test statistics. We also describe a bootstrap procedure in order to get the critical values of the test.

Censored Regression by the Method of Average Derivatives

◆ Xuewen Lu, University of Calgary; Murray D. Burke, University of Calgary

University of Calgary, Dept. of Mathematics and Statistics, Calgary, AB T2N 1N4 Canada

lux@math.ucalgary.ca

Key Words: *asymptotic normality, average derivative, nonparametric regression, censoring, data transformation, local polynomial*

This paper proposes a technique [termed censored average derivative estimation (CADE)] for studying estimation of the unknown regression function in nonparametric regression models with randomly censored samples.The CADE procedure involves three stages: first, transform the censored data into synthetic data or pseudo-responses using the IPCW (Inverse Probability Censoring Weighted) technique; secondly, estimate the average derivatives of the regression function; and finally, approximate the unknown regression function by an estimator of univariate regression using techniques for one-dimensional nonparametric censored regression. The CADE provides an easily implemented methodology for modeling the association between the response and a set predictor variables when data are randomly censored. It also provides a technique for "dimension reduction" in nonparametric censored regression models. The average derivative estimator is shown to be root-N consistent and asymptotically normal. The estimator of the unknown regression function is a local linear kernel regression estimator and is shown to converge at the optimal one-dimensional nonparametric rate. A Monte Carlo study is presented.

Outlier Detection and Treatment in the Current Employment Statistics Survey

◆ Julie B. Gershunskaya, Bureau of Labor Statistics; Larry L. Huff, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212

gershunskaya.julie@bls.gov

Key Words: *robust estimation, one-step M-estimator, ratio estimator, influence curve*

The Current Employment Statistics (CES) Survey uses a weighted link relative estimator to make estimates of employment at various levels of industry and area detail. The estimates are produced monthly approximately three weeks after the reference date of the survey. Sometimes outliers combined with relatively large probability weights result in influential reporters that cause estimates of smaller domains to be very unstable. An employment figure reported to the survey may be considered typical for a relatively large estimation domain; however, it may be unusual and highly influential for a more detailed industry and area domain. The focus of the current simulation study is to explore feasibility of using a robust estimation technique in a simple and automated way to detect and treat outliers during the short time frame allotted for monthly survey processing. Results are evaluated based on the deviation of the estimates from the true population levels.

427 Case Studies ▲

General Methodology

Thursday, August 12, 10:30 am–12:20 pm

Clustering Models to Improve Forecasts in Retail Merchandising

◆ Nitin Patel, Cytel Software Corp.

Cytel Software Corp., 675 Massachusetts Ave., Cambridge, MA 02139

nitin@cytel.com

Key Words: *forecasting, retail merchandising, clustering algorithms*

Forecasting short-term demand for hundreds of items is a routine activity in retail merchandising. Methods used in practice employ a parametric demand model for each item. Our approach clusters items into homogenous groups and models parameter estimates for each item as following a multivariate normal distribution. We will describe two clustering algorithms based on this approach. We will report on experience with applying these algorithms to several real and simulated datasets.

Elevated Infant Death Rates a Product of Chance

◆ Winston A. Richards, Pennsylvania State University, Harrisburg

Pennsylvania State University, Harrisburg, 777 W Harrisburg Pike, Middletown, PA 17104

Ugu@psu.edu

Key Words: *probability map, CUSUM chart, Three Mile Island*

Attention has been focused on the south central Pennsylvania region since the occurrence of the nuclear accident at the Three Mile Island Electricity Power Generating Plant in March of 1979. Today, there is continuing discussion on what may have been the the adverse effects of this accident on the population and immediate environment. The probability map and a technique borrowed from quality control (the CUSUM chart) are used to investigate the observed elevation of infant death rates in Dauphin County after the accident. TMI is located in Dauphin County.

Variations in Road Accident Counts: An Explanatory Time-series Model

◆ Filip A.M. Van den Bossche, Limburgs Universitair Centrum; Tom Brijs, Limburgs Universitair Centrum; Geert Wets, Limburg University Centre

Limburgs Universitair Centrum, Universitaire Campus - Gebouw D, Diepenbeek, 3590 Belgium

filip.vandenbossche@luc.ac.be

Key Words: *time-series analysis, traffic accident counts, seasonality*

Monthly traffic accident counts show a yearly recurring pattern. The number of accidents for some months always seems to be higher or lower than average. We investigate a range of factors that can explain this variation in monthly accident counts. Apart from legal, economic and climatologic factors, monthly accident statistics are subject to calendar effects and seasonal variation. The number of accidents may vary with the number of days per month or with the number of times each day of the week occurs. Accident counts are also affected by seasonal variation. Factors like weather and holidays vary strongly with the seasons, leading to a changing level of traffic safety. Although these factors have an influence on the reported accident counts, they can never be changed to increase traffic safety and should therefore be taken into account when modelling traffic accidents. We include their effects on Belgian road accident counts in an explanatory time-series model that can be used to predict traffic safety.

A Generalized Estimating Equations Approach to Analyzing a Replicated cDNA Microarray Experiment

◆ Susan J. Simmons, University of North Carolina, Wilmington; Meng Wu, University of North Carolina, Wilmington; Ann Stapleton, University of North Carolina, Wilmington

University of North Carolina, Wilmington, NC 28403

simmonssj@uncw.edu

Key Words: *generalized estimating equations, microarray*

We investigate the use of a GEE model in a cDNA microarray experiment that contains, on average, three replicated spots per gene on each array. The microarrays are created from two different lines of maize with information regarding UV treatment versus control on each array. This methodology allows us to look at possible differences in gene expression involved in the interaction of treatment and line, as well as by treatment alone and line alone.

Standardization and Denoising Algorithms for Mass Spectra to Classify Whole-organism Bacterial Specimens

◆ Somnath Datta, University of Georgia; Glen Satten, Centers for Disease Control and Prevention; Hercules Moura, Centers for Disease Control and Prevention; Adrian Woolfitt, Centers for Disease Control and Prevention; John Barr, Centers for Disease Control and Prevention

University of Georgia, Dept. of Statistics, Athens, GA 30602

datta@stat.uga.edu

Key Words: *mass spectrometry, normalization, pre-processing, Random Forest, classification*

Application of mass spectrometry in proteomics is a breakthrough in high-throughput analyses. Early applications have focused on protein expression profiles to differentiate amongst various types of tissue samples (e.g., normal vs. tumor). We use mass spectra to differentiate between whole-organism samples of bacteria. The raw spectra are similar to spectra of tissue samples, raising some of the same statistical issues (e.g., nonuniform baselines and higher noise associated with higher baseline), but are substantially noisier. As a result, new pre-processing procedures are required before these spectra can be used for statistical classification. We introduce novel pre-processing steps that can be used with any mass spectra. These include a standardization step and a denoising step. The noise level for each spectrum is determined using only data from that spectrum. After applying these preprocessing steps, we used the Random Forest program to classify 120 mass spectra into four bacterial types. The method resulted in extremely low prediction errors in the training samples and zero prediction error in a test dataset we created from the whole dataset.

428 Engineering Applications

Section on Physical and Engineering Sciences

Thursday, August 12, 10:30 am–12:20 pm

A Survey of Design, Analysis, and Reporting of Results in Key Comparisons

◆ Adriana Hornikova, National Institute of Standards and Technology; William F. Guthrie, National Institute of Standards and Technology

National Institute of Standards and Technology, U.S. Dept. of Commerce, Statistical Engineering Division, 100 Bureau Dr., stop 8980, Gaithersburg, MD 20899-8980

adriana.hornikova@nist.gov

Key Words: *interlaboratory studies, experiment design, key comparison, MRA*

Key comparisons are international interlaboratory studies used to establish the degree of equivalence between national measurement standards. These studies, which are carried out by National Measurement Institutes (NMIs), are time-consuming, but necessary to facilitate international trade. Since the signing of the Mutual Recognition Arrangement in 1999, 27 key comparisons in a wide range of metrological areas have been completed, with results posted in the International Comparisons Database (ICDB). Now, with a rich set of real comparisons for reference, an opportunity to answer some of the questions about the theory and methods for conducting key comparisons has arisen. To take advantage of this opportunity, staff in the Statistical Engineering Division are summarizing the work on currently completed key comparisons and will develop recommendations to make it easier to carry out these comparisons in the future and to learn as much as possible from the results.

Empirical-mechanistic Modeling for Monitoring Multiproduct Manufacturing Process Applied to the Fabrication of Integrated Circuits

◆ Spencer B. Graves, PDF Solutions

PDF Solutions, 333 W. San Carlos St., Suite 700, San Jose, CA 95110

spencer.graves@prodsyse.com

Key Words: *modeling spatial variations, statistical process control (SPC), Bayesian sequential updating, Kalman filtering, monitoring low-volume, multiproduct production, optimizing preventive maintenance and equipment replacement cycles*

Many manufacturing facilities produce multiple products (sometimes with low volumes or short life cycles) using the same or similar equipment. This presentation will discuss conceptual tools for decomposing variability into product and process specific components. This includes tracing variations to the structure and geometry of a product and to specific types of equipment malfunctions. Applications include optimization of preventive maintenance and equipment replacement cycles. These concepts will be illustrated in studying spatial variations across a wafer in the fabrication of integrated circuits (ICs/computer chips). Parameters estimated from appropriate models can be monitored over time, across products, and between early and mature production, decomposing variability between product and equipment. Simple models may not be available for other phenomena such as scratches, whose presence may be more easily detected in residuals. Parameter estimation can be built into Bayesian sequential analyses/Kalman filters. Multiple Kalman filters running simultaneously can provide sensitive data-mining tools tied to manufacturing physics and management.

Statistical Modeling of Flow Instabilities in an Engineering System

◆ Carolyn B. Morgan, Hampton University; Morris H. Morgan, III, Hampton University; Adrian Cox, Hampton University

Hampton University, Queen and Tyler St.s, Hampton, VA 23668

carolyn.morgan@hamptonu.edu

Key Words: *time series, modeling, autocorrelation, analysis*

Spouted beds are popular systems for many engineering applications. Recent experimental investigations have been conducted to develop an understanding of the ranges of fluid and particle conditions that give rise to local instabilities in these and similar combusting systems. These instabilities often give rise to large bed disturbances that, under certain conditions, grow into macro-scale disturbances that affect combustion efficiencies and performance. This paper will describe the development of the two limiting ordinary differential equation models that describe such flow conditions and system performance. Prior experimental work has focused on the connection between the system fluid/particle mechanical stability and combustor operating conditions. Time series analysis is used to characterize the effect of internal process flow rates, mixer volume, reactor size, and electric current levels on system stability. The present work shows that the autocorrelation function is a useful statistical tool for characterizing the onset of such chaotic type flows. When the voidage is chaotic, information about the past behavior is lost.

Approximating Stationary Distribution of Atom Traps Using Twin Processes

◆ James H. Yen, National Institute of Standards & Technology

National Institute of Standards & Technology, 100 Bureau Dr. Stop 8980, Gaithersburg, MD 20899—898

yen@nist.gov

Key Words: *birth-death process, stochastic process, Poisson process*

Physicists are building magneto-optical traps into which they can inject a stream of atoms. Ideally, they would like to control the process, by varying the flow rate or blocking the flow, so that there is exactly one atom in the trap as much as possible. However, that they can only look to block the flow at finite time intervals complicates the stochastic analysis. We introduce a “Twin Process” that is not the same as the trap population process but has approximately the same stationary distribution.

Automated Detection of Evoked Responses Using Wavelets

◆ Jonathan D. Norton, University of Arkansas for Medical Sciences

University of Arkansas for Medical Sciences, 4301 West Markham,
#820, Little Rock, AR 72205

NortonJonathanD@uams.edu

Key Words: *wavelets, neuroscience, time-series analysis, nonstationary, signal detection*

Neural evoked responses, which are studied with such technologies as electro- and magneto-encephalography (EEG and MEG), are traditionally detected by averaging time points around the onset of the stimulus. These averages are analyzed using largely subjective criteria. Automatic, statistically valid detection of evoked responses is complicated by temporal dependence resulting from the high resolution of the EEG/MEG trace. By decorrelating the sensor time series and transforming it toward normality, the discrete wavelet transform (DWT) allows the analyst to test for an association between a stimulus and a sensor time series with appropriate degrees of freedom. Wavelet-based evoked response detection is demonstrated on a study that employed SARA (SQUID Array for Reproductive Assessment), a 151-channel fetal MEG system. Recordings were obtained from pregnant subjects while tones were played against the abdomen. The stimulus time series was compared to the MEG channels by taking the DWT of each. A nonparametric correlation was then computed between the stimulus and MEG wavelet coefficients at an appropriate scale. In this manner, a significance level was determined for each sensor.

Classification Using Time-frequency Representations

◆ Justin W. Davis, Baylor University

Baylor University, One Bear Place #97328, Waco, TX 76798

Wade_Davis@Baylor.edu

Key Words: *time-frequency representations, discriminant analysis, nonstationary, time series, wavelet packets, pattern recognition*

Techniques for discrimination and clustering of serial data have been employed in many different scientific areas. Applications include speech pattern recognition, the categorization of seismic records, and the classification of radar signals. Since many such time series exhibit nonstationary dynamics, a need exists for classification and clustering procedures that exploit temporal frequency shifts. Characterizing frequency dynamics is possible using a time-frequency representation (TFR). Various transforms can be used to obtain a TFR for a time series, and each method has its advantages and disadvantages. Several discriminants are proposed to be used in conjunction with TFRs to classify nonstationary series. The effectiveness of such methods are discussed, and an application of these methods to real data is given.

429 **Quality Control** ▲

Section on Quality and Productivity

Thursday, August 12, 10:30 am–12:20 pm

Optimal and Asymptotically Optimal CUSUM Rules for Changepoint Detection in the Brownian Motion Model with Multiple Alternatives

◆ Olympia Hadjiliadis, Columbia University; George V. Moustakides, INRIA

Columbia University, 618 Mathematics Bldg., MC 4403, 2990
Broadway Ave., New York, NY 10027

oh66@columbia.edu

Key Words: *2-CUSUM, changepoint detection, ARL, multiple alternatives*

Statistical surveillance finds application in different areas such as quality control, computer networks, monitoring of health events, etc. This work employs the Brownian motion model in which observations are taken sequentially. The objective is to detect a change in the constant drift by means of a stopping rule when there are multiple possibilities for such a change. As a performance measure an extended Lorden's criterion is proposed. The goal is to minimize the worst case detection delay subject to a false alarm constraint on the in-control ARL. When the drifts have the same sign, the CUSUM rule designed to detect the smallest in absolute value drift, is proven to be optimum. If the drifts have opposite signs then a specific 2-CUSUM rule is shown to be asymptotically optimal as the in-control ARL tends to infinity. In particular, when the drifts are equal in absolute value, the difference in performance between the unknown optimal rule and the proposed scheme remains uniformly bounded, although both quantities tend to infinity; for unequal in absolute value drifts the asymptotic optimality is even stronger since the corresponding difference tends to zero.

Distribution-free Statistical Process Control with Bootstrap Cumulative Sum Control Charts

◆ Snigdhansu Chatterjee, University of Minnesota

University of Minnesota, School of Statistics, 313 Ford Hall, 224
Church St. SE, Minneapolis, MN 55455

snigdhansu@rediffmail.com

Key Words: *CUSUM, bootstrap, distribution-free control charts*

We propose a statistical process control technique for cases where both the in-control and out-of-control distributions are unknown. We use a test statistic based on the cumulative sum conditional on the last time it was zero, and estimate a sequence of control limits using bootstrap. Classically unconditional cumulative sums are used and a single control limit is set assuming normality. The proposed procedure requires much weaker assumptions. Numerical experiments show that our method is competitive with known techniques, and quite superior to them when standard assumptions about the in-control and out-of-control distributions of the process are not satisfied.

On Detection of Changes in Count Data

◆ Emmanuel Yashchin, IBM Research
IBM Research, IBM, T.J. Watson Research Ctr., Box 218, Yorktown Heights, NY 10598
yashchi@us.ibm.com

Key Words: *categorical data, control schemes, CUSUM, sequential analysis*

The problem of monitoring count data is encountered in a wide range of practical settings. For example, in the field of manufacturing, problems of this type include monitoring defect rates, levels of contamination or data integrity. This paper discusses an approach to monitoring in situations where the underlying parameters of the model are subject to abrupt changes of unpredictable magnitude at some unknown points in time. We derive detection schemes based on the likelihood-ratio approach and discuss their performance and issues related to their design and analysis. The paper also discusses problems related to online estimation of the parameters in the presence of abrupt changes. It contains several examples that illustrate use of the proposed techniques in semiconductor industry.

New Control Charts for the Mean of the Poisson Distribution

◆ John Aleong, University of Vermont
University of Vermont, Dept. of Mathematics and Statistics, Burlington, VT 05495
jaleong@zoo.uvm.edu

Key Words: *Brown, Brown/Edgeworth expansions, simulations*

The problem of the control charts for the mean of the Poisson distribution (c-charts) will be revisited. The c-chart is based on the standard Wald confidence interval. Brown et al. (2003) and Aleong (2003) discuss the erratic behavior of the coverage probability of the Wald confidence interval. Recent results by Brown, Cai, and DasGupta (2000, 2001, 2002, 2003), have shown theoretically, using Edgeworth expansions and simulations, the eccentric behavior of the Wald confidence interval for mean of the of the natural exponential family using a quadratic variance function. Aleong (2003) gave new control charts for the binomial proportion. Using the coverage probabilities and expected length of the intervals, Brown et al. compared the Wald confidence interval with other intervals. Equal tail Jeffrey interval and the likelihood ratio interval are better than the Wald interval, which is the basis of the c and u charts. Hence the standard control limits, which are based on the Wald interval, are unlikely to represent actual three-sigma limits. The Wald intervals are particularly prone to erratic behavior.

A Comparison of Spectral Control Charts

◆ Laura A. McSweeney, Fairfield University
Fairfield University, Math and CS Dept., 1073 N. Benson Rd., Fairfield, CT 06824
lmcsweeney@mail.fairfield.edu

Key Words: *spectral control chart, periodogram, average run length, Monte Carlo simulations, statistical process control*

Statistical process control is used in industrial settings to improve productivity and the quality of goods. Extensions of the popular Shewhart control chart have been developed, and of interest are the spectral control charts used for the detection of new behavior that is periodic in nature. The first spectral control chart developed for detecting periodic behavior was proposed by Beneke, Leemis, Schlegel and Foote (1988). This control chart is based on Fisher's test statistic and as such the chart has difficulty detecting cyclic behavior corresponding to a frequency in between Fourier frequencies. Tatum (1996) introduced a modified control chart which showed improvement over the first spectral control chart. Two additional spectral control charts will be constructed based on the periodogram tests developed by Siegel (1980) and Bolviken (1983). The results of comparisons of these four control charts using Monte Carlo simulations to determine estimates of the average run lengths to detect the introduction of cyclic behavior will be presented for various conditions like simple vs. compound periodicities or Fourier frequencies vs. non-Fourier frequencies.

Statistical Process Control on Biochemical QC Data

◆ Nien Fan Zhang, National Institute of Standards and Technology
National Institute of Standards and Technology, 100 Bureau Dr., Stop 8980, Gaithersburg, MD 20899
zhang@nist.gov

Key Words: *autocorrelation, control chart, stationary process, time series*

Daily quality control (QC) measurements of common biochemical and hematological quantities were recorded during several months while methods and analyzers showed no signs of malfunctioning. Usually, it is assumed that QC data may be described as i.i.d. In this case an X chart and/or an EWMA chart are the proper control charts to use. When autocorrelation is presented, the traditional control charts may be inefficient. An alternative control chart, the EWMAST chart proposed in Zhang (1998) has been developed for stationary process data. The EWMA and the EWMAST chart were applied to each of the 11 QC data series. In six of the 11 series, significant process autocorrelations were demonstrated. The results show that the conventional EWMA chart may give false alarms in the presence of autocorrelation while the EWMAST chart gave few false alarms.

430 Generalized Regression and Linear Models ⌘

General Methodology

Thursday, August 12, 10:30 am–12:20 pm

Modeling College Application and Matriculation Likelihood Using Logistic Regression Scoring

◆ Douglas M. Andrews, Wittenberg University
Wittenberg University, Dept. of Math and Comp. Science, Springfield, OH 45501
DAndrews@Wittenberg.Edu

Key Words: *college admissions, logistic regression, modeling, scoring systems*

To help a small-college admissions staff focus its energies more efficiently, a scoring system was developed to use all available information on prospective high school students to predict which prospects are more likely to apply and to matriculate. Variable selection and weights were determined through a mixture of logistic regression and ad hoc analysis, guided by staff intuition and experience. Model factors include demographic and geographic variables, measures of academic ability, and vocational and personal interests. The resulting system predicts application propensity quite sensitively, but is much less successful at predicting matriculation.

Quantile Dispersion Graphs for Poisson Regression Models

◆ Siuli Mukhopadhyay, University of Florida; Andre I. Khuri, University of Florida

University of Florida, 103 Griffin/Floyd Hall, Dept. of Statistics, Gainesville, FL 32611-8545

siuli@ufl.edu

Unlike linear models, optimal designs for generalized linear models depend on the unknown parameters of the fitted model. This dependence problem causes a great difficulty in the construction of designs, and some prior knowledge of the parameters is needed. A graphical technique is proposed for comparing and evaluating designs for a log-linear Poisson model. Quantiles of the scaled mean square error of prediction are obtained within the region of interest, R. For a given design these quantiles depend on the parameters. Plots of the maxima and minima of the quantiles, over a subset of the parameter space, produce the so-called quantile dispersion graphs(QDGs). The use of QDGs provides a convenient technique for evaluating and comparing designs for GLMs. They provide information concerning the quality of prediction of designs and its sensitivity to the model's parameter values. The robustness of the QDGs to the form of the link function, will also be discussed.

Multiple Spherical Regression

◆ Tom Downs, Rice University; Kathy Downs, Research Statistics, Inc.

proftddowns@msn.com

Key Words: *Moebius mapping, polar decomposition, stereographic projection, vectorcardiogram*

Complex variables, Moebius mappings and stereographic projection, used historically to study non-Euclidean geometry, are used herein to construct non-Euclidean links for regressing one spatial direction on several others. Each spatial direction is a point on a unit sphere centered at the origin of an xyt rectangular coordinate system whose xy plane is treated as an extended complex plane. A link function m for regressing the mean direction of the dependent direction v on a vector u of p independent spatial directions is defined as the spherical projective image of the complex number $w = a'z/c'z$, where z is a (p+1)-vector of homogeneous coordinates for the complex projective images of the p spatial direction components of u, while a and c are linearly independent complex-valued parameter vectors. Properties and interpretations of the link function m are obtained from the polar decomposition of the 2x(p+1) matrix M with rows a′ and c′. Simple spherical regression ensues when p=1, and has been studied by Downs. Parameter estimation and inference are illustrated with vectorcardiogram data.

Evaluation of Approximate P Values in Logistic Regression

◆ Xiaohong Zhang, Iowa State University; Lynn R. LaMotte, Louisiana State University

Iowa State University, Dept. of Statistics, 308 Snedecor Hall, Ames, IA 50011

xhzhang@iastate.edu

Key Words: *approximate p value, exact p value, chi-square statistics, logistic regression*

The logistic regression model is widely used in applications where the response variable of interest takes binary values. Exact p values for testing coefficients equal to zero can be obtained from binomial probabilities. When the number of predictor variables in regression models or observations is considerably large, acquiring exact p values may cause intensive computational difficulties. In such cases, p values obtained from the chi-squared distribution provide a way of approximation. We explore the accuracy of this approximation method in this work. Fisher-Scoring method is used to simulate SAS built-in procedures to get estimated p values and test statistics. The exact p values are obtained by computing the probabilities from binomial distributions. The datasets with different numbers of predictors, different values of x, and different number of observations of each x are used as examples. Agreement of approximate p values and exact p values is good in most cases. Approximate p value is more conservative than exact p value in the overall performance. Decreasing the number of predictors orincreasing the number of observations of each x can increase the agreement.

Minimum Distance Estimation for the Logistic Regression Model

◆ Howard Bondell, Rutgers University

Rutgers University, Statistics Dept., Hill Center, Piscataway, NJ 08854

bondell@stat.rutgers.edu

Key Words: *logistic regression, robustness, minimum distance estimation, goodness of fit, case-control study*

A new class of estimation procedures for the logistic regression model is introduced. The estimates are constructed via a minimum distance approach after identifying the model under case-control sampling with a semiparametric biased sampling model. In simple logistic regression, using the weighted Cramer-Von Mises distance measures, the resulting estimators can be highly efficient, yet strictly robust in the sense that their influence function remains bounded. Hence these procedures are less sensitive than the MLE to outlying observations. The approach is shown to remain applicable if the sampling was done prospectively via its equivalence to a particular residual process in the standard setting. Based on this equivalence, the asymptotic normality of the class of procedures is derived. In order to maintain affine equivariance in multiple regression, a projection-based extension is proposed. The estimates are shown to compare favorably to the MLE and existing robust

modifications in small sample simulations and real data examples. Finally, this approach also yields a natural goodness-of-fit measure, and hence may be used to test the model assumptions.

Path Analysis with Logistic Regression

◆ Scott Menard, University of Colorado

University of Colorado, Institute of Behavioral Science #9, 910 28th St., Boulder, CO 80303

Scott.Menard@Colorado.edu

Key Words: *logistic regression, path analysis, indirect effects, nonrecursive models, recursive models*

Early attempts to blend logit analysis with path analysis for causal modeling, beginning with the work of Leo Goodman more than 30 years ago, were largely unsuccessful, abandoned because of technical difficulties in the quantification and interpretation of indirect effects involving nonquantitative variables. Now, with the development of alternative standardized coefficients for multiple logistic regression analysis, it is possible to model causal relationships using path analysis in which logistic regression is used by itself, or in a model that mixes logistic and linear regression for a combination of dichotomous, nominal, and interval/ratio/continuous variables. This makes possible an intuitively plausible quantification of indirect effects and decomposition of explained variance. The model for Path Analysis with Logistic Regression (PALR) will be illustrated, beginning with simple bivariate relationships and extended to recursive and nonrecursive models.

431 Application of Regression and Linear Models ▲

Business and Economics Statistics Section

Thursday, August 12, 10:30 am–12:20 pm

Regional Mobility and Unemployment in Germany—An Analysis with District-level Data

◆ Anette Haas, Institute for Employment Research

Institute for Employment Research, Regensburger Str. 104, D - 90478 Nuernberg, Germany

anett.haas.@iab.de

Key Words: *regional mobility, labor mobility, unemployment*

The role of mobility is central to the debate on reducing unemployment. A further question is to what extent a lack of mobility enforces regional disparities. Using a micro dataset containing information about two cohorts we analyse the impact of regional unemployment (district-level) to regional employment duration. As an alternative to the frequently used logit analysis approach we focus to duration time analyses. We use Cox Regression (Breslow Method for ties) and piecewise constant models to find out the impact of regional unemployment rate for duration working in a special region. Additionally, we could differentiate between voluntary and involuntary mobility. The results of this comparison show a contrary influence of the regional unemployment rate. Our results confirm the lower mobility of women and that the younger cohort exhibits higher mobility rates. We also compare downward/upward moves (defined as wage losses/gains after mobility) and could not find evidence for influence of regional unemployment rate to wage growth. This yields us to the conclusion that high levels of regional unemployment inhibit mobility because of a lack of vacancies.

Further Analysis of Financial and Other Factors as Predictors of Freshman-to-sophomore Retention

◆ Andrew W. Braunstein, Iona College; Donn Pescatrice, Iona College; Mary Lesser, Iona College

Iona College, Dept. of Finance, Business Economics, and Legal Studies, Hagan School of Business, 715 North Ave., New Rochelle, NY 10801

abraunstein@iona.edu

Key Words: *retention, logistic*

A previous presentation compared expected freshman-to-sophomore retention behavior for students taking a first-year experience course to actual retention results for those without exposure to such a course. Two additional years worth of data are now available. The results of logistic regression analysis using various economic, financial, academic, and demographic variables indicate that the first-year experience course to some extent achieved its goals.

The Use of Interactions in Analyzing Gender/Race Differences in Compensation

◆ Gholam H. Borhani, Economic Research Service; Chris Haan, Economic Research Service

Economic Research Service, 9009 Eagles Ridge Dr., Tallahassee, FL 32312

hborhani@ersgroup.com

Key Words: *interaction, compensation, gender, race*

Statistical techniques are often used in litigation and regulatory settings to determine whether there are statistical differences in compensation between protected group and unprotected groups of employees. Multiple regression analysis commonly used to assess differences in compensations of protected and unprotected groups after other factors that might affect compensations are taken into account. Due to the complexities involved in compensation decision-making process and the myriad factors that may affect the compensation decisions, it becomes very tempting to include many different factors in a regression model. With more than one independent variable in the model, however, it is quite possible that there may be interactions between these variables. Therefore, when the regression analysis is used to decide whether there are compensation differences the interactions between the independent variables has to be also assessed. We present a simple hypothetical example, in which we show how ignoring interactions between independent variables in the regression analysis may lead the researcher to erroneous conclusions.

Robust Estimation of Electric Utility Betas

◆ James B. McDonald, Brigham Young University

Brigham Young University, 130 Faculty Office Building, PO BOX 22363, Provo, UT 84602-2363

james_mcdonald@byu.edu

Key Words: *electric utilities, stock returns, robust estimation, beta, skewness, kurtosis*

Similarly with nonregulated firms, the distribution of electric utility stock returns and capital asset pricing model regression errors are characterized by skewness and kurtosis (fat tails). Electric utility stock returns, however, tend to be positively skewed due to lax regulation of rates of return on the book value of common equity (ROE). Electric utility rates of return on common equity have a strong lower boundary and a weak upper boundary due to lax regulation. Stock returns for unregulated firms or portfolios, such as the S&P 500 are often negatively skewed and characterized by kurtosis. This paper reviews the skewness and kurtosis of electric utility returns, and applies four flexible probability density functions (pdf) in estimating the Capital Asset Pricing Model and compares the results with the OLS and LAD estimators. Estimation based on the flexible pdf's provides the potential for improved efficiency. OLS is not the best estimator of betas. Using simulations, flexible pdf's yield results at least as efficient as OLS in the presence of skewness or kurtosis and similar results when the errors are normally distributed.

Model Selection and Parameter Estimation for Dynamic Panel Data

◆ Dylan Small, University of Pennsylvania; Tze Leung Lai, Stanford University

University of Pennsylvania, 3514 Lancaster Ave., Apt. 416, Philadelphia, PA 19104

dsmall@wharton.upenn.edu

Key Words: *dynamic panel data, model selection*

Panel data has important advantages over purely cross-sectional or time-series data in studying many economic phenomenona, because it contains information about both the intertemporal dynamics and the individuality of the entities being investigated. A commonly used class of dynamic models for panel studies identifies the parameters of interest through moment restrictions. Two important problems that arise in such models are the following: (1) it may not be clear a priori whether certain moment restrictions are valid; and (2) some of the moment restrictions may only "weakly" identify the parameters of interest, providing little information about these parameters and making inference based on conventional asymptotic theory misleading. A procedure based on empirical likelihood for choosing among possible estimators and selecting variables in this setting is developed. The advantages of the procedure over other approaches in the econometric literature are demonstrated through theoretical analysis and simulation studies.

Estimation of Partially Additive Linear Models with an Application to Gasoline Demand

◆ Dawit Zerom, University of Alberta; Sebastiano Manzan, University of Amsterdam

University of Alberta, School of Business, Room 2-43 Business Bldg., Edmonton, AB T6G 2R6 Canada

dawit@ualberta.ca

Key Words: *partially additive linear models, kernel, gasoline demand, semiparametric efficient*

This paper proposes a kernel-based procedure for the estimation of a partially additive linear model (PALM). The main contribution is the development of a semiparametric efficient estimator for the linear part of the PALM and the derivation of its asymptotic properties. Because the definition of the estimator only involves simple smoother matrices, it is computationally convenient to analyze large datasets. It is also shown that the proposed estimator is asymptotically more efficient than an estimator that ignores the additive structure. A Monte Carlo investigation indicates that the efficiency gain by the proposed estimator is more pronounced when the number of additive nonparametric components is increased; suggesting that the proposed estimator can also be used as a way to deal with the curse of dimensionality problem. The paper includes an application to U.S. gasoline demand using household data (1991 and 1994) from the Residential Transportation Energy Consumption Survey.

Inconsistency of Estimate of the Degree of Freedom of Multivariate Student-t Disturbances in Linear Regression Models

◆ Wai C. Ip, Hong Kong Polytechnic University

Hong Kong Polytechnic University, Dept. of Applied Mathematics, Hung Hom, Kowloon, Hong Kong

mathipwc@polyu.edu.hk

Key Words: *characteristic function, Chebychev's inequality, variance*

In most theoretical and applied research on linear regression models, disturbance terms are traditionally assumed to have a normal distribution. However, it is well known in practical situations that disturbance terms can have distributions with fatter tails than a normal distribution. In cases where the multivariate student-t distribution is employed, the variance of the estimates and confidence intervals of regression coefficients will depend on the degree of freedom of the t-distribution. Thus, these are not computable in practice where the degree of freedom is unknown. A moment estimator has been suggested in the literature for the degree of freedom of the jointly multivariate student-t distribution of the disturbances in a linear regression. We will show that the distribution of the moment estimate is independent of the true value of the degree of freedom and the moment estimate converges to infinite in probability as the sample size goes to infinite. Our results show that the moment estimate does not provide any information on the degree of freedom of the disturbance distribution.

432 Design of Experiments II

Section on Physical and Engineering Sciences

Thursday, August 12, 10:30 am–12:20 pm

An Empirical Comparison of Several Popular Designs for Computer Experiments

◆ Ofelia Marin, The Ohio State University; William Notz, The Ohio State University; Thomas J. Santner, The Ohio State University

The Ohio State University, Dept.t of Statistics, 1958 Neil Ave., Columbus, OH 43210

ofelia@stat.ohio-state.edu

Key Words: *EBLUP, MSPE, ML, REML, prediction accuracy, computer experiments*

We perform an empirical comparison of several popular designs for computer experiments. The comparisons are in terms of the prediction accuracy of a cubic model fit using least squares and four Empirical Best Linear Unbiased Predictors (EBLUPs) obtained by using either the Matern or Power Exponential correlation function with correlation parameters estimated by maximum likelihood (ML) or restricted maximum likelihood (REML). The prediction accuracy was determined by computing the mean squared prediction error (MSPE) over sets of 50 random surfaces obtained as slightly contaminated cubic functions, using the Hartman 6 family of test functions and as EBLUP-type interpolators generated using the krigifier of Trosset (1999) and Trosset and Padula (2000). Among the designs studied are Latin hypercube designs, uniform designs, Sobol and Niederreiter sequences, D-optimal designs, good lattice points and scrambled nets.

Design and Analysis of Probability of Visual Detection for Surface Dents

◆ Sabyasachi Basu, Boeing Company; Fritz Scholz, Boeing Company; Julio L. Peixoto, Boeing Company; Thompson Jeffrey, Boeing Company

Boeing Company, PO Box 3707, MC-7L-22, Seattle, WA 98124-2207

sabyasachi.basu@boeing.com

Key Words: *design of experiments, presenting statistical results, logistic regression, probability of detection*

We discuss the design and anlysis of the probability of visual detection (Visual POD) of the surface dents. The experimental factors chosen for this study were surface shininess and distance from which the inspections were made. Several inspectors were used in the study. The results that were obtained from the experiments and the POD curves are presented into an Excel-based tool. We will discuss the experimental design, the analysis, the results, and the contraints that we dealt with when doing this study.

Some Comments on "Staircase Method" Stress Tests

◆ Robert Kushler, Oakland University

Oakland University, Dept. of Mathematics and Statistics, Rochester, MI 48309-4485

kushler@oakland.edu

Key Words: *reliability, experimental design*

The "staircase method" involves decreasing the stress level for the next test unit if the current unit failed, and increasing the stress level otherwise. The goal of the study is to estimate the parameters and selected quantiles of the underlying strength or "tolerance" distribution. Some issues involved in the use of this method will be discussed, and some recommendations for the design of such studies will be given.

Design and Analysis of Probability of Subsurface Crack Detection

◆ Julio L. Peixoto, Boeing Company; Sabyasachi Basu, Boeing Company; Fritz Scholz, Boeing Company; Gregory Warren, Boeing Company

Boeing Company, 500 W Roy St. #W-103, Seattle, WA 98119

julio.l.peixoto@boeing.com

Key Words: *probability of detection, design of experiments, industrial statistics, aeronautics, SAS, Excel*

We discuss the methods used to obtain the optimum nonstandard design for probability of detection (POD) experiment with several factors and several inspectors. SAS was used to generate these designs. We also generated the crack lengths and locations for each of the plates used for the experiment. Results obtained from these POD experiments are presented. The models and the POD curves were captured into an Excel-based tool.

A Classification of Two-level Designs

◆ Tena I. Katsaounis, The Ohio State University

The Ohio State University, Columbus, OH 44907

katsaounis.1@osu.edu

Key Words: *factorial design, orthogonal array, balanced array, irregular design, information matrix, similarity measure*

The combinatorial properties of a design govern its statistical properties. This talk will examine a method of classifying designs into the well known-classes of orthogonal, partially balanced, PB1, and extended PB1 arrays.The classification method is based on the concept of index sets that characterize the combinatorial structure. A further concept of minimal representation will be introduced as a way to classify designs uniquely. Applications of index sets are discussed for the calculation of the elements of the information matrix and of similarity of designs.

A Repeated Measures Split-plot Experiment to Investigate the Effect of Cabin Pressure on Comfort and Well-being of Airline Passengers

◆ Stephen P. Jones, Boeing Company; I-Li Lu, Boeing Company

Boeing Company, 6709 Elderberry Ave. SE, Snoqualmie, WA 98065

stephen.p.jones@boeing.com

Key Words: *repeated measures, aircraft cabin, split-plot experiment, hypoxia*

Although there is large body of knowledge concerning the effects of altitude on humans, most of it involves healthy young people at altitudes higher than that encountered in pressurized aircraft cabins. To understand the effect of cabin pressure on passenger comfort during prolonged flight, the Boeing Company, collaborating with Oklahoma State University, conducted an experiment with people who reflect the age and gender distribution characteristic of passengers and cabin crew on commercial airlines that simulated five different interior cabin pressures. We summarize the experimental design process, discuss issues related to the implementation of the experiment and the development of statistical methods for the analysis of the data, and present the preliminary findings of the experiment.

433 Telephone Surveys ▲

Section on Survey Research Methods, Social Statistics Section

Thursday, August 12, 10:30 am–12:20 pm

Bias from Excluding Households without Telephones in Random Digit Dialing Surveys: Follow-up

◆ Mourad Touzani, Mathematica Policy Research, Inc.; John W. Hall, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., PO Box 2393, Princeton, NJ 08543

mtouzani@mathematica-mpr.com

Key Words: *coverage bias, in-person interviewing, telephone interviewing, RDD*

Surveys conducted by telephone are by their nature restricted to only households with telephone service. While a high proportion of U.S. households have telephone service, there is still concern that the exclusion of households without telephone service may result in biased survey estimates especially where those estimates are correlated with income or other factors associated with lack of telephone service. This paper builds on research using data from the first round of the Community Tracking Survey (CTS) which has an RDD component and an area probability sample for households without telephones. The earlier analysis concluded that some estimates from the CTS would be biased if interviews were not conducted among households without telephones. Using data from Round 3 of the CTS and the approach employed earlier, estimates of potential bias will be evaluated based on the size of the bias relative to the estimates and to their standard errors, and how the bias translates into aggregate estimates, such as persons affected and dollars spent. We will compare estimates of bias for Round 3 to those from Round 1 to see if the potential for bias increased or decreased.

Cellular Telephone Use Among Households with No or with Substantial Interruptions in Land-line Service

◆ John W. Hall, Mathematica Policy Research, Inc.; Mourad Touzani, Mathematica Policy Research, Inc.; Yuhong Zheng, Mathematica Policy Research, Inc.

Mathematica Policy Research, Inc., PO Box 2393, Princeton, NJ 08543

jhall@mathematica-mpr.com

Key Words: *cellular telephones, nontelephone housholds, in-person interviewing*

This paper examines the prevalence and use of cell phones among households who lack or have reported substantial interruptions in traditional land-line service and compares cell phone use in those households to that of other households with uninterrupted (or minimally interrupted) land-line telephone service. Data for this paper were collected as part an in person screening of 4,000 households for the Community Tracking Survey (CTS) to identify those with recent significant interruptions in telephone service. These households were sampled using area probability methods from a random subset of 12 (urban) sites of the 60 sites (PSUs) in the CTS sample. Due to the sampling strategy used, households screened have lower incomes than the general population. As part of screening, questions were added that asked about household ownership and use of cell phones. From these questions and others we can classify households as having no telephone service, having only cellular service, having only land-line service and having both land-line and cell service. The latter two groups can be further characterized by the length of any interruption in their land-line service.

An Assessment of Interrupted Telephone Service Adjustment

◆ Mansour Fahimi, RTI International; Lily Trofimovich, RTI International; Ali Mokdad, Centers for Disease Control and Prevention; Michelle Town, Centers for Disease Control and Prevention; William Garvin, Centers for Disease Control and Prevention; Lina Balluz, Centers for Disease Control and Prevention; Henry Wells, RTI International; Paul Levy, RTI International

RTI International, 6110 Executive Blvd., Suite 420, Rockville, MD 20852

fahimi@rti.org

Key Words: *telephone sampling, interrupted service, post-stratification, behavioral risk factors*

The exclusion of households without telephone is a source of systematic bias in telephone surveys. While lack of telephone service among households is about 5% nationally, this rate is higher among households of lower levels of socioeconomic status. Using the interruption in telephone service as a surrogate for lack of service, we have used a weighting adjustment methodology to compensate for non-telephone coverage in the 2003 Behavioral Risk Factors Surveillance System (BRFSS). Inherent to this methodology is the assumption that persons living in households

with interrupted telephone service are similar to those living in households without a telephone. We discuss the construction of the ITSA for BRFSS and evaluate its merits. We will examine the extent of variance inflation that results from applying this additional layer of adjustment and contrast that against the potential gains in bias reduction on selected variables from BRFSS.

Analysis of Nonresponse in Cross-sectional RDD Surveys

◆ Randal S. ZuWallack, ORC Macro International Inc.; Ronaldo Iachan, ORC Macro International Inc.; Leslyn Hall, ORC Macro International Inc.

ORC Macro International Inc., 126 College St., Burlington, VT 05401

randal.s.zuwallack@burlington.orcmacro.com

Key Words: *response propensity, auxiliary data, bias*

As a general problem, the measurement effect due to survey nonresponse is a difficult topic to address, primarily since little is known about the nonrespondents. One approach to evaluating unit nonresponse is to compare the sample distributions to known demographic, socioeconomic, and geographic distributions. Assuming the selected sample is representative of the population from which it was selected, deviations from the population distributions may be attributed to differential nonresponse. While inexpensive, this macrolevel approach evaluates unit nonresponse for large domains. Patterns of nonresponse may not be discovered in the expanse of the measures. Another approach to evaluating nonresponse targets individual nonresponding units directly. This microlevel approach provides highly informative information concerning the nonrespondents, but is very costly, typically involving personal follow-up visits or re-interview surveys to establish patterns of nonresponse. We discuss an alternative approach that strikes a balance between the two extremes, with application to the Department of Housing and Urban Development (HUD) Fair Market Rent Areas Surveys.

Assessing Nonresponse in Telephone Surveys

◆ Ronaldo Iachan, ORC Macro International Inc.; Randal S. ZuWallack, ORC Macro International Inc.

ORC Macro International Inc., 11785 Beltsville Dr., Calverton, MD 20705

ronaldo.iachan@orcmacro.com

Key Words: *nonresponse, RDD surveys, propensity models, longitudinal*

Assessing survey nonresponse bias is especially difficult when little is known about the nonrespondents. Nonresponse weight adjustments take advantage of geographic, demographic, and socioeconomic information known for the population to reduce the bias potential. In many telephone surveys, such as the Department of Housing and Urban Development (HUD) Areas and Regions Fair Market Rent (FMR) surveys, addressing nonresponse is further complicated by the scarcity of data that can be used for screening and survey respondents. In the HUD surveys, both survey questionnaires ask only survey eligibility and rent-related questions, and have extremely exclusive eligibility criteria. This paper addresses three questions related to nonresponse in these surveys: (a) Are the gross rent estimates biased? (b) Do we see differential nonresponse rates? And (c) Which methods can best reduce nonresponse bias? The analyses make use of Census data extracted at the tract level, and used in propensity-type models. In the Regions Surveys, a longitudinal component permits additional insights into the components of nonresponse, and additional modeling of nonresponse.

Mode Effects in the Canadian Community Health Survey: A Comparison of CAPI and CATI

◆ Martin St-Pierre, Statistics Canada; Yves Béland, Statistics Canada

Statistics Canada, R.H. Coats Bldg., 16th floor, Tunney's Pasture, Ottawa, ON K1A 0T6 Canada

martin.st-pierre@statcan.ca

Key Words: *health survey, mode effects*

The regional component of the Canadian Community Health Survey (CCHS) makes use of multiple sample frames (area and telephone) and data collection modes (CAPI and CATI). Between the CCHS 2003 and the CCHS 2001, the mix of sample frames and collection modes has changed considerably for various reasons; in fact, the CCHS 2003 had more CATI than the first survey. It is anticipated that such a change could impede the comparability of the estimates over the two surveys if differences in the response behavior are observed between the two modes of collection. Therefore, with the objective of evaluating the impact of this change on the estimates of key health-related indicators, a study of the effect of the two collection methods (CAPI and CATI) on the responses was carried on within the CCHS 2003. For the purpose of the study, 11 health regions across Canada were chosen and, for each region, two samples of similar size were selected from the same sample frame and assigned to the two modes of collection. This paper presents the methodology of the study and shows the impact of the two modes of collection on the estimates of several health-related indicators.

434 Economic and Business Topics in Government Statistics ▲

Business and Economics Statistics Section, Section on Government Statistics

Thursday, August 12, 10:30 am–12:20 pm

Generalized Beta Functions for the Distribution of Income

◆ Jungsywan Sepanski, Central Michigan University

Central Michigan University, Mathematics PAERCE214, Mount Pleasant, MI 48858

sepan1jh@cmich.edu

Key Words: *beta distribuiton, beta-normal, beta-exponential*

Many distributions have been used to model the distribution of income. This paper considers a class of generalized beta distribution including the beta-normal distribution introduced by Eugene et al. (2002) and beta-exponential distribution. The two distributions are to fit to U.S. family income and their relative performance is compared.

Missing Income Information in Panel Data: Incidence, Imputation, and Its Impact on the Income Distribution

◆ Joachim R. Frick, DIW Berlin; Markus M. Grabka, DIW Berlin

DIW Berlin, SOEP, Koenigin-Luise-Str. 5, Berlin, 14195 Germany

jfrick@diw.de

Key Words: *item nonresponse, imputation, income inequality, panel data, missing data*

Based on data from the German Socio-Economic Panel study (SOEP), this paper deals with the selectivity of missing income data due to item-nonresponse and the various longitudinal imputation strategies applied to this panel dataset. We analyze the impact of imputation on the personal income distribution and on income mobility. Comparing income inequality measures based on truly observed data to those derived from all observations, we find an increase in inequality due to imputation. This effect appears to be relevant in both tails of the distribution. Longitudinal analyses show a positive correlation of item nonresponse on income data over time, and also provide evidence of item nonresponse being a predictor of subsequent unit-nonresponse. Applying mobility indicators provide evidence of income mobility being understated using truly observed information only. Multivariate models show that survey-related factors (e.g., interview mode) and indicators for complex income structures are significantly correlated with item nonresponse. In conclusion, our results indicate the selectivity of item nonresponse on income questions in social surveys and push the necessity for adequate imputation.

Imputation for the 2002 Employee Retirement Census

◆ Loretta A. McKenzie, U.S. Census Bureau; David Kellerman, U.S. Census Bureau; Terri L. Craig, U.S. Census Bureau

U.S. Census Bureau, FOB#4 Rm 3104, 4700 Silver Hill Rd., Washington, DC

Loretta.A.McKenzie@census.gov

Key Words: *imputation, retirement*

The state and local government Public Employee Retirement Systems is an annual survey conducted by the Census Bureau. In 2002, a census of retirement systems was done as a part of the Census of Governments, which is conducted at five-year intervals. The survey coverage includes public employee retirement systems administered by state and local governments throughout the nation. The survey provides revenues, expenditures, financial assets, and membership information for public employee retirement systems. These retirement systems are a source of current and future income for thousands of retirees and maintain large volumes of investment holding. This paper discusses the imputation procedures for handling unit nonrespondents in the Public Employee Retirement Systems phase of the 2002 Census of Governments. The Retirement Survey Imputation Planning Group was formed to produce an imputed file for the 2002 Census of Governments to address problems due to a change in accounting systems, and to improve the imputation process. The Governments Accounting Standard Board practice of reporting market value rather than book value is being phased in over a number of years.

Analysis of Divergence between Chained CPI-U and Regular CPI-U for the All_US-All_Items Indexes (2000-2002)

◆ Owen J. Shoemaker, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE Rm 3655, Washington, DC 20212

shoemaker_o@bls.gov

Key Words: *superlative index, Tornqvist, stratified random group*

In February 2004, the BLS calculated and published its third annual set of C-CPI-U indexes for the 12 months of 2002. The C-CPI-U (Chained Consumer Price Index - Urban) is calculated and published every year, with a one-year lag, using a Tornqvist formula, and its set of weights are updated yearly so that a unique set of monthly weights are available for both time t as well as for time t-n. The C-CPI-U can thus be labeled a "superlative" index. By contrast the Regular CPI-U uses weights that are, at a minimum, at least two years old, and uses a combination (hybrid) of Geomeans and Laspeyres formulas as its final estimator. The set of All_US-All_Items Chained C-CPI-U index results continue to diverge (lower) from Regular CPI-U index results. We investigate the nature of this divergence. We also analyze the two different weight structures, possible response biases, and the standard errors that we calculate for these indexes.

Characteristics of Survival: Longevity of Business Establishments in the Business Employment Dynamics Data

◆ Amy E. Knaup, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Room 4840, Washington, DC 20212

knaup.amy@bls.gov

Key Words: *survival rates, births*

An examination of the U.S. economy reveals that about 60% of business establishments which opened in second quarter of 1998 were still in existence 24 months later. The largest drop in surviving businesses occurs within the first six months, when 13% of establishments are no longer operating. This paper uses the data from the new Business Employment Dynamics series to characterize opening establishments for the period of 1998-2000 according to two-digit NAICS sector, location, employment levels, ownership status, and survival rates based upon these characteristics. Survival rates across two-digit NAICS sectors vary from 53% in the educational services sector to 74% in the utilities sector. Contributions to the number of opening establishments, however, vary more widely, from one percent to nearly 50% of establishments. Both survival rates and employment depend heavily on location and ownership status, while time to peak employment seems unaffected by either.

Representations of Inequality in the Context of Free Trade and Globalization

◆ Hugh Peach, H. Gil Peach & Associates LLC; Howard Reichmuth, Scan America; Ryan N. Miller, Scan America

H. Gil Peach & Associates LLC, 16232 NW Oakhills Dr., Beaverton, OR 97006

hgilpeach@scanamerica.net

Key Words: *job structure, income inequality, globalization, graphical approach, method*

Some years ago Amstat News published an article on Lenin and statistics. Lenin used graphical methods to show the extreme inequality of land ownership in Czarist Russia. With degradation of the job structure of the U.S., due to globalization, the graphical representation of income inequality approximates the graph of land ownership in pre-revolutionary Russia. Of course, wealth and income are different but both may be taken as proximate indicators of social well-being. We contrast standard graphical approaches to representation of inequality. We introduce the “income donut” representation and then develop the method of “sector maps,” a graphical approach to demonstrate the increasing unaffordability of natural gas, electric, and water bills in the context of increasing deterioration of jobs and incomes.

The Advanced Technology Program: Evaluation of Best Practices and Results

◆ Stephanie Shipp, National Institute of Standards and Technology; Connie Chang, Advanced Technology Program; Lorel Wisniewski, ATP

National Institute of Standards and Technology, Advanced Technology Program, 100 Bureau Dr., Gaithersburg, MD 20899-4710

stephanie.shipp@nist.gov

Key Words: *Advanced Technology Program, private returns, program evaluation, R&D surveys, social returns, evaluation toolkit*

The Economic Assessment Office of the Advanced Technology Program (ATP) measures the economic impact of program funding of high-risk, enabling technologies, and seeks to increase understanding of underlying relationships between technological change and economic phenomena. The National Academy of Sciences has praised ATP’s evaluation program as “one of the most rigorous and intensive efforts of any U.S. technology programs.” ATP’s evaluation efforts were put into place for four reasons: to meet external requests for ATP program results; to use evaluation as a management tool to meet program goals and to improve program effectiveness; to understand ATP’s contribution to the U.S. innovation system, and to develop innovative methodologies to measure the impact of public R&D investment. To do this, ATP economists track progress throughout each project’s life and into the post project period by conducting surveys, compiling data, producing statistical analyses, undertaking economic studies, and commissioning external studies with consultants and research economists. This paper will describe the evolution of ATP’s evaluation activities and more.

435 Efficient Methods for Incomplete Data ▲ ⌘

SSC

Thursday, August 12, 10:30 am–12:20 pm

Estimating Survival Function under Dependent Truncation

◆ Lajmi Lakhal, Universite Laval

Universite Laval, Departement de Mathematiques et Statistique, Cite Universitaire, Quebec, PQ G1K 7P4 Canada

lakhal@mat.ulaval.ca

Key Words: *survival estimator, semi-survival copulas, dependent truncation, martingales, Archimedean copulas*

Left-truncated data arise when failure is observed only when it is greated or equal to some random quantity. The product-limit estimator for data subject to random truncation rely on the testable assumtion of quasi-independence between failure time X and truncation time T. We present a method to estimate the failure survival function S_X when it is related to the truncation time cumulative distribution function F_T by a known archimedean semi-survival copula. After defining the semi-survival copulas, we will show that the survival estimator can be expressed in terms of the unknown probability $c=P(X > T)$ estimator. This estimator as well as its asymptotic properties are presented. Finally, we will investigate our survival estimator properties.

Efficient Estimation in the Missing Censoring Indicator Model

◆ Sundar Subramanian, University of Maine

University of Maine, 5752 Neville Hall, Orono, ME 04469-5752

subraman@math.umaine.edu

Key Words: *functional delta method, Wiener process, efficiency bound, kernel density estimator, independent increments, almost sure representation*

The “inverse-probability-of-nonmissingness” estimator of a survival function is an attractive alternative to the currently existing estimators in the random censorship model when the censoring indicator is missing at random. The estimator is based on well-known empirical estimators and a kernel estimate for the conditional probability of nonmissingness of the censoring indicator. Asymptotic normality, weak convergence, asymptotic efficiency, and the approximate mean squared error of the estimator as a function of the bandwidth are discussed.

The Cost of Dichotomization: A Case Study

◆ Yodit Seifu, Novartis Pharmaceuticals

Novartis Pharmaceuticals, 1 Healthcare Plaza, East Hanover, NJ 07936-1080

yodit.seifu@pharma.novartis.com

Key Words: *logistic regression, efficiency*

Researchers desiring to model glycemic control must decide between treating HbA1C as a continuous variable or discretizing it. We compare the statistical precision obtainable by modeling HbA1C on this two scales. We analyze a dataset containing information on patients with diabetes, including HbA1C levels and various explanatory variables, including indicators of where they were treated. We show mathematically that in general one loses precision when one analyzes data on a dichotomized scale that might otherwise have been analyzed continuously. This is verified by comparing the precision of estimates from these two models. Hence, one ought to analyze similar datasets without dichotomizing the response variable.

Efficient Estimation in Regression Analysis with Missing Data in Two-phase Studies

◆ Yang Zhao, University of Waterloo; Jerald F. Lawless, University of Waterloo; Donald L. McLeish, University of Waterloo

University of Waterloo, Dept. of Statistics and Actuarial Science, 200 University Ave. W, Waterloo, ON N2L 3G1 Canada

y3zhao@uwaterloo.ca

Key Words: *two-phase study, missing data, semiparametric maximum likelihood estimator*

Regression analysis that involves incomplete observations can utilize auxiliary information to provide more efficient estimates of regression parameters. We recommend an efficient EM algorithm that can be used to estimate the parameters in regression models with some covariates missing at random. The advantages of this method in dealing with continuous missing variables with fine surrogates are discussed. Extensive simulation studies to compare it with other available method will be provided.

A Robust Transmission/Disequilibrium Test for Incomplete Family Genotypes

◆ Gulhan Alpargu, University of Massachusetts; Paola Sebastiani, Boston University; Maria M Abad, University of Granada; Marco Ramoni, Harvard Medical School

University of Massachusetts, Dept. of Mathematics and Statistics, Amherst, MA 01003

alpargu@math.umass.edu

Key Words: *TDT, SNP, missing genotype, robust inference*

Several solutions have been proposed to extend the Transmission Disequilibrium Test (TDT) to deal with missing parental genotype. All these solutions remedy the problem of missing parental genotype, while offsprings with missing genotypes are typically not taken into account. However, the treatment of such reconstructed parental genotypes can introduce bias if the underlying missing data mechanism is informative. We propose an extension to the TDT, called robust TDT (rTDT), which can cope with incomplete genotypes on both parents and children. RTDT computes minimum and maximum values of TDT that are consistent with all possible completions of the missing data. RTDT is applied to a database of markers of susceptibility to Crohn Disease, and shows that only two of the 11 markers originally associated to phenotype do not depend on assumptions about the missing data mechanism.

Estimation of Nonlinear Models with Berkson Measurement Errors

◆ Liqun Wang, University of Manitoba

University of Manitoba, Dept. of Statistics, Winnipeg, MB R3T 2N2 Canada

wangl1@cc.umanitoba.ca

Key Words: *nonlinear regression, errors in variables, method of moments, minimum distance estimator, simulation-based estimator, semiparametric models*

This paper is concerned with general nonlinear regression models where the predictor variables are subject to Berkson-type measurement errors. The measurement errors are assumed to have a general parametric distribution, which is not necessarily normal, while the distribution of the random error in the regression equation is nonparametric. A minimum-distance estimator is proposed based on the first two conditional moments of the response variable given the observed predictor variables. A simulation-based estimator is also constructed to overcome the possible computational difficulty of minimizing an objective function which involves multiple integrals. Consistency and asymptotic normality for both estimators are derived under general conditions.

436 Bayesian Estimation

Section on Bayesian Statistical Science

Thursday, August 12, 10:30 am–12:20 pm

Improved Estimation of Parameters for Linear Models with Respect to the Linex Loss Function

◆ Marvin H.J. Gruber, Rochester Institute of Technology

Rochester Institute of Technology, Dept. of Mathematics and Statistics, 85 Lomb Memorial Dr., Rochester, NY 14623

mjgsma@rit.edu

Key Words: *optimum affine estimator, comparison of efficiency, parametric functions, mean square error, translation*

Linear Bayes estimators with respect to the squared error loss function include mixed estimators, ridge- and contraction-type estimators, and other shrinkage estimators as special cases. The Bayes risk of these estimators with respect to the linex loss function may be substantially improved upon by a suitable translation. For normal priors and populations the optimum translation is subtraction of the average mean squared error. Application of these results to different kinds of shrinkage estimators will be explored.

Optimality of Median Probability Model in Generalized Linear Models

◆ Samiran Ghosh, University of Connecticut; Dipak K. Dey, University of Connecticut

University of Connecticut, 215 Glennbrook Rd., U-4120, Dept. of Statistics, Storrs, CT 06269

samiran@stat.uconn.edu

Key Words: *Bayes factor, deviance loss, Poisson regression, binomial regression, GLM, predictive distribution*

Goal of optimal model selection is of two fold. First select a model which fits the data well and second use selected model for future prediction. Under Bayesian philosophy it is commonly perceived that the optimal predictive model is the model with highest posterior probability. Recently Berger and Barberi (2003) have shown that for linear models with normal error structure, median probability model is the optimal predictive model which often differs from highest probability model. We have investigated their findings in generalized linear model (GLM). First, we have developed some optimality conditions for GLM, which will generalize the optimality theory of median probability model. In particular we have considered Binomial and Poisson case in much detail. Finally real data examples are used to illustrate the proposed methodology.

Bayesian Monotone Smoothing

◆ Farideh H. Dehkordi-Vakil, Western Illinois University

Western Illinois University, Stipes Hall 431M, Macomb, IL 61455

mffhd@wiu.edu

Key Words: *monotone smoothing, Bayesian, regression, WinBugs*

The data from many disciplines call for regression on smooth monotone functions. As is well known, the classical pool-adjacent-violators technique of monotonizing does not yield a curve that is necessary smooth, and the classical spline-fitting produces a smooth function, but it is not guaranteed to be monotone. This situation has much improved, during the recent years, thanks to the contributions of such authors as J.O. Ramsay and Xuming He and Peide Shi. The methods these authors use is based on sophisticated optimization techniques and requires codes that are specifically designed for the purpose of this problem. We present a Bayesian approach to this problem, which offers distinct advantage over the optimization methods on at least two grounds. Bayesian framework offers the possibility of a direct estimation of the regression function through sampling methods without the use of optimal criterion. This, in turn, makes our method computationally simpler, for it allows us to use the widely available software, WinBugs, developed for the implementation of Markov chain Monte Carlo method. No special code is needed.

Bayesian Estimation of Generalized Lambda Distributions Using MCMC

◆ Michele A. Haynes, University of Queensland; Robert King, University of Newcastle; Kerrie Mengersen, University of Newcastle

University of Queensland, UQ Social Research Centre, Bldg. 39A, Brisbane, 4072 Australia

m.haynes@uq.edu.au

Key Words: *generalized lambda distributions, Bayesian estimation, MCMC, posterior distributions*

The Generalized Lambda Distributions (GLD), a family of flexible distributions defined through a quantile function, have previously been estimated using moment-based methods. Recently, a transformation-based procedure for estimation known as the starship method has been shown to perform favorably. This procedure does not rely on computation of the likelihood but focuses on how well the reversely transformed data fits a base distribution (in the case of the GLD, the uniform). We investigate a Bayesian procedure for the estimation of the FMKL-parameterization of the GLD which is well defined for all values of the two shape parameters. MCMC provides a simulation-based alternative for deriving posterior estimates of the parameters through a Bayesian framework. This approach is also advantageous as it allows us to place sensible prior distributions on the parameters to govern the shape of the posterior distributions during the simulation process. We simulate sample data from symmetric and asymmetric distributions and estimate parameters for the GLD using both MCMC and starship estimation procedures for comparison.

Bayesian Estimation of a Covariance Matrix with Applications on Mixed Effects Models

◆ John S.J. Hsu, University of California, Santa Barbara; Steven C.W. Hsu, National College of Physical Education and Sports

University of California, Santa Barbara, Dept. of Statistics and Applied Probability, Santa Barbara, CA 93106

hsu@pstat.ucsb.edu

Key Words: *Laplacian approximation, hierarchical Bayes, Metropolis-Hastings MCMC algorithm*

Leonard and Hsu (1992) suggested a flexible class of prior distributions for the covariance matrix of a multivariate normal distribution. They proposed a normal prior on the log-covariance matrix and calculated the exact posterior moments using importance sampling methods. We propose a Laplacian approximation to the posterior moments. The approximation is reasonably accurate when compared with the exact result, and the calculation takes only seconds of computer time. We further consider the same prior structure for the covariance matrix in the mixed effects model. Exact posterior moments for the parameters in the model can be calculated via a Metropolis-Hastings MCMC algorithm. A subset of the High School and Beyond survey data is analyzed and will be reported.

A Piecewise Linear Generalized Poisson Regression Model for Longitudinal Data

◆ John C. Kern, II, Duquesne University

Duquesne University, 600 Forbes Ave., 440 College Hall, Pittsburgh, PA 15282

kern@mathcs.duq.edu

Key Words: *overdispersion, underdispersion, generalized Poisson, piecewise linear*

We apply a piecewise linear generalized Poisson regression model to logitudinal frequency data collected from subjects randomly allocated among experimental groups. Bayesian inference on the parameters of this model—which include the knot locations of the piecewise regression curve—is conducted using MCMC techniques. Use of the generalized Poisson distribution allows for application to overdispersed or underdispersed data. We fit this model using data from a clinical trial examining the effectiveness of acupuncture as a treatment for menopausal symptoms.

Incorporating Bayesian Prior Information into CART

◆ Ming Yin, Constella Health Sciences; Eric Harvey, Constella Health Sciences; Patrick W. Crockett, Constella Health Sciences

Constella Health Sciences, 2605 Meridian Pkwy., Suite 200, Durham, NC 27713

myin@constellagroup.com

Key Words: *CART, Bayes, Bayesian*

The Classification and Regression Tree (CART) algorithm is a hierarchical method for partitioning data into increasingly more homogenous groups. In standard CART analyses, the data drives the rules selected at each splitting node. In some cases, information about the relative importance of the rules may be available from other sources. We propose a Bayesian method which allows the inclusion of such information into CART analyses. Other Bayesian CART methods have been proposed but only allow the inclusion of uniform priors for the splitting rule assignment. We develop a stochastic algorithm for a CART model search which incorporates prior information on the rule importance. A method for placing priors on the tree space is described. Possible permutations of a small tree are summarized for illustrative purposes. We compare the Bayesian CART algorithm with the standard CART algorithm and a CART analysis which includes prior information on the rule importance via a weighting scheme. Including the prior information has a noticeable impact on the rules selected and leads to the reduction of variance for observations belonging to a given terminal node.

437 Nonparametric Methods ⌘

Biopharmaceutical Section

Thursday, August 12, 10:30 am–12:20 pm

A Compatible Nonparametric Estimate of Placebo Effect after Estimating the Shift in Multicenter Clinical Trials

◆ Mushfiqur Rashid, U.S. Food and Drug Administration

U.S. Food and Drug Administration, CDER, Division of Biometrics 2, 5600 Fishers Lane, PKLN 9B07, HFD-715, Rockville, MD 20857

rashidm@cder.fda.gov

Key Words: *r-estimate, placebo effect, shift, rank test*

In clinical trials, usually the shift (between treatment difference) is estimated by Hodges-Lehmann estimate based on Mann-Whitney and Wilcoxon test. But the placebo effect (or control effect) is usually estimated by the median of the data from the placebo (control group) group and the drug effect is estimated by median of the data from drug group. However, the difference of these two medians is not equal to estimate of the shift (using on Mann-Whitney and Wilcoxon test). One could use the difference between the two medians to estimate the shift. However, the standard shift estimate (using on Mann-Whitney and Wilcoxon test) will be more efficient than the estimate based on the difference of placebo median and drug median. Compatible nonparametric estimates of placebo effect, drug effect, and shift will be developed in a multicenter trial setting.

A SAS Procedure for the Stuart-Maxwell Chi-Square Test and Its Application to Clinical Trials

◆ Jiang-Ming J. Wu, Forest Research Institute

Forest Research Institute, Harborside Financial Center, Plaza V, Jersey City, NJ 07311

jiang-ming.wu@frx.com

Key Words: *categorical data analysis, contingency tables, matched pairs, SAS macros, Stuart-Maxwell chi-square test*

The Stuart-Maxwell chi-square test is a useful test procedure in categorical analysis that can be applied to $r \times r$, $r \geq 2$, matched pairs contingency tables. It also includes the well known McNemar's test as a special case when $r=2$. In the pharmaceutical industry, the SAS® language is the dominant software to be used for generating clinical tables, listings, and graphs. Unfortunately, there is no developed SAS procedure that can perform such a test. The author has written a SAS macro that can be used to perform this test. Examples of applying this SAS macro to analyze both an artificial dataset and a real world clinical dataset are presented. In addition to the macro itself, another advantage of this SAS macro is that it can be applied to some ill-conditioned sparse matrix cases where in theory these kinds of situations can not be determined. Discussions regarding some special cases as well as the developed SAS code that performs the Stuart-Maxwell chi-square test are also included and the developed SAS macro is free to all interested users.

A New Exact Test for Homogeneity of Risk Difference

◆ Colleen Kelly, San Diego State University; Karen Lo, San Diego State University

San Diego State University, Dept. of Mathematics and Statistics, 5500 Campanile Dr., San Diego, CA 92182-7720

kelly@math.sdsu.edu

Key Words: *multicenter clinical trials, risk difference, permutation test, bootstrap test, Type I error*

A new exact test for homogeneity of risk differences is proposed. Our proposed methodology is to develop randomization and bootstrapped tests to improve the performance of the asymptotic tests discussed in Lui and Kelly (2000). We first focus on a matched pairs design and then extend the test to the unmatched case. We investigate the performance of these new exact tests through Monte Carlo simulations, and compare their performance of that of the weighted least squared test statistic, which has an asymptotic chi-square distribution. In particular, we are interested in the performance when each center has a small sample size.

Some Modifications on the Application of the Exact Wilcoxon-Mann-Whitney Test

◆ Xiaohui Luo, Merck & Co., Inc.; Xun Chen, Merck & Co., Inc.

Merck & Co., Inc., RY34-A316, PO Box 2000, Rahway, NJ 07065-0900

edmund_luo@merck.com

Key Words: *discreteness, ties, Hodge-Lehmann procedure, permutation, variance adjusted statistic, small sample*

Despite of its lasting popularity, the application of the exact Wilcoxon-Mann-Whitney (WMW) test has never been free of inappropriateness or unnecessary inefficiency in practice. The goal of this paper is to exemplify some modifications that can be conducted to the application of the exact WMW test in small sample cases to achieve more precise and more powerful estimations and tests. At first, we propose a modified procedure to replace the current generalized Hodges-Lehmann procedure in Proc-StatXact 5 to correctly derive the exact WMW-test-based interval estimates for the shift parameter in the presence of ties. We then propose a variance adjustment for the exact WMW test to achieve more powerful exact test as the variance adjusted WMW statistic has less discrete distribution than the conventionally unadjusted WMW statistic. We find that the proposed variance adjusted exact method is noticeably less conservative than the conventional exact WMW method, especially when ties in observation are not very heavy.

Bias-corrected Spearman Rank Correlation Coefficient

◆ William W.B. Wang, Merck & Co., Inc.; Li Huiling, Columbia University; Ivan S.F. Chan, Merck & Co., Inc.

Merck & Co., Inc., 785 Jolly Rd., Bldg. C, Mailstop: UNA-102, Blue Bell, PA 19422

william_wang@merck.com

Key Words: *Spearman's Rank Correlation, Pearson Correlation, bias correction, confidence interval coverage, vaccine clinical trial*

Spearman rank correlation coefficient is a nonparametric statistic to measure the strength of association between two random variables, such as those for the cellular and humoral immune responses often measured in vaccine clinical trials. Even though Spearman correlation has been commonly used as a robust alternative to Pearson correlation, its statistical properties are neither popularly known nor well understood. We investigate the bias of Spearman rank correlation coefficient as an estimator of the population correlation coefficient. We point out that substantial bias may exist in some typical clinical data situations (and even with large sample size), resulting in poor coverage of the associated confidence interval. We propose a method to obtain the bias-corrected Spearman correlation and the associated confidence interval. Our simulation results show that this new method can dramatically reduce the bias and yield a confidence interval with proper coverage.

The Use of James-Stein Estimators in Tests of Homogeneity of the Risk Difference

◆ Purnima Rao-Melacini, McMaster University; Colleen Kelly, San Diego State University

McMaster University, 584 Governer's Rd., Dundas, ON L9H 5E3 Canada

raomelp@mcmaster.ca

Key Words: *multicenter data, treatment difference, meta-analysis, institution effect, shrinkage estimators*

The evaluation of the effectiveness of a new treatment as compared to previous therapies requires the design of a clinical trial. In this context, a multicenter clinical experiment is one in which two or more institutions agree to follow the same protocol. In medical and pharmaceutical research, responses are often measured on a dichotomous scale and then efficacy may be measured with the risk difference. To summarize the risk difference across centers, the estimated risk difference for each center must be comparable. Lipsitz et al. (1998) have proposed several tests of homogeneity of the risk difference in sets of 2x2 tables for sparse data, to which Lui and Kelly (2000) made several important changes and recommendations. We investigate using James-Stein estimates in weighted least squares statistics of homogeneity. These estimates shrink the individual proportions towards the overall mean and thus avoid the problems encountered when risk estimates are zero or one. In Monte Carlo simulations, we see that these estimates have a lower mean squared error with respect to the MLE, particularly when the mean sample size per treatment is small.

Which P Value is Correct? An Application of Randomization Test Procedure in a Lung Transplant Trial

◆ Yulan Li, Novartis Pharmaceuticals

Novartis Pharmaceuticals, One Health Plaza, East Hanover, NJ 07936

yulan.li@pharma.novartis.com

Key Words: *parametric test, randomization test, permutation, exact p value*

Generally, in inferential statistical data analyses, p values are obtained using the parametric tests assuming the test statistics follow certain parametric distributions. However, this assumption is not always relaiable, that is, p values obtained based on the assumption could be very different from those obtained using the randomization test procedures. Thus, investigation of parametric test procedures can be crucial to the conclusions. Permutation and design-based randomization test procedures are assessed to determine the correct p value. A real example in transplant trial was used for illustration.

438 Cross-over Designs

Biopharmaceutical Section

Thursday, August 12, 10:30 am–12:20 pm

Confounded Cross-over Designs for Factorial Experiments

◆ Sourav Santra, Northern Illinois University; Sudhir Gupta, Northern Illinois University

Northern Illinois University, Division of Statistics, DU 366, DeKalb, IL 60115

santra@math.niu.edu

Key Words: *change-over design, confounding, cross-over design, extra-period design, factorial experiment, repeated measures design*

Although cross-over designs for comparing treatments have been considered by several authors, factorial structure in cross-over designs has been studied only recently. Often some higher order interactions can be safely assumed to be zero in factorial experiments. Confounded cross-over designs for symmetrical factorial experiments with number of levels a prime or power of a

prime are studied in this paper. In addition to having factorial structure, all estimable direct main effects and interactions are estimated with a constant variance, and all estimable residual main effects and interactions are estimated with a constant variance in these designs. Extra-period confounded cross-over designs are also studied.

Estimating the Variance Components in a Replicated Cross-over Study

◆ Guoyong Jiang, Cephalon, Inc.; Jianrong Wu, St. Jude Children's Research Hospital

Cephalon, Inc., 145 Brandywine Pkwy., West Chester, PA 19380

jjiang@cephalon.com

Key Words: *block compound symmetry, linearly patterned covariance matrix, maximum likelihood estimation, restricted maximum likelihood estimation, saturated model, uniform within sequences*

In a uniform within-sequence replicated cross-over experiment, the covariance matrix of a vector variable of measurements on each subject typically possesses a block compound symmetric structure. We derive explicit representations for the maximum likelihood and restricted maximum likelihood estimators of the variance components. They can be expressed as known linear combinations of elements of the sample covariance matrix from the original observations. We illustrate the new method by applying it to an example.

A Comparative Study of Statistical Methods for Population Bioequivalence in 2x2 Cross-over Design

◆ Sang-Gue Park, Chung-Ang University; Wan-Youn Yang, Kyungwon University

Chung-Ang University, Dept. of Statistics, Seoul, 156-756 Korea

spark@cau.ac.kr

Key Words: *bioequivalence, cross-over design, population bioequivalence, prescribability*

The Food and Drug Administration (FDA) issued a second-draft guidance in August 1999 on the subject of in-vivo bioequivalence, which is based on the concepts of individual and population bioequivalence. The intention of this guidance is to replace the 1992 guidance that requires that in-vivo bioequivalence be demonstrated by average bioequivalence. The test for population bioequivalence in the latest FDA guidance document on statistical considerations for bioequivalence studies has been known as very conservative. The FDA test is compared with three tests—Lee, Shao, and Chow (2002), Chow, Shao, and Wang (2003), McNally, Iyer, and Mathew (2002)—for assessing population bioequivalence in 2x2 cross-over design through simulation study.

Cross-over Design and Model for Drug Abuse Potential Study

◆ Ling Chen, U.S. Food and Drug Administration; Jun Shao, University of Wisconsin

U.S. Food and Drug Administration, HFD 705, 5600 Fishers Lane, Rockville, MD 20857

chenli@cder.fda.gov

Key Words: *drug abuse, mixed carryover effect, self-evaluated response, Williams design*

In drug abuse potential clinical studies, there is large between-subject variability in the endpoint measurements based on self-evaluated responses. In addition, these studies are small, because of the difficulty in recruiting appropriate study subjects. The designs for such studies are thus typically cross-over, with self-control. The design for a study reviewed recently, comprised four replicates of a 6x6 Williams square crossover design with 24 subjects. Since the responses from the primary variables were very subjective and washout periods were not sufficient among the treatments, a serious mixed carryover effect occurred in the study. I will propose an improved model for similar drug abuse potential studies.

Comparison of Variances of Response Data in a Cross-over Study

◆ Jiannong Wang, Novartis Pharmaceuticals

Novartis Pharmaceuticals, Biometric - US CD&MA, One Health Plaza, East Hanover, NJ 07936-1080

jim.wang@pharma.novartis.com

Key Words: *cross-over study, variance comparison, correlated data*

Usually the response data from different periods in a cross-over study will be considered to be correlated. Since the assumption of independence of two samples is not satisfied, the traditional F-test techniques will not be applicable when the variances of numeric responses need to be compared under certain circumstances. Some research results on how the data can be analyzed will be presented. First, distribution of the (correlated) F-statistics will be investigated. Some tables of the percentile values will be provided. Then bootstrapping methodology will be applied on same examples. The results from bootstrapping methods will be checked against the tables of the percentile values.

An Application of Repeated Measures Analysis in a Two-period, Longitudinal Cross-over Study

◆ Xiaojing Wang, Merck & Co., Inc.; Guanghan F. Liu, Merck & Co., Inc.; Duane Snavely, Merck & Co., Inc.; Al Getson, Merck & Co., Inc.

Merck & Co., Inc., 10 Sentry Pkwy., BL3-2, Blue Bell, PA 19422

xiaojing_wang@merck.com

Key Words: *cross-over study, longitudinal method, repeated measures, analysis of variance*

In the early stages of clinical development, cross-over designs are often used to obtain adequate study power with a smaller number of patients. To take full advantage of the crossover design, it requires an assumption of no carry-over effect. Without carry-over effect, the treatment comparison can be made within subject. The study gains more power because within subject variance is usually smaller than between subject variance. When the primary efficacy variable is measured several times within a period, a longitudinal method may be considered for the analysis of the cross-over study data. This method will have some advantage over a conventional statistical analysis method since all observed data can be used in the analysis regardless of whether a patient

completes or discontinues prematurely from the study. Statistical tests for sequence and period effects as well as for the treatment comparison of the primary endpoint will be discussed. Simulations will be used to illustrate the method and will be compared with the conventional analysis of variance model.

Nonparametric Confidence Intervals for Tmax: Methods Comparison for 3x3 Crossovers

◆ Susan A. Willavize, Pfizer Inc.; Elizabeth A. Morgenthien, Quincunx Group Biostatistical Consulting

Pfizer Inc., PGRD, Groton Laboratories—MS8260-2221, 1 Eastern Point Rd., Groton, CT 06340

Susan_A_Willavize@groton.pfizer.com

Key Words: *3x3 cross-over trial, nonparametric confidence interval, Hodges-Lehmann confidence interval, Tmax*

Tmax, the time associated with the maximum plasma drug concentration, is a continuous random variable in theory, but actually discrete, because it is equated to a nominal sampling time. The European guidance on bioavailability and bioequivalence (2001) indicates that, when Tmax is analyzed, a nonparametric confidence interval (CI) method should be used for the pairwise comparisons of interest. For 3x3 cross-over designs, we propose a new median-scaled (MS) method of adjusting for periods in a Hodges-Lehmann type CI. A simulation study was done comparing MS with a previously described stratified ranks method. The normal theory (NT) and the nonparametric CI approaches (without adjustment for periods) were also compared. Concentration data were simulated using (Pharsight's) TS-2 software. Results show that sparse sampling and period effects cause increases in CI length. NT can be liberal (i.e., less than nominal coverage) if there is a true treatment effect. All the nonparametric methods are conservative and among them MS is least conservative and has shortest CIs. SR is most conservative and has very long CIs.

439 Multiple and Composite Outcomes in Clinical Trials

Section on Statistics in Epidemiology, Biometrics Section, ENAR

Thursday, August 12, 10:30 am–12:20 pm

Composite Outcomes in Cluster-randomized Clinical Trials: Experiences from the Community Hypertension Assessment Trial

◆ Lehana Thabane, McMaster University; Keith O'Rourke, Ottawa Health Research Institute; Janusz Kaczorowsk, McMaster University; Larry Chambers, University of Ottawa; Tina Karwalajtys, McMaster University; Cheryl Levitt, McMaster University; William Hogg, University of Ottawa; Lisa Dolovich, McMaster University

McMaster University, Centre for Evaluation of Medicines, 105 Main St. East, Level P1, Hamilton, ON L8N 1G6 Canada

thabanl@mcmaster.ca

Key Words: *composite outcomes, controlled clinical trials, hypertension, community-based intervention, blood pressure, cluster-randomization*

The choice of appropriate primary and secondary endpoints is one of the many crucial elements in designing a clinical trial. This presentation will examine the use of composite endpoint(s), created by combining several endpoints, in a community-based cluster-randomized clinical trial. We will use our experiences from the CHAT study to examine the challenges of choosing a composite endpoint. The CHAT study is a paired cluster-randomized trial to assess the effectiveness of community pharmacy blood pressure (BP) sessions, with feedback to physicians, patients and pharmacists, on monitoring and management of BP among older adults. It is a pragmatic trial intended to assess the effectiveness of the intervention in real-life setting and therefore relies on health record review as the primary source of data. There are no standardized BP measurements at the start and end of the trial. As a result, there are several methodological challenges that include choice of (preferably one) primary composite (or individual) endpoint that is valid, reliable, and sensitive to change based on chart audits.

Alternative Analyses of Risk Factors for Composite Endpoints

◆ Robert J. Glynn, Brigham & Women's Hospital; Bernard A. Rosner, Harvard Medical School

Brigham & Women's Hospital, 900 Commonwealth Ave., Boston, MA 02215

rglynn@rics.bwh.harvard.edu

Key Words: *polytomous regression, competing risks, survival analysis, model building, goodness-of-fit, likelihood methods*

Clinical trials and observational studies commonly consider composite endpoints, such as total mortality, all cancer or cardiovascular disease, to provide a broad evaluation of the benefits or risks of an intervention. Polytomous logistic regression and competing risk survival analysis offer alternative approaches to compare the overall and component-specific effects of risk factors and identify heterogeneity in these effects. We compared these approaches for evaluation of risk factors for a composite endpoint including myocardial infarction, stroke, and venous thromboembolism in a 20-year follow-up of 18,662 participants in the Physicians' Health Study. Strengths of both approaches include ready implementation in available software, likelihood-based strategies for comparing effects across components, interpretable parameters, and ability to include time-varying covariates. Polytomous logistic regression has the advantage of being fully parametric with accessible measures of goodness of fit and explained variance. Survival analysis better accounts for censoring and accommodates variables such as age that may have nonproportional effects on outcomes over time.

Analysis Strategies for Multiple-outcome Measures

◆ Sarah Baraniuk, University of Texas; Lemuel A. Moye, University of Texas School of Public Health

University of Texas, School of Public Health, 1200 Herman Pressler, Suite E835, Houston, TX 77030

sbaraniuk@sph.uth.tmc.edu

Key Words: *multiple-outcome measures, multiplicity, clinical trials, experiment-wise error rate*

Multiple-outcome measures are common in clinical trials, particularly in cardiovascular clinical trials where generally more than one type of outcome is measured (different types of strokes, myocardial infarctions, and death). The outcomes are often correlated to varying degrees and differ in severity. The motivation of this work stems from a clinical trial involving different stroke severity measures, ending with the construction of strategies that aid the analysis of multiple outcome measures.

Mixture Survival Models for Multiply Censored Data with Multiple Outcomes

◆ Nong Shang, Centers for Disease Control and Prevention

Centers for Disease Control and Prevention, Atlanta, GA

nms6@cdc.gov

Key Words: *survival analysis, mixture model, log-rank test, censoring*

When applying a treatment to a population, such as in a clinical trial, there may be multiple outcomes at the end of study. For example, the treatment may have no effects to some patients, while has short-term effects as well as long-term effects to the other patients, resulting in a mixture of three groups of patients. The complex nature of the trials may also bring in all types of censoring. Linear mixture survival models are applied to estimate the proportions and the survival functions of all subpopulations. Log-rank based tests are proposed to compare different treatment arms for each of the proportions and the survival functions.

Analyzing Health-related Quality-of-life Measures Using General Mixed-data Models

◆ Alexander de Leon, University of Calgary; Daniel C. Bonzo, Serono, Inc.

University of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4 Canada

adeleon@math.ucalgary.ca

Key Words: *general location model, conditional grouped continuous model, generalized linear model, generalized estimating equation (GEE)*

Clinical trials typically involve various continuous measures as main efficacy endpoints and a host of health-related quality of life (HRQOL) endpoints as important secondary endpoints. The use of such measures presents an analytic challenge on how to best exploit simultaneously the existing relationship between the various continuous, ordinal, and nominal variables that form the endpoints. Two types of models used in the past to handle such a setting were general location models and conditional grouped continuous models. We introduce generalized linear models based on the so-called general mixed-data model developed by de Leon and Carriere (2003) for simultaneously analyzing continuous efficacy and HRQOL endpoints. To avoid the difficulties associated with fully likelihood-based approaches, we employ generalized estimating equations (GEEs) to estimate the model parameters. We illustrate the applicability of the proposed methodology on clinical data.

The Impact of Twin Births on Inference in Randomized Clinical Trials of Pre-term Infants

◆ Michele L. Shaffer, Pennsylvania State University; Kristi L. Watterberg, University of New Mexico

Pennsylvania State University, College of Medicine, 600 Centerview Dr., Suite 2200, Hershey, PA 17033

mshaffer@hes.hmc.psu.edu

Key Words: *correlated data, mixed-effects model, generalized estimating equations*

The multiple-gestation rate in pre-term births is considerably higher than in full-term births so that twin births can account for an appreciable percentage (10-20%) of a study sample for a trial targeted at pre-term infants. Failure to account for the correlation within complete twin pairs may impact the precision of estimated treatment effects or lead to incorrect decisions regarding the significance of treatment. The current clinical literature rarely mentions if twins were taken into consideration for trial design or analysis. Well-known methods exist for analyzing correlated data, but the operating characteristics of these methods are not fully understood for such a mix of correlated and independent data. A simulation study is conducted to compare mixed-effects models and generalized estimating equations to standard methods, which assume all subjects are independent, for continuous outcomes in two-armed clinical trials. Data from an NICHD Neonatal Research Network trial are used as motivation.

440 Bayesian Computation and Data Augmentation

Section on Statistical Computing, Section on Bayesian Statistical Science

Thursday, August 12, 10:30 am–12:20 pm

Fully Bayesian Computing

◆ Jouni Kerman, Columbia University; Andrew Gelman, Columbia University

Columbia University, 611 Mathematics Bldg., Dept. of Statistics, New York, NY 10027

kerman@stat.columbia.edu

Key Words: *Bayesian inference, statistical computation, computational methods*

A fully Bayesian computing environment calls for the possibility of defining vector and matrix objects that may contain both random and deterministic quantities, and syntax rules that allow treating these objects much like any variables or numeric arrays. Working within the statistical package R, we introduce a new object-oriented framework based on general vector object types that can accommodate any type of object, including new random variable objects that are implicitly represented by simulations. We seek to be able to manipulate posterior simulation objects conveniently and transparently and provide a basis for further development of methods and functions that can access these objects directly. We illustrate the use of this new programming environment with several examples of Bayesian computing, including posterior predictive checking and the manipulation of posterior simulations.

On Real-parameter Evolutionary Monte Carlo Algorithm

◆ Gopika R. Goswami, Harvard University

Harvard University, Science Center, 7th Floor, 01 Oxford St., Cambridge, MA 02138

goswami@stat.harvard.edu

Key Words: *exchange operation, cross-over operation, Markov chain Monte Carlo, real-parameter Evolutionary Monte Carlo, snooker algorithm, temperature ladder*

Real-parameter Evolutionary Monte Carlo algorithm (EMC) has been proposed as an effective method not only for sampling from multi-modal distributions but also for stochastic optimization. The authors have shown by various well-chosen examples that it works better than parallel tempering, it has a certain ability to learn from the past and it improves mixing by sampling along a temperature ladder. Our goal is to improve the performance of this method and provide some practical implementation guidelines. First, we introduce some new exchange moves which equips the algorithm with more learning capability and makes it faster while preserving the quality of the samples (as measured by integrated autocorrelation time) produced from the target distribution. Second, we introduce some premiliminary run based tools for contructing the temperature ladder. Lastly, we generalize a result originally proved about the validity of the conditional (or line) sampling step in the context of the snooker algorithm which is a type of move involved in the EMC.

The Gibbs Sampler for Bayesian Analysis on Censored Categorical Data

◆ Thomas J. Jiang, National Chengchi University; Li-Wen Ko, Kaohsiung Municipal Vocational High School of Commerce

National Chengchi University, Dept. of Mathematical Sciences, Taipei, 116 Taiwan

jiangt@math.nccu.edu.tw

Key Words: *Bayesian inference, Gibbs sampler, quasi-Bayes method, censored categorical data*

Bayesian analysis on censored categorical data frequently involves high-dimensional computations that have no closed form. The quasi-Bayes method is shown to provide a solution to this computational difficulty. The Gibbs sampler provides an alternative solution. Here, we shall describe and explore the application of the Gibbs sampler to this Bayesian analysis. Comparisons and discussions between the quasi-Bayes method and the Gibbs sampler shall also be given.

Ranking the Sources of Numerical Error in MCMC Computations

◆ William F. Guthrie, National Institute of Standards and Technology; Hung-kung Liu, National Institute of Standards and Technology; Donald Malec, National Institute of Standards and Technology; Grace Yang, National Institute of Standards and Technology

National Institute of Standards and Technology, 100 Bureau Dr., Stop 8980, Gaithersburg, MD 20899-8980

will.guthrie@nist.gov

Key Words: *MCMC computations, numerical error, Bayesian analysis, reference data*

Like all statistical computations, the computations in Markov Chain Monte Carlo analyses are subject to many sources of numerical error. The sources of numerical error that arise in the computations associated with a simple Bayesian model for data sets from the NIST Statistical Reference Datasets (StRD) web site will be studied. The different sources of numerical error will be compared and ranked with respect to their impact on the total numerical error.

Hyperplane Inflation for Markov Chain Monte Carlo in Locally Nested Spaces

◆ Giovanni Petris, University of Arkansas

University of Arkansas, Fayetteville, AR 72701

GPetris@uark.edu

Key Words: *Markov chain Monte Carlo, model mixing, model selection, variable selection*

We propose an approach to transdimensional MCMC based on geometric considerations. The main tool is provided by a family of functions that, by inflating low-dimensional hyperplanes, maps two nested models onto nonoverlapping regions of a unique auxiliary space. A sampler in the auxiliary space can then move smoothly between the regions representing the different models, avoiding in this way the need for “jumps” and corresponding jump proposals. Transdimensional MCMC using the proposed approach can be carried out in a semiautomatic way. An example will be discussed.

An Update on the NIST Statistical Reference Datasets for MCMC

◆ Hung-kung Liu, National Institute of Standards and Technology; William F. Guthrie, National Institute of Standards and Technology; Donald Malec, National Institute of Standards and Technology; Grace Yang, National Institute of Standards and Technology

National Institute of Standards and Technology, Statistical Engineering Division, 100 Bureau Dr., Stop 8980, Gaithersburg, MD 20899

liu@nist.gov

Key Words: *Statistical Reference Datasets (StRD), Markov chain Monte Carlo (MCMC), numerical accuracy, floating point arithmetic, Bayesian analysis*

In the Statistical Reference Datasets project, NIST provided datasets on the web (*www.itl.nist.gov/div898/strd/index.html*) with certified values for assessing the accuracy of software for univariate statistics, linear regression, nonlinear regression, and analysis of variance. A new area in statistical computing is the Bayesian analysis using Markov chain Monte Carlo. Despite of its importance, the numerical accuracy of the software for MCMC is largely unknown. We have recently updated the StRD web site with the six new datasets for Bayesian model fitting using MCMC algorithms. We will discuss some results obtained using these datasets that challenge the conventional wisdom that longer simulations lead to improved approximation of the posterior distribution.

441 Fixing the Data: Post-survey Edit and Disclosure Control ◬

Section on Survey Research Methods

Thursday, August 12, 10:30 am–12:20 pm

Electronic Data Collection and Data-editing at Data Capture

◆ Paula E. Weir, EIA

EIA, 19343 Mill Dam Place, Lansdowne, VA 20176

Paula.Weir@eia.doe.gov

Key Words: *electronic data reporting (EDR), editing, internet data collection (IDC), fatal edits, query edits, self-administered surveys*

Many surveys that historically were paper/mail surveys have begun offering the option of electronic data reporting. As a result, the potential for editing data at data capture has greatly increased. This paper explores that editing potential for electronic data reporting through computer self-administered questionnaires (CSAQ) via web surveys, downloadable software, and email attachments with examples from the Energy Information Administration's surveys. The presence and the extent of fatal and query edits that are implemented at the initial data entry and capture, versus those implemented in the traditional data-editing stage is investigated in relationship to: (1) the amount of development resources dedicated; (2) the sophistication of the electronic option selected; (3) the security of the transmission that is required; (4) the quality of the data that is required; and (5) the amount of respondent burden that is acceptable.

Implicit Linear Inequality Edits Generation and Error Localization in the SPEER Edit System

◆ Maria M. Garcia, U.S. Census Bureau

U.S. Census Bureau, 4700 Silver Hill Rd., Stop code 9100, Washington, DC, DC 20233-9100

Maria.M.Garcia@census.gov

Key Words: *error localization, editing, Fellegi-Holt model*

The Census Bureau's SPEER editing system applies the Fellegi-Holt model to economic establishment surveys under ratio edits and a limited form of balancing. If the complete set of explicit and implicit edits is available, then Fellegi-Holt methods have the advantage that they determine the minimal number of fields to change so that a record satisfies all edits in one pass through the data. For most surveys implicit edits are not generated because the generation requires days-to-months of computation. In some situations when implicit edits are not available Fellegi-Holt systems use integer programming methods to solve the error localization problem directly and slowly. With only a small subset of the needed implied edits, the current version of SPEER applies ad hoc heuristics that finds error-localization solutions that are not optimal for as much as five percent of the edit-failing records. We present recent modifications to the SPEER editing system that maintain its exceptional speed and do a better job of error localization. The new SPEER has an auxiliary program for generating implicit linear inequality edits.

Using a Quadratic Programming Approach to Solve Simultaneous Ratio and Balance Edit Problems

◆ Katherine Thompson, U.S. Census Bureau; James T. Fagan, U.S. Census Bureau; Brandy Yarbrough, U.S. Census Bureau; Donna Hambric, U.S. Census Bureau

U.S. Census Bureau, 2700 Karen Dr., Chesapeake Beach, MD 20732-3841

katherine.j.thompson@census.gov

Key Words: *nonlinear programming, constrained optimization*

Often, items collected by the Economic Census are first subjected to ratio edit tests and associated corrections to ensure consistency within the questionnaire and within the industry and are then subsequently "balanced" to ensure additivity with other data items. With this two-step procedure, the balanced data often fail to satisfy the original ratio edit tests. In most cases, this does not pose a data quality problem. However, there are situations in which it is undesirable to "undo" the original ratio edit corrections to satisfy additivity: examples include marginal totals in two-dimensional balance complexes and derived items that combine previously ratio-edited items with unedited items. We present an approach that first uses nonlinear (quadratic) programming to find data "adjustment" solutions that satisfy such simultaneous balancing and ratio editing problems and then performs controlled rounding to obtain integer solutions.

Effects of Rounding Continuous Data Using Specific Rules

◆ Joe Fred Gonzalez, Jr., National Center for Health Statistics; Jay J. Kim, National Center for Health Statistics; Lawrence H. Cox, National Center for Health Statistics; Myron Katzoff, National Center for Health Statistics

National Center for Health Statistics, 3311 Toledo Rd. Room 3121, Hyattsville, MD 20782

jgonzalez@cdc.gov

Key Words: *rounding, integer, variance, uniform distribution, disclosure risk, posterior probability*

Data such as incomes are frequently rounded. Rounding may be done to protect the confidentiality of records in a file or to enhance readability of the data, or by the notion that the digits subject to rounding are inconsequential. The rounding may not have any effect on the bias of an estimator, but may have a large impact on variance. Integers can be expressed as $x=qB+r$, where q is the quotient, B is the base, and r is the remainder. B is a constant, but q and r are random variables. We use four rules for rounding "r" above to observe the effects of rounding on bias and variance. We will assume a uniform distribution on r, but no specific distributional assumption will be made on "q." When $q=0$, we will show that the variance after rounding is three times the variance before rounding. As the variance of q gets larger, the effect of rounding on the variance decreases. Disclosure risk in terms of the posterior probability $P(x|qB)$ will also be shown.

Use of an Audit Program to Improve Confidentiality Protection of Tabular Data at the Bureau of Labor Statistics

◆ Randall Powers, Bureau of Labor Statistics; Stephen Cohen, Bureau of Labor Statistics

Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

powers_R@bls.gov

Key Words: *individually identifiable data, audit, confidentiality, tabular data*

U.S. Statistical Agencies develop disclosure avoidance processes to ensure that individually identifiable data can not be detected in the published tabular tables from confidential data. The Disclosure Audit System (DAS) is software that uses linear programming methodologies to audit the effectiveness of disclosure avoidance processes to prevent identification of individual responses. Currently, all data released by programs at the Bureau of Labor Statistics are subject to heuristic disclosure analysis algorithms which ensure that data users outside the Bureau can't ascertain the values of individually respondent data. The Covered Employment and Wages Survey program publishes quarterly and annual counts of employment and wages reported by employers covering 98% of U.S. jobs, available at the national, state, MSA, and county levels by North American Industry Classification System (NAICS) codes. This paper will evaluate the effectiveness of confidentiality procedures used to protect the confidentiality of tabular data in the Covered Employment and Wages Survey. We will discuss how we applied the DAS software to evaluate these confidentiality procedures, and the ramifications of our findings.

Evaluating Alternative Disclosure Limitation Strategies for Public Use Data Files: The California Health Interview Survey's Approach

◆ M. Leeann Habte, University of California, Los Angeles; Jenny Chia, University of California, Los Angeles

University of California, Los Angeles, 10911 Weyburn Ave., Suite 300, Center for Health Policy Research, Los Angeles, CA 90024

lhabte@ucla.edu

Key Words: *privacy, confidentiality, statistical disclosure limitation, risk assessment, key variables*

Adequate protection of respondent confidentiality is crucial for surveys that release microdata to the public, yet uniform standards for disclosure limitation are lacking. Theoretical and practical approaches to risk assessment and disclosure control vary. This paper will describe the California Health Interview Survey's (CHIS) confidentiality policies and data disclosure limitation techniques for its public use data files. It will compare CHIS' approach to those of select national surveys, including the National Health Interview Survey and the National Immunization Survey. Emphasis will be placed on the factors CHIS used in evaluating the applicability of particular disclosure limitation strategies, such as differential methods of participant selection and the nature of survey information collected. The paper will discuss the implications of these strategies for researchers using CHIS data.

Limiting the Risk of Data Disclosure by Using Swapping Techniques in Variance Estimation

◆ Sylvia M. Dohrmann, Westat; Randy Sitter, Simon Fraser University; Leyla Mohadjer, Westat; Jill M. Montaquila, Westat; Lester R. Curtin, Centers for Disease Control and Prevention

Westat, 1650 Research Blvd., Rockville, MD 20850

sylviadohrmann@westat.com

Key Words: *confidentiality, replication, Taylor Series linearization, area sample*

The National Health and Nutrition Examination Surveys (NHANES) are one of the series of health-related programs conducted by the National Center for Health Statistics (NCHS). Beginning in 1999, NHANES became an annual survey with data release to the public every two years. NHANES 1999-2002 was fielded in approximately 57 locations comprising 54 PSUs that were primarily single counties. The small number of PSUs in the sample, together with characteristics available in the data files and outreach efforts aimed at increasing public awareness of the survey, led to concerns about the risk of PSU identification in the public release of the data. NCHS reviewed the data to identify disclosure risks, and requested research on variance estimation approaches to mask the PSUs. This paper will present the methods used to create alternative sets of PSU and stratum identifiers for variance estimation, the comparison of the variance estimates from each of the approaches considered, and a discussion of how the swapping approach is an improvement over the previously released method for estimating variances in NHANES.

442 The Internet, Animation, Simulation, Visualization, and Formulation to Improve Teaching

Section on Statistical Education

Thursday, August 12, 10:30 am–12:20 pm

The Evolution of a Revolution: The Impact of the Internet on Statistical Education

◆ Cynthia A. Gargano, Merck & Co., Inc.

Merck & Co., Inc., 10 Sentry Pkwy., BL3-2, Bluebell, PA 19422

cynthia_gargano@merck.com

Key Words: *internet, education*

The internet has revolutionized the way we communicate, educate, conduct business, and collect information in our personal and professional lives. Just as the computer took statistical computation to a new level, the internet has changed the way statistics is taught in the classroom environment. This paper takes a look at the evolution of the internet in statistics education.

Creating and Using Online Simulations in Statistical Education and Training

◆ Thomas B. Barker, Rochester Institute of Technology

Rochester Institute of Technology, 98 Lomb Memorial Dr., Center for Quality and Applied Statistics, Rochester, NY 14623

tbbeqa@rit.edu

Key Words: *simulation, online learning*

Discovery learning is more and more an ingrained part of the way students of all ages participate in the educational process. By drawing out ideas from the student, the process guide (formerly known as the teacher) allows that student to develop the ideas of the topic and thus gives the student ownership of the ideas. Of course, careful guidance is incumbent upon the process guide who assures that the emerging ideas fit the accepted norms of truth and current practice. This paper will show how, by using online (web based) exercises, students can seek out and discover statistical methods that will enable problem-solving. In most cases, the students are subjected to some preliminary "sage on the stage" lectures to set a framework for their own learning extensions. The process described shows both course design and learning outcomes.

Animations that Demonstrate Statistical Concepts

◆ Edward R. Mansfield, University of Alabama; B. Michael Adams, University of Alabama; Michael Conerly, University of Alabama

University of Alabama, Box 870226, Dept. of ISM, Tuscaloosa, AL 35487-0226

emansfie@cba.ua.edu

Key Words: *animation, teaching, pedagogy, presentations*

The use of computer animation can be a valuable asset in explaining the logic behind many concepts and techniques that are taught in statistics courses. As presentation tools such as PowerPoint become more powerful, more flexibility exists for communicating with students. Several animated graphical illustrations that we have created for in-class use will be demonstrated. Most of our animations work particularly well in large lecture classes at the introductory level. Some of the examples to be shown include using the normal distribution table, power of a test, simulations associated with confidence intervals, and the effect of changes in the mean and/or variability of a system on the likelihood of meeting a customer's needs. The animations that we present at JSM and others will be made available as PowerPoint files to those who participate at the meetings.

Visualizing Confidence Intervals

◆ Ronald L. Tracy, Oakland University

Oakland University, Dept. of Economics, Rochester Hills, MI 48309

tracy@oakland.edu

Key Words: *graphics, computers, teaching, mean, standard deviation, variance*

Students have difficulty understanding confidence intervals. They often mistakenly think that $P(Xl< =\mu< =Xu) = 0.95$ means that there is a 0.95 probability that the parameter mu lies in the interval [Xl, Xu]. Although textbooks and instructors valiantly try to correct this interpretation, the notation does nothing to discourage it. The software Visualizing Confidence Intervals illustrates the correct interpretation of a confidence interval through simulation. The software allows the user to create confidence intervals for m using either the known standard deviation s or the estimated standard deviation s. It also enables students to create a confidence interval for the population variance s2. By changing the sample size the user can investigate its affect on a confidence interval. Although the default distribution of the X's is normal, the distribution can be changed to uniform or a variety of skewed distributions. This feature enables the user to investigate the effect the underlying distribution has on the interpretation of a confidence interval. Other features include a help system with search capability and "hot text."

Demonstrating Multicollinearity: A Spreadsheet Simulation

◆ David P. Doane, Oakland University; Mark W. Isken, Oakland University

Oakland University, Dept. of Decision & Info Sciences, Rochester, MI 48309-4493

doane@oakland.edu

Key Words: *multicollinearity, simulation, spreadsheet, collinearity, variance inflation*

Spreadsheet simulations are used to demonstrate how colinearity and multicollinearity lead to variance inflation and instability of least squares estimates in linear regression. The user specifies the relationships among predictors, true model parameters, and random error magnitude. In single sample mode, individual samples are taken and the estimation results are displayed numerically and visually. In replication mode, we create empirical distributions of fitted coefficients compared with the orthogonal reference distribution to reveal the meaning of variance inflation and its consequences for estimating individual predictor coefficients.

A Hierarchical System for Analyzing and Formulating Statistical Problems

◆ Norma F. Hubele, Arizona State University; George Runger, Arizona State University; Sarah Brem, Arizona State University

Arizona State University, Dept. of Industrial Engineering, PO Box 875906, Tempe, AZ 85287-5906

hubele@asu.edu

Key Words: *teaching statistics, problem-solving, statistical concepts*

The central issue in teaching statistics is to impart to students the underlying principles of statistical reasoning, rather than a cookbook of rote procedures. Instructors expend considerable energy to formulate, illustrate, and motivate conceptual and practical elements. This research supplements others by focusing on ways to make the underlying problem structure salient to students through homework and classroom word problems. The structural approach to problem-solving helps students to understand why two problems that appear different because they do not share surface features, are really "the same thing." A hierarchical classification system useful for dissecting the complexity of problem statements is presented. By applying this system, instructors can deliberately instruct students how to distinguish between surface features of a particular application of statistical concepts and the underlying statistical concepts themselves. If a student can see the structure of the problem, then they can see how to extend their existing knowledge to novel problems and domains that incorporate those same structures. Examples are given from introductory engineering statistics class work.

443 Examples and Tools for Probability and Mathematical Statistics

Section on Statistical Education

Thursday, August 12, 10:30 am–12:20 pm

Some Variations of the Hat Problem

◆ Marepalli B. Rao, University of Cincinnati; Rupa Mitra, Minnesota State University, Moorhead

University of Cincinnati, Center for Genome Information, 3223 Eden Ave., Cincinnati, OH 45267-0056

marepalli.rao@UC.EDU

Key Words: *probability, optimization, strategy*

The Hat Problem has been circulating among mathematics and statistics departments all over the world for quite some time. This is a fascinating probability problem which can be posed in any introductory probability class. The problem can be stated as follows. Belinda, Miranda, and Yolanda step on to a stage in order to win a million dollar prize collectively. Each of the members of the team is fitted by the host with a hat, which is either black or white. Each member can see the color of the hats of the other members but not her own. Each member is queried by the host separately to guess the color of her own hat. The response could be black, white, or pass. The team will get a million dollars prize provided at least one guesses the color of the hat and guesses right. The participants could get into a huddle before coming on to the stage, formulate a stragegy of responses, and then execute the strategy on the stage. The puzzle is to find a strategy which maximizes the probability of winning the prize. In this presentation, we discuss some variations of this problem.

The Doomsday Argument: An Intriguing Example of Bayes Theorem

◆ Dean Nelson, University of Pittsburgh, Greensburg

den@pitt.edu

Key Words: *doomsday, Bayes theorem*

An argument based on Bayes theorem was put forward in the late 1980s that purports to show that the human race will die out sooner rather than later. Because of this gloomy conclusion, it was called the Doomsday Argument. Because the argument is based on a simple application of Bayes theorem and because it reaches a disturbing conclusion with intrinsic interest to anyone in the human race, the Doomsday Argument makes a useful example to initiate discussion in an undergraduate probability theory class. The disturbing conclusion leads one to question the assumptions used to assign probabilities, what a probability model can infer, or whether a model is even appropriate. If the assumptions of the first model are accepted, a similar conclusion in the opposite direction in time can be shown, that the human race is young rather than old.

Major Portion Analysis

◆ Mary S. Fowler, Worcester State College; Joseph B. Kadane, Carnegie Mellon University

Worcester State College, 15 Northwoods Rd., Groton, MA 01450

m_fowler@charter.net

Key Words: *truncation, law, oil, royalties*

It is often difficult to offer students real-life examples requiring the use of calculus-based probability theory. We present such an example, concerning the valuation of oil for royalty purposes due to Indian tribes on leases let by the Department of the Interior. All oil sales made at prices below the median are to be valued at the median price to determine the royalty to be paid. This leads to an interesting truncation problem that is addressed using basic probability theory.

Finite Population Probability-generating Functions

◆ George R. Terrell, Virginia Polytechnic Institute and State University; Steven J. Kathman, GlaxoSmithKline

Virginia Polytechnic Institute and State University, Statistics Dept., Blacksburg, VA 24061

terrell@vt.edu

Key Words: *discrete distributions, hypergeometric variables, occupancy problems*

Probability-generating functions (PGF) are an elegant device for deriving probability densities and moments for non-negative integer-valued random variables. Unfortunately, they cannot be expressed in elementary form for a number of important families, such as the hypergeometric. We propose an alternative form of generating function, the Finite Population Probability-generating Function (FPPGF), that takes an elementary form in many cases where the PGF does not. Interestingly, the FPPGF is as simple as the PGF for some important families, such as the binomial. It facilitates derivation of probability densities and moments in a similar way. We demonstrate its application to cases that were formerly considered difficult.

Using APPL

◆ Diane Evans, Rose-Hulman Institute of Technology

Rose-Hulman Institute of Technology, Mathematics Dept., 5500 Wabash Ave., Terre Haute, IN 47803

diane.evans@rose-hulman.edu

Key Words: *computer algebra system, post-calculus statistics, probability distributions*

This talk shows how the Maple-based software tool APPL, "A Probability Programming Language," can be utilized in the classroom for performing random variable manipulations and calculations (e.g., PDF/CDF conversions, convolutions, distribution plot comparisons) for both well-known and arbitrary distributions. I have integrated the use of APPL into my post-calculus introductory statistics and probability courses at Rose-Hulman Institute of Technology. The student response has been positive; they are able to (1) minimize time spent on tedious calculations and devote it to the problem solving process, (2) test their own ideas and conjectures about random variable behavior, and (3) visually compare their numerical answers to plots obtained with APPL. I am able to provide convincing evidence of well-known theorems, such as the Central Limit Theorem, through exact random variable manipulation instead of simulations. Using APPL has allowed me to give students an intuitive glimpse behind statistical and probabilistic theory while having time for illustrating this theory in applications. This talk will illustrate the usefulness and power of the APPL language.

Use of R as a Toolbox for Mathematical Statistics Education

◆ Nicholas J. Horton, Smith College; Elizabeth R. Brown, University of Washington

Smith College, Dept. of Mathematics, Northampton, MA 01063-0001

nhorton@smith.edu

Key Words: *statistical computing*

The R language, a freely available environment for statistical computing and graphics is widely used in many fields. This “expert-friendly” system has a powerful command language and programming environment, combined with an active user community. We discuss how R is ideal as a platform to support experimentation in mathematical statistics, both at the undergraduate and graduate levels. Using a series of case studies and activities, we describe how R can be utilized in a mathematical statistics course as a toolbox for experimentation. Examples include the calculation of a running variance, maximization of a nonlinear function, resampling of a statistic, simple Bayesian modeling, sampling from multivariate normal, and estimation of power. These activities, often requiring only 10-20 lines of code, offer the student the opportunity to explore statistical concepts and experiment. In addition, they provide an introduction to the framework and idioms available in this rich environment.

Statistical Properties of the Slope Mean

◆ Lewis VanBrackle, Kennesaw State University

Kennesaw State University, Mailbox 1204, 1000 Chastain Rd., Kennesaw, GA 30144

lvanbrac@kennesaw.edu

Key Words: *regression, estimation*

In a recently submitted paper, Ji and Kicey examined the invariance properties of means in general and the “slope mean” in particular. In the context of regression through the origin, the slope mean is defined as the tangent of the mean of the angles between the x-axis and the lines from the origin through the data points. Examining the statistical properties of the slope mean is an excellent exercise for undergraduate mathematics and statistics majors. In the process of evaluating the statistical properties of the slope mean and comparing them to the properties of other slope estimators, such as the average of the slopes of the lines from the origin through the data points and the usual ordinary least squares estimator, students can apply techniques they have learned in a variety of mathematics and statistics courses. This presentation will show how the Fundamental Theorem of Calculus, the Central Limit Theorem, the Gauss-Markov Theorem, Taylor’s series approximations, calculation of probabilities by numerical integration and simulation can all be applied in deriving the statistical properties of the slope mean

2004 ASA PROCEEDINGS

Eligibility

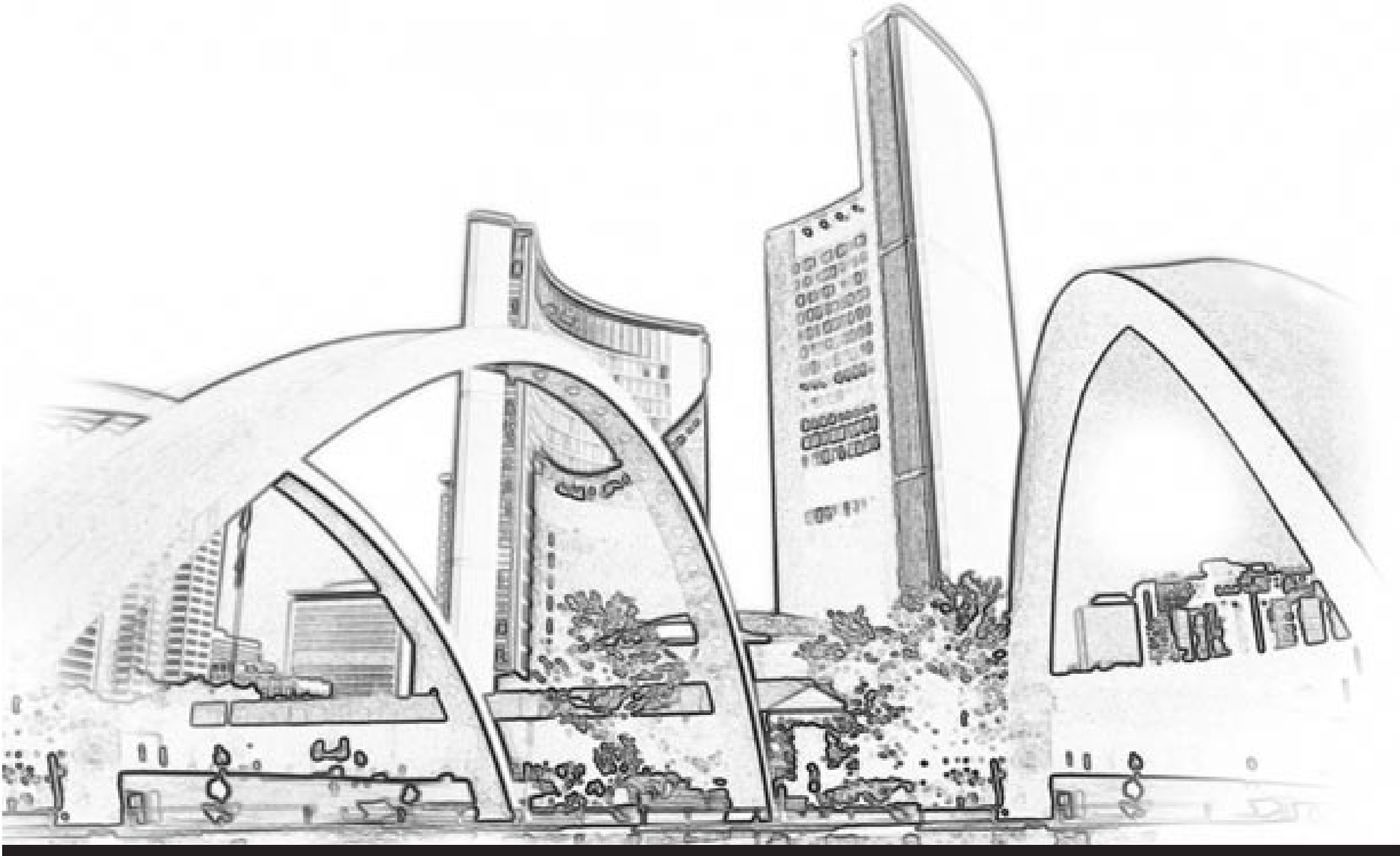
Papers must have been presented at the Joint Statistical Meetings, including poster sessions, or an approved outside conference to be authorized for publication in the Proceedings. If your paper was withdrawn or not orally presented, it is not eligible.

If the **PRIMARY** (first listed) sponsor of your session is one of the following Sections, you are eligible to publish in the 2004 ASA Proceedings:

- Bayesian Statistical Science
- Biometrics Section-to include ENAR & WNAR
- Biopharmaceutical Section
- Business & Economic Statistics Section
- Statistical Computing Section
- Section on Statistical Consulting
- Section on Statistical Education
- Section on Statistics & the Environment
- Section on Statistics in Epidemiology
- Section on Government Statistics
- Section on Statistical Graphics
- Section on Health Policy Statistics
- Section on Nonparametric Statistics
- Section on Physical & Engineering Sciences
- Section on Quality & Productivity
- Section on Risk Analysis
- Social Statistics Section
- Section on Statistics in Sports
- Section on Teaching Statistics in Health Sciences
- Section on Survey Research Methods

For more details, deadlines, and formatting instructions, please go to *www.amstat.org/meetings/jsm/2004*, click on the “Program” tab, and select “Proceedings.”

CONTINUING EDUCATION COURSES



August 7–8, 2004

CE_01C

Sample Survey Methods: Recent Developments and Applications (2-day Course, Day 1)

8:15 am–4:15 pm (August 7)

Instructor(s): Jon N.K. Rao, Carleton University; Sharon Lohr, Arizona State University

ASA

This two-day course will be an introduction to and appraisal of recent developments in sample survey methods. Several topics of practical interest will be covered: survey design and inferential issues, analysis of survey data with emphasis on re-sampling methods, design and estimation in dual frame surveys, variance estimation under imputation for missing data, small area estimation. Applications of the methods in U.S. and Canadian surveys and evaluations of software packages for survey data will also be presented. Prerequisite knowledge is a basic course in survey sampling at the level of W. G. Cochran's 1977 book *Sampling Techniques* or S. Lohr's 1999 book *Sampling: Design and Analysis. Exposure* to basic courses in mathematical statistics and linear models will also be helpful but are not essential.

CE_01C

Sample Survey Methods: Recent Developments and Applications (2-day Course, Day 2)

8:15 am–4:15 pm (August 8)

Instructor(s): Jon N.K. Rao, Carleton University; Sharon Lohr, Arizona State University

ASA

August 7, 2004

CE_02C

Generalized Linear Latent and Mixed Models (1-day Course)

8:00 am–4:00 pm

Instructor(s): Anders Skrondal, Norwegian Institute of Public Health; Sophia Rabe-Hesketh, University of California, Berkeley; Andrew Pickles, University of Manchester

Biometrics Section

Generalized linear mixed (or multilevel) models are useful for longitudinal data, cluster-randomized trials, surveys with cluster-sampling, genetic studies, meta-analysis and many other applications. The random coefficients in generalized linear mixed models are latent variables representing between-cluster variability and inducing within-cluster correlations. Latent variables are also often used to represent true values measured with error, a typical example being diet (continuous latent variable) or diagnosis (categorical latent variable). Measurement models relating measured variables to latent variables can be used to investigate the properties of measurement instruments or diagnostic tests. Such measurement models can also form part of structural equation models relating latent variables to other latent or observed variables. An important application is regression with covariate measurement error. Finally, latent variables can be used to model the dependence between different processes, for instance the response of interest in a clinical trial and the (nonignorable) drop-out process.

All these models have very a similar structure. However, this is not commonly recognized due to disparate terminologies and lack of communication between methodologists in different disciplines, for instance (biostatisticians, econometricians and psychometricians. Taking a unified view is beneficial since developments for one model-type are often applicable to other model-types. Furthermore, the same software can often be used to estimate seemingly different models.

The course will be structured in three parts: (1) generalized linear mixed models, (2) measurement models and (3) structural equation models. In each part, we start with the simplest version of the model, motivating each extension through examples. Methods of estimation and prediction will also be surveyed. We then consider real applications, specifying models to address the research question, interpreting parameter estimates and providing further insight and model diagnostics using graphical displays of both data and model predictions. We will use consistent notation throughout, emphasizing the communalities between model-types.

This course will benefit statisticians and graduate students in statistics familiar with generalized linear models.

CE_03C

Introduction to Clinical Trials (1-day Course)

8:00 am–4:00 pm

Instructor(s): L. Jane Goldsmith, University of Louisville

Biometrics Section

Clinical trials are the heart of biomedical research. Trials for new drugs and other novel treatment developments comprise a large amount of research activity in the drug and medical equipment industry. Academic research efforts in medicine, nursing, and dentistry often culminate in clinical trials as the gold standard for the determination of effective treatment methods for humans. Clinical research has an interesting history and nomenclature as well as standards for design, ethics, and reporting. Statistical methods are used in clinical trials for sampling and randomization strategies for treatment assignment, as well as for study design and analysis of data. Specialized statistical methods are sometimes used, but standard statistical theory can often be applied once the statistician gains the right background knowledge. This one-day seminar serves to introduce statisticians to the praxis of clinical trials.

This seminar will benefit an experienced statistician who wants to learn about clinical trial research and also a new or student statistician who wishes to learn the role of biostatisticians in clinical research.

CE_04C

Practical Bayesian Methods for Evaluating Clinical Interventions (1-day Course)

8:00 am–4:00 pm

Instructor(s): David Spiegelhalter, Institute of Public Health

Biopharmaceutical Section, Section on Bayesian Statistical Science

We may define a Bayesian approach as "the explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a health care evaluation". Such a perspective is claimed to be more flexible than traditional methods, more efficient in using all available evidence, and more useful in providing predictions as a basis for decisions for specific patients, planning research, or public policy. This course is intended to provide a nontechnical review of the essential ideas of Bayesian analysis as applied to the evaluation of health care interventions, brief comparisons with conventional approaches, a wide range of applications on real data, presented in a standardized format, and advice on modeling issues such as prior specification.

The emphasis throughout will be on practical examples: software and code to carry out all the analyses will be freely downloadable from the web.

CE_05C

Modern Data Mining (1-day Course)

8:15 am–4:15 pm

Instructor(s): David Banks, Duke University

ASA

This short course begins with an introductory overview of data mining: its scope, the classical approaches, and the heuristics that guided the initial development of theory and methods. Then the course moves towards the treatment of more modern issues such as boosting, overcompleteness, and large p/small n problems. This leads to a survey of currently popular techniques, including random forests, support vector machines, and wavelets. The main focus is upon regression inference, as this is a paradigm problem that informs all data mining applications, but we also discuss clustering, classification, and multidimensional scaling.

The only prerequisites for the course are a basic knowledge of applied multivariate inference and a general level of statistical knowledge comparable to a weak master's degree. Any math will focus upon conveying general insight rather than specific details.

CE_06C

Statistical Leadership: From Consultant to Effective Leader (1-day Course)

8:15 am–4:15 pm

Instructor(s): Roger Hoerl, General Electric Company ; Ron Snee, Tunnell Consulting; Bill Parr, University of Tennessee; Angela Patterson, General Electric Company

Section on Quality and Productivity, ASA

Many statisticians in consulting roles would like to enhance their impact on the organizations they consult with. While statisticians' contributions may be very significant, all too often the statistician is viewed as playing a supporting, rather than leading role, in project success. Similarly, the statistician is often viewed as someone passively giving advice, or performing narrow technical tasks predetermined by others, rather than proactively making things happen and delivering results. The purpose of this workshop is to enhance the effectiveness of statisticians in business, industry, and pharmaceuticals, particularly as it relates to exhibiting leadership. The goal is to help statisticians transition from being viewed as passive consultants to being viewed as proactive leaders within their organizations. In preparation for this workshop, several senior business leaders, CEOs, and other executives were interviewed to obtain their insights into how statisticians could have more impact. In addition, recent statistics graduates were polled to obtain their perspectives on their current roles, and what skills they feel are necessary to succeed. These insights will be shared and discussed at the workshop. The attendees will participate in breakouts to determine specific changes needed, and how to go about making them happen.

August 8, 2004

CE_07C

Regression Modeling Strategies (1-day Course)

8:00 am–4:00 pm

Instructor(s): Frank E. Harrell, Jr., Vanderbilt University School of Medicine

ASA

All standard regression models have assumptions that must be verified for the model to have power to test hypotheses and for it to be able to predict accurately. Of the principal assumptions (linearity, additivity, distributional), this course will emphasize methods for assessing and satisfying the first two. Practical but powerful tools are presented for validating model assumptions and presenting model results. This course provides methods for estimating the shape of the relationship between predictors and response using the widely applicable method of augmenting the design matrix using restricted cubic splines. Even when assumptions are satisfied, overfitting can ruin a model's predictive ability for future observations. Methods for data reduction will be introduced to deal with the common case where the number of potential predictors is large in comparison with the number of observations. Methods of model validation (bootstrap and cross-validation) will be covered, as will auxiliary topics such as modeling interaction surfaces, efficiently utilizing partial covariable data by using multiple imputation, variable selection, overly influential observations, collinearity, and shrinkage. The methods covered will apply to almost any regression model, including ordinary least squares, logistic regression models, and survival models.

CE_08C

Designing Clinical Trials (1-day Course)

8:00 am–4:00 pm

Instructor(s): Peter F. Thall, University of Texas M.D. Anderson Cancer Center

Section on Bayesian Statistical Science, Biopharmaceutical Section

This one-day short course will cover an array of statistical methods for clinical trials that I have designed during the past 13 years working as a biostatistician at M.D. Anderson Cancer Center (MDACC). These are practical, sequential, outcome-adaptive Bayesian designs, with heavy reliance on computer simulation to establish frequentist operating characteristics. An underlying theme will be that each clinical trial typically has unique aspects that require the design to be tailored to accommodate its particular structure and goals. Each design will be illustrated in the context of one or more specific applications, in addition to presentation of the general underlying probability model and statistical methodology. The illustrative designs have been applied to conduct single-institution clinical trials at MDACC and at other medical centers, as well as multi-institution trials, and several are currently ongoing. The topics listed in the outline will be covered as time permits. While the orientation is toward oncology trials, since that is my main experience, nearly all of the methods have a much broader application.

CE_09C

Bootstrap Methods and Permutation Tests for Doing and Teaching Statistics (1/2-day Course)

8:00 am–12:00 pm

Instructor(s): Tim C. Hesterberg, Insightful Corporation

Section on Statistical Computing, Section on Statistical Education

Early in Stat 101 we teach that robustness is important. Yet later in the course, and too often in statistical practice, we ignore those lessons, and use simple means and least-squares regression together with Normal-based inferences, even though the corresponding assumptions are violated.

Bootstrapping and permutation tests (BPT) let us check the accuracy of common procedures, and the results are surprising. We'll see how inaccurate Normal-based methods are in the presence of even moderate skewness. The old rule of trusting the CLT if n>30 is just old.

BPT let us more easily do inferences for a wider variety of statistics (e.g. trimmed means, robust regression) for data collected in a variety of ways (e.g. stratification). BPT provide output we may graph in familiar ways (like histograms) to help students and clients understand sampling variability, standard errors, p-values, and the Central Limit Theorem (CLT)-not just in the abstract, but for the data set and statistic at hand.

We'll look at applications from a variety of fields, including telecommunications, finance, and biopharm.

CE_10C

Design and Analysis of Gene Expression Studies Using Microarrays (1Day Course)

8:15 am–4:15 pm

Instructor(s): Robert J. Tempelman, Michigan State University; Guilherme J.M. Rosa, Michigan State University

ASA

Many applied statisticians collaborating with genomic researchers have been overwhelmed by a plethora of conflicting recommendations on design and analysis that pervade the microarray data analysis literature. Our course integrates fundamentally important concepts in experimental design and mixed model ANOVA in helping statisticians and their collaborators evaluate design and power considerations for their own studies. The distinction between experimental replication and technical replication or subsampling in these evaluations is particularly addressed as these two levels of replication are often confused with each other in many power and analysis methods papers. Our course is based on materials that we have presented in 4+ different national and international workshops, primarily to biologists. Therefore, our presentation emphasizes sources of variation that have been recognized by biologists and need to be considered by their statistical collaborators. Along those lines, a basic introduction to genomics research with a general overview of different microarray technology platforms with implications for design and analysis is provided as well. A fundamental background in experimental design is required whereas training in mixed model analysis and some experience with SAS software is desirable but not required.

CE_11C

Longitudinal and Incomplete Data (1-day Course)

8:15 am–4:15 pm

Instructor(s): Geert Verbeke, Katholieke Universiteit Leuven; Geert Molenberghs, Limburgs Universitair Centrum

Section on Statistics in Epidemiology

A general introduction to longitudinal data and the linear mixed model for continuous responses will be presented. The topic will be approached from the modeler's and practitioner's points of view, with emphasis on model formulation, inference and parameter interpretation. Illustrations will be given based on the SAS procedure MIXED.

When the response of interest is categorical, the linear mixed model concepts can be extended towards generalized linear mixed models. An alternative approach is the use of generalized estimating equations (GEE). A lot of emphasis will be put on the fact that the regression parameters in both types of models have different interpretations, and illustrations will be based on the SAS procedures GENMOD and NLMIXED.

Finally, when analyzing longitudinal data, one is often confronted with incomplete observations, i.e., scheduled measurements have not been made, due to a variety of (known or unknown) reasons. It will be shown that, if no appropriate measures are taken, missing data can cause seriously biased results, and interpretational difficulties.

Throughout the course, it will be assumed that the participants are familiar with basic statistical modeling, including linear and generalized linear models.

CE_12C

Methods for Designing and Analyzing Mixture Experiments (1-day Course)

8:15 am–4:15 pm

Instructor(s): Greg F. Piepel, Battelle/Pacific Northwest National Laboratory; John A. Cornell, University of Florida

Section on Physical and Engineering Sciences

Mixture experiments involve changing the proportions of product components and observing the changes in the product's characteristics. Mixture component proportions cannot be varied independently (as in factorial experiments) because they must sum to 1.0 for each run in the experiment. Mixture experiments are very useful in many product development areas, including foods, materials, fertilizers, textile fibers, drugs, and many others.

The short course will provide an overview of methods used in designing mixture experiments and analyzing the resulting data. Topics to be covered include: (1) designs for simplex-shaped and irregular-shaped regions (the latter resulting from additional constraints on the component proportions), (2) various types of mixture models for fitting mixture data, (3) graphical techniques for interpreting component effects, (4) including process variables and/or a total amount variable in mixture experiments, and (5) graphical and analytic methods for developing mixtures with optimum properties. Numerous examples will be used to illustrate the topics discussed.

The course is designed for anyone (statistician or nonstatistician) wanting to know about statistical methods for designing mixture experiments and analyzing the resulting data. Prerequisites are an understanding of elementary statistics concepts and some previous exposure to experimental design and least squares regression.

CE_13C

Disease Mapping and Risk Assessment for Public Health (1/2-day Course)

1:00 pm–5:00 pm

Instructor(s): Andrew B. Lawson, University of South Carolina

Section on Statistics and the Environment, ASA

This course provides participants with a basic introduction to the concepts and methods useful in the analysis of geo-referenced small area health data, with emphasis on public health and epidemiological applications. The course covers basic spatial epidemiology concepts; statistical background; geographical information systems uses; examples of public health applications such as cluster detection, resource allocation and ecological analysis.

August 9, 2004

CE_14C

Statistical Rules of Thumb (1/2-day Course)

8:00 am–12:00 pm

Instructor(s): Gerald van Belle, University of Washington

Puget Sound Chapter (PUGECH), ASA

This course is based on the recently released book, Statistical Rules of Thumb (Wiley, 2002), by Gerald van Belle. A statistical rule of thumb is defined as a widely applicable guide to statistical practice-with sound theoretical basis. Characteristics include intuitive appeal, elegance, and transparency. A rule states not only what is important but, by implication of what is not included, makes an assertion about what is less important.

For this workshop three areas of rules have been selected: design of a study, its analyses, and reporting it. Rules covered under design include selection of study design; three principles of design: randomization, blocking, factorial treatment structure; and sample size calculations. Analysis deals with the following rules: analysis should follow design, parametric analyses are preferred, plan for multiple comparisons, and every analysis has a graphical analogue. Rules of thumb with respect to reporting of results deal with when to use words, when tables, when graphs. All rules will be illustrated and appropriate references will be supplied. Participants will be encouraged to apply the rules presented and to think of additional rules of thumb for their area of expertise.

CE_15C

Introduction to Survey Quality (1-day Course)

8:00 am–4:00 pm

Instructor(s): Paul P. Biemer, RTI International and UNC-CH; Lars E. Lyberg, Statistics Sweden

Section on Survey Research Methods, ASA

The course will span a range of topics dealing with the quality of data collected through the survey process. Total survey error, as measured by the mean squared error and its component parts, is the primary criterion for assessing the quality of the survey data. The course begins with a discussion of total survey error and its relationship to survey costs and provides a number of measures of survey quality that will be used throughout the course. Then the major sources of survey error are discussed in some detail. In particular, we examine (a) the origins of each error source (i.e., its root causes), (b) the most successful methods that have been proposed for reducing the errors emanating from these error sources, and (c) methods that are most often used in practice for evaluating the effects of the source on total survey error.

The target audience for the course is persons who perform tasks associated with surveys and may work with survey data but are not necessarily trained survey researchers. These are survey project directors, data collection managers, survey specialists, statisticians, data processors, interviewers, and other operations personnel who would benefit from a better understanding of the concepts of survey data quality, including sampling error and confidence intervals, validity, reliability, mean squared error, cost-quality trade offs in survey design, nonresponse error, sampling frame error, measurement error, specification error, data processing error, methods for evaluating survey data, and how to optimize survey quality through the optimal use of survey resources.

CE_16C

Statistical Methods in Diagnostic Medicine (1-day Course)

8:15 am–4:15 pm

Instructor(s): Xiao-Hua A. Zhou, University of Washington

Section on Statistics in Epidemiology

The purpose of this one-day short course is to provide a comprehensive account of statistical methods for design and analysis of diagnostic studies, including sample size calculations, estimation of the accuracy of a diagnostic test, comparison of accuracies of competing diagnostic tests, and regression analysis of diagnostic accuracy data. Additionally, some recently developed methods for correction of verification bias and imperfect reference bias, and meta-analysis methods will be discussed.

This short course will be based on the instructor's recent textbook entitled "Statistical Methods in Diagnostic Medicine" that was published by Wiley & Sons in year 2002

CE_17C

Splines, Knots, and Penalties: The Practice of P-spline Smoothing (1-day Course)

8:15 am–4:15 pm

Instructor(s): Brian Marx, Louisiana State University; Paul Eilers, Leiden University

ASA

Smoothing has become an established part of the statistical toolbox and it is finding its way into most of the statistical software packages. There exists a wide choice of methods, like smoothing splines, regression splines, kernel smoothing and local likelihood. It is difficult for the novice to make an informed choice.

The main theme of our course is the use of a combination of local basis functions (B-splines) and roughness penalties: the P-spline approach. This smoother stays as close as possible to established regression methodology, with many advantages:

- compact computations;
- compact results (a small set of coefficients);
- easily computed diagnostics: cross-validation or AIC, standard errors, effective degrees of freedom;
- parametric models are obtained as limiting cases of heavy smoothing;
- straightforward extension to non-normal data, along the lines of the generalized linear model;
- useful for density estimation, with well-behaved boundaries;
- P-splines are ideal building blocks for generalized additive models, varying-coefficient models and signal regression.

We present the basic theory and illustrate it with many practical examples: scatterplot smoothing, trend lines, density estimation, dose-response curves, smooth high dimensional coefficient vectors, generalized additive models, and mixed models.

There will be no hands-on computer exercises, but we will supply a CD with software, example data and demonstration scripts for S-Plus (or R freeware) and Matlab. Their use will be illustrated during the course. After taking this course one will be able to go home to analyze his/her own data.

A background in multiple regression and familiarity with basic matrix operations is needed.

CE_18C

Bayesian Inference (1-day Course)

8:15 am–4:15 pm

Instructor(s): Bruno Sansó, University of California, Santa Cruz

Section on Bayesian Statistical Science

Bayesian methods have become increasingly popular with the advent of fast computational algorithms for the exploration of high dimensional probability distributions. The Bayesian paradigm provides a coherent framework to build models of high complexity, incorporate quantitative and structural prior information and account for all uncertainties in a probabilistic way. This course reviews the bases of Bayesian inference.

The course will start by presenting the basic elements of statistical inference that uses likelihood functions. We will then consider the problem of specifying prior distributions, proceed by describing the tools for both pointwise and interval estimation and prediction and present the Bayesian theory of hypothesis testing and model comparison.

Finally we will review the elements of modern computational methods used in the applications of Bayesian models.

The course targets students or professionals with a good knowledge of statistics that want to learn or refresh their knowledge of basic Bayesian inference. The level of mathematical sophistication will be kept as low as possible. Calculus and basic probability theory are considered a pre-requisite.

August 10, 2004

CE_19C

Modeling Financial Derivatives for the Management of Risk (1/2-day Course)

8:00 am–12:00 pm

Instructor(s): Sean Chen, Fordham University

Section on Risk Analysis, ASA

This half-day course introduces basic stochastic models for financial derivatives such as options and futures, which are important instruments in risk management, and the statistical issues involved in the implementation of these models. The story of financial derivatives is told from two perspectives, pricing and trading, which are balanced through Arbitrage Pricing Theory. The course combines theoretical and practical aspects of option pricing and trading, using real world examples for illustration. Statistical issues such as model diagnosis, volatility estimation, and computing/sampling methods are discussed.

Prerequisite for the course is basic knowledge of probability theory, including mean, variance, conditional probability, normal distribution, binomial distribution.

CE_20C

Modeling and Analysis of Categorical Data with Overdispersion (1-day Course)

8:00 am–4:00 pm

Instructor(s): Jorge G. Morel, Procter & Gamble Company; Nagaraj K. Neerchal, University of Maryland, Baltimore County

Biopharmaceutical Section

The aim of the course is to present a general overview of the phenomenon of overdispersion relative to the binomial/multinomial and Poisson distributions, as well as to provide different methods (quasi-likelihood, likelihood, generalized estimating equations, and generalized linear mixed models) to cope with this problem. Several practical examples will be shown to illustrate the available methodology to model categorical data with overdispersion. Some of the main examples will be analyzed using the SAS Procedures GENMOD and SURVEYLOGISTIC, and the SAS Macro GLIMMIX. The course is at the level of an applied master's degree. It will be also accessible to those with a bachelor's degree and adequate work experience. Basic knowledge of the binomial, Poisson and multinomial distributions, logistic and Poisson regressions, maximum likelihood estimation and Fisher's information matrix is needed. Introductory knowledge of SAS would be helpful to take full advantage of the SAS codes and the annotated outputs.

CE_21C

Intermediate/Advanced Bayesian Hierarchical Modeling (1-day Course)

8:00 am–4:00 pm

Instructor(s): David Draper, University of California, Santa Cruz

ASA

This course provides coverage of intermediate and advanced topics arising in the formulation, fitting, and checking of hierarchical or multilevel models from the Bayesian point of view. Hierarchical models (HMs) arise frequently in four main kinds of applications:

- HMs are common in fields such as health and education, in which data — both outcomes and predictors — are often gathered in a nested or hierarchical fashion: for example, patients within hospitals, or students within classrooms within schools. HMs are thus also ideally suited to the wide range of applications in government and business in which single- or multi-stage cluster samples are routinely drawn, and offer a unified approach to the analysis of random-effects (variance-components) and mixed models.
- A different kind of nested data arises in meta-analysis in, e.g., medicine and the social sciences. In this setting the goal is combining information from a number of studies of essentially the same phenomenon, to produce more accurate inferences and predictions than those available from any single study. Here the data structure is subjects within studies, and as in the clustered case above there will generally be predictors available at both the subject and study levels.
- When individuals—in medicine, for instance—are sampled cross-sectionally but then studied longitudinally, with outcomes observed at multiple time points for each person, a hierarchical data structure of the type studied in repeated-measures or growth curve analyses arises, with the readings at different time points nested within person.
- Hierarchical modeling also provides a natural way to treat issues of model selection and model uncertainty with all types of data, not just cluster samples or repeated measures outcomes. For example, in regression, if the data appear to exhibit residual variation that changes with the predictors, you can expand the model that assumes constant variation, by embedding it hierarchically in a family of models that span a variety of assumptions about residual variation. In this way, instead of having to choose one of these models and risk making the wrong choice, you can work with several models at once, weighting them in proportion to their plausibility given the data.

The Bayesian approach is particularly effective in fitting hierarchical models, because other model-based methods—principally involving maximum likelihood—often do not capture all relevant sources of uncertainty, leading to overconfident decisions and scientific conclusions.

In this course the basic principles of Bayesian hierarchical modeling are reviewed, with emphasis on practical rather than theoretical issues, and intermediate- and advanced-level ideas are illustrated with real data drawn from case studies involving complicated applications of HMs in cluster sampling and mixture modeling. The course is intended for applied statisticians with an interest in learning more about intermediate and advanced topics in hierarchical modeling in general, and the Bayesian analysis of such models in particular. An understanding of probability and statistics at the level typically required for a master's degree in statistics provides sufficient mathematical background.

This course is intended to be a follow-on from an introductory treatment of Bayesian hierarchical modeling, so I will assume that participants have background in Bayesian methods and hierarchical modeling at the level of the first six chapters of the textbook by Gelman et al. (*Bayesian Data Analysis*, second edition, 2003) or equivalent.

CE_22C

Missing Data Methods in Regression Models (1-day Course)

8:00 am–4:00 pm

Instructor(s): Joseph G. Ibrahim, University of North Carolina, Chapel Hill; Ming-Hui Chen, University of Connecticut

Section on Bayesian Statistical Science

Statistical inference with missing data is a very important problem since missing values are frequently encountered in practice. In fact, most statistical problems can be considered incomplete because not all variables are observed for each unit (or possible unit) in a study. For example, randomization in a clinical trial generates missing values since the outcome that would have been observed had a subject been randomized to a different treatment group is not observed. Missing values can be both planned and unplanned. Unplanned missing data can arise when study subjects fail to report to a clinic for monthly evaluations, when respondents refuse to answer certain questions on a questionnaire, or when data is lost. On the other hand, data can be missing by design in a randomized clinical trial or in a Latin square experimental design. Although the problems associated with incomplete data are well-known, they are often ignored, and the analysis is restricted to those observations with complete data. This method of analysis is still the default method in most software packages despite the development of statistical methods that handle missing data more appropriately. In particular, likelihood-based methods, multiple imputation, methods based on weighted estimating equations, and fully Bayesian methods have gained increasing popularity since they have become more computationally feasible in recent years. In this short course, we examine each of these methods in some detail, and compare and contrast them under various settings. In particular, we will examine missing covariate and response data in generalized linear models, random effects models, and survival models. Ignorable missingness as well as nonignorable missingness will be presented for theses models, as well as frequentist and Bayesian methods for analysis. The newly developed statistical package XMISS (Cytel Software) will be used and demonstrated for several real data examples. In addition, live demos of the XMISS software and data analysis using the various models will be given using an LCD projector.

The course presents a balance between theory and applications, and for each class of methods and models discussed, detailed examples and analyses from case studies are presented whenever possible. The applications are all essentially from the health sciences including cancer, AIDS, epidemiology, and the environment. Overall, this course will be applied in nature and will focus on the applications of frequentist and Bayesian methods for research problems arising in the medical sciences. Live demo real data examples will be given using the XMISS software.

CE_23C

Statistical Methods in Bioinformatics (1-day Course)

8:15 am–4:15 pm

Instructor(s): Jun S. Liu, Harvard University

Section on Bayesian Statistical Science

A substantial core of computational biology (or bioinformatics) methods has been developed during the past three decades to meet the need of biological scientists for data storage, data retrieval, and data analysis. The databases of DNA and protein sequences contain millions of sequences, many completed genomes, and more are coming rapidly. DNA microarray data are being produced at a phenomenal speed. Protein arrays are being developed. High throughput structural data are being produced. Analysis of these data using bioinformatics tools has played a key role in several recent advances and will play increasingly important roles in future biomedical researches.

A main problem that motivated early research in computational biology is protein sequence analysis. Recently, because of the dramatic increase in many types of biological data due to the human genome project and other high-throughput projects, the scope of bioinformatics research has been extended to embrace diverse topics such as microarray analysis, protein classification, regulatory motif analysis, RNA analysis etc.

The sheer amount and variety of the molecular biology data have already presented a major challenge to all quantitative researchers. A distinctive feature of these data, be they microarray images, DNA sequences or protein structures, is that there is a large body of biological knowledge associated with them. This makes standard data mining or statistical analysis tools less effective. Incorporating relevant scientific knowledge into the development of statistical or computational analysis tools is the key to success.

This short course is intended to provide coverage of some key developments of bioinformatics in the past thirty years with an emphasis on topics of recent interest. Topics include: pair-wise sequence analysis, local alignment, dynamic programming, BLAST, multiple sequence alignment, Gibbs motif sampler, gene regulation, hidden Markov models, context-free grammars, protein structure analysis, comparative genomics, model-based microarray analysis, clustering methods for microarrays, phylogenetic trees, etc.

CE_24C

Writing Effectively: Communicating with Nonstatisticians (1-day Course)

8:15 am–4:15 pm

Instructor(s): Thomas A. Lang, Tom Lang Communications; Ralph G. O'Brien, Cleveland Clinic Foundation

Section on Teaching Statistics in the Health Sciences

Even the best statistical work has value only when it is communicated effectively to relevant audiences. Most often, this communication occurs through technical reports and scientific articles. Participants will learn several techniques for communicating scientific research, especially statistical information, to non-statisticians. The presentation has four core topics: 1) the Perspective of Communication, 2) The Statistical Report and the Scientific Article, 3) Reporting Statistics in Medicine, and 4) Written Communication Skills. Instruction will be by lecture-discussion and by individual and group activities.

There are no prerequisites. The workshop is not a review of English composition and is not intended to improve the English capabilities of nonnative English speakers.

CE_25C

Statistical Analysis and Data Display (1-day Course)

8:15 am–4:15 pm

Instructor(s): Richard Heiberger, Temple University; Burt Holland, Temple University

Section on Statistical Graphics

The course is based on the presenters' forthcoming book, *Statistical Analysis and Data Display: An Intermediate Course*, due from Springer later this year.

We emphasize the graphical display of data. We show graphs, how to construct and interpret them, and how they relate to the tabular outputs that appear automatically when a statistical program "analyzes" a data set. The graphs are not automatic and so must be requested. Gaining an understanding of a data set is always more easily accomplished by looking at appropriately drawn graphs than by examining tabular summaries. In our opinion, graphs are the heart of most statistical analyses; the corresponding tabular results are formal confirmations of our visual impressions. We use graphs for statistical research and analysis, that is for exploring and designing new techniques of analysis, as well as for analysis.

Prerequisites: Some knowledge of regression and analysis of variance. Familiarity with statistical software.

CE_26C

Analysis of Multivariate Survival Data: An Introduction to Frailty Models (1-day Course)

8:15 am–4:15 pm

Instructor(s): Philip Hougaard, H. Lundbeck A/S

Section on Statistics in Epidemiology, Biometrics Section, ASA

The course starts with a brief introduction to survival data to get a common notation. Frailty models for univariate data are described showing the mathematics behind mixture evaluations. The univariate data frailty model can model heterogeneity and extend standard models (like the proportional hazards to nonproportional hazards). Multivariate data examples are described and classified in six types. Probability mechanisms for dependence and correlation-like measures of dependence are described. The shared frailty model for multivariate data is the central model, allowing both covariate effects and dependence. The course emphasizes choice of model, advantages and disadvantages of each model, interpretation and applications. Many applications are considered, including the survival of twins. The course covers parametric and nonparametric models and proportional hazards models and accelerated failure time models. Also recurrent events data are considered. Finally extensions of the model are described briefly to discuss shortcomings of assuming shared frailty. Software examples show gamma frailty models fitted by Splus. Course book: Hougaard *Analysis of multivariate survival data*, Springer 2000.

Pre-requisites: Standard survival data methods (Censoring, truncation, Nelson-Aalen, Cox proportional hazards, accelerated failure times)

CE_27C

Meta-analysis: Statistical Methods for Combining the Results of Independent Studies (1/2-day Course)

1:00 pm–5:00 pm

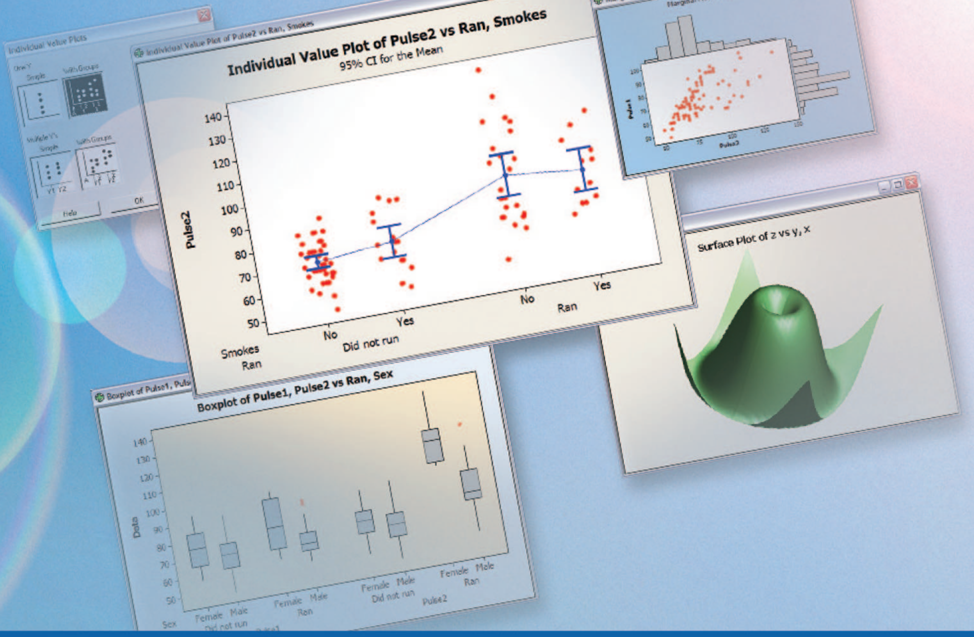
Instructor(s): Ingram Olkin, Stanford University

ASA

Meta-analysis enables researchers to synthesize the results of a number of independent studies designed to determine the effect of an experimental protocol such as an intervention, so that the combined weight of evidence can be considered and applied. Increasingly meta-analysis is being used in the health sciences, education and economics to augment traditional methods of narrative research by systematically aggregating and quantifying research literature. The information explosion in almost every field coupled with the movement towards evidence based decision making and cost-effective analysis has served as a catalyst for the development of procedures to synthesize the results of independent studies.

INDEX

CE INSTRUCTORS INDEX



Individual Value Plot of Pulse2 vs Ran, Smokes
95% CI for the Mean
Pulse2
Smokes
Ran
No
Yes
Did not run
Ran
Boxplot of Pulse1, Pulse2 vs Ran, Sex
Data
Female
Male
Did not run
Ran
Pulse2
Sex
Surface Plot of z vs y, x