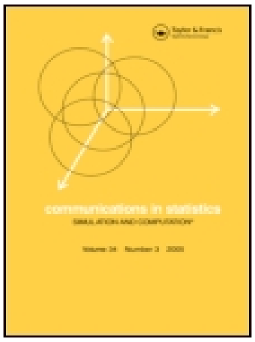

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

An Optimal Confidence Region for the Largest and the Smallest Means from a Multivariate Normal Distribution

Shu-Fei Wu & Hubert J. Chen

To cite this article: Shu-Fei Wu & Hubert J. Chen (2016) An Optimal Confidence Region for the Largest and the Smallest Means from a Multivariate Normal Distribution, Communications in Statistics - Simulation and Computation, 45:3, 952-967, DOI: 10.1080/03610918.2014.889160

To link to this article: https://doi.org/10.1080/03610918.2014.889160

Accepted author version posted online: 23 Jun 2014.
Published online: 03 Nov 2015.

Communications in Statistics—Simulation and Computation®, 45: 952–967, 2016

ISSN: 0361-0918 print / 1532-4141 online
DOI: 10.1080/03610918.2014.889160

An Optimal Confidence Region for the Largest and the Smallest Means from a Multivariate Normal Distribution

SHU-FEI WU[1] AND HUBERT J. CHEN[2]

[1]Department of Statistics, Tamkang University, Tamsui, Taiwan R.O.C.
[2]Department of Statistics, The University of Georgia, Athens, Georgia, USA

An optimal confidence region is proposed for obtaining the largest and the smallest means of a multivariate normal distribution with a common unknown variance and a non-negative correlation coefficient. The confidence region is compared to a class of asymmetric confidence regions, and the results show that it is uniformly better. Further, a design-oriented two-stage confidence region with a fixed width is also given. Finally, the optimal confidence region is applied to an experiment to measure the treatment effectiveness of a physical therapy with four independent groups and the result reveals that the proposed confidence region can provide some useful information.

Keywords Confidence region; Fixed-width; Largest (smallest) mean; Least favorable configuration; Student *t*

Mathematics Subject Classification 46N30

1. Introduction

The topic of interval estimation for the largest mean of several independent normal populations under certain random variable has been studied by many researchers in the past. When the common population variance is known, Saxena and Tong (1969) and Dudewicz (1972) analyzed symmetric and asymmetric confidence intervals that were not optimal. Later on, the optimal problem was solved by Dudewicz and Tong (1971) in their work on optimal confidence interval for the largest mean and then Tong (1973) provided percentage points for it. When the common population variance is unknown, Chen and Dudewicz (1973) proposed a class of confidence intervals for the largest normal mean. Saxena (1976) proposed a confidence interval for the largest mean based on a large sample approximation. Chen and Chen (1999) proposed a nearly optimal confidence interval which improves the one by Chen and Dudewicz (1973) for the largest mean. No optimal solution was found until a recent breakthrough by Chen and Chen (2004) who developed an optimal confidence interval for the largest normal mean under homoscedasticity to conclude a longtime investigation in this area. In situations where the population variances are unequal and unequal, Chen and Wen (2006) proposed an optimal confidence interval for the largest normal mean. For

Received September 23, 2013; Accepted January 6, 2014
Address correspondence to Hubert J. Chen, Department of Statistics, The University of Georgia, Athens, GA, USA; E-mail: hjpcheney@yahoo.com

several correlated normal populations or equivalently a multivariate normal distribution of a random vector, with a common non-negative correlation coefficient, Chen et al. (2008) proposed an optimal confidence interval for the largest population mean and it was applied to U.S. stock mutual fund evaluation. To extend the aforementioned inference to a more general case, Chen and Wu (2011) proposed a confidence region for the largest and smallest population means of several independent normal populations under heteroscedasticity. Based on the result of Chen et al. (2008), we propose, in this paper, an optimal confidence region for the largest and the smallest means from a multivariate normal distribution, or equivalently, of several correlated normal populations. By ”optimal” we mean that the individual interval width in the component of the confidence region is the shortest one at some LFC among the class of confidence regions obtained by Bonferroni inequality. More specifically, the optimal confidence region is obtained by the following steps: For an individual interval on the largest population mean with a fixed expected interval width, one first attains a least favorable configuration (LFC) of the means such that the infimum of the coverage probability of the interval for the largest population mean and the smallest one, respectively, over the set of all possible values of means and variance is achieved, and then locates the optimal choice of the critical values so that the maximum of the infimum of the coverage probability attains a preassigned probability level. Finally, the Bonferroni inequality is employed to obtain the confidence region for the largest and smallest populations means. Each individual interval component in the proposed confidence region is asymmetric about its best mean due to the fact that the largest sample mean overestimates the largest population mean and the bias increases as the number of populations increases as argued by Dudewicz (1972). Therefore, in the confidence region, it is necessary to shift more of the individual interval for the largest (smallest) population mean to the left (right) of the largest (smallest) sample mean. By taking a negative sign of all observations, the largest population mean becomes the smallest one, and consequently, the smallest sample mean turns out to underestimate the smallest population mean. Owing to such symmetry the calculation of the confidence region becomes feasible. In real world problems, a confidence region for both the largest and the smallest population means can tell how good and how bad about the selected best and the worst ones are in ranking and selection problems. In Section 2 we provide the technical part of an optimal confidence region for the largest and smallest population means under a multivariate a normal population with a common non-negative correlation coefficient. An algorithm to obtain the critical values for the confidence region is given and tables of needed critical values are calculated. In addition, if an experimenter specifies a desired fixed width for each individual interval in the confidence region, a design-oriented two-stage asymmetric confidence region for the largest and the smallest population means under correlated case is proposed in Section 3, and thereafter the needed sample size can be determined. In Section 4, the proposed optimal confidence region is compared to a class of nonsymmetric confidence regions formulated by intercepting a lower and an upper interval; we can show that the proposed optimal confidence region is uniformly better (in the sense of a shorter interval width) than the one by intercepting two one-sided confidence intervals. In Section 5, this method is applied to treatment effectiveness of a physical therapy among four independent groups and the result reveals that the proposed confidence region can provide some useful information on real world problems. At last, in Section 6, a summary and conclusion is made to conclude the findings of this research. It is recommended that the joint confidence region for the largest and the smallest population means be employed in real world problems whenever the best and the worst scenarios are simultaneously interested.

2. An Optimal Confidence Region for the Largest and Smallest Means

Let $\{(X_{1l}, ..., X_{kl}), l = 1, ..., n\}$ be a random vector sample of size n drawn from a k-variate normal distribution with a mean vector of $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)$, a common unknown variance of σ^2 and a non-negative correlation coefficient of ρ. Let $(\bar{X}_1, \ldots, \bar{X}_k)$ be the sample mean vector and $Cov = \{s_{ij}, i, j = 1, \ldots, k\}$ the sample covariance matrix based on the random vector sample of size n. Then, the well-known unbiased variance estimate of σ^2 is defined by

$$S^2 = \frac{\{1 + (k-2)\rho\} \sum_{i=1}^{k} s_{ii} - 2\rho \sum_{i>j} s_{ij}}{k(1-\rho)\{1 + (k-1)\rho\}} \tag{1}$$

with $\nu = k(n-1)$ degrees of freedom (d.f.), independent of $\bar{X}_i$'s (see Johnson and Kotz, 1972; Johnson and Wichern, 2002). Let $\bar{X}_{[1]} \le \bar{X}_{[2]} \le ... \le \bar{X}_{[k]}$ and $\mu_{[1]} \le \mu_{[2]} \le ... \le \mu_{[k]}$ be, respectively, the ordered sample means and ordered population means in the mean vectors where the $\bar{X}_{[i]}$'s are known sample mean values while $\mu_{[i]}$'s are unknown ordered population means. For a prespecified probability $P^*(1/k^2 < P^* < 1)$, consider a confidence region for the largest mean $\mu_{[k]}$ and the smallest mean $\mu_{[1]}$ simultaneously by

$$I = \{\mu_{[k]} \in I_1 \text{ and } \mu_{[1]} \in I_2\}, \tag{2}$$

where $I_1 = (\bar{X}_{[k]} - d_1 S/\sqrt{n}, \bar{X}_{[k]} + d_2 S/\sqrt{n})$ is a confidence interval for $\mu_{[k]}$ and $I_2 = (\bar{X}_{[1]} - d_3 S/\sqrt{n}, \bar{X}_{[1]} + d_4 S/\sqrt{n})$ is a confidence interval for $\mu_{[1]}$, and d_1, d_2, d_3 and d_4 are critical values, such that the coverage probability of the largest mean $\mu_{[k]}$ being included in the interval I_1 and the smallest mean $\mu_{[1]}$ being included in the interval I_2 is at least P^*. To be more specific, the coverage probability of the confidence region for $\mu_{[k]}$ in I_1 and $\mu_{[1]}$ in I_2 can be expressed by

$$\begin{aligned}
& P(\mu_{[k]} \in I_1, \mu_{[1]} \in I_2) \\
&= P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n}, \bar{X}_{[1]} - d_3 S/\sqrt{n} < \mu_{[1]} < \bar{X}_{[1]} + d_4 S/\sqrt{n}) \\
&\ge P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n}) \\
&+ P(\bar{X}_{[1]} - d_3 S/\sqrt{n} < \mu_{[1]} < \bar{X}_{[1]} + d_4 S/\sqrt{n}) - 1,
\end{aligned} \tag{3}$$

where the above probability lower bound (3) is obtained by use of Bonferroni inequality. Let $Y_{il} = -X_{il}$, then the random vector $(Y_{1l}, ..., Y_{kl})$ has a k-variate normal population with a unknown mean of $-\mu_i$, a common unknown variance of σ^2, and the same non-negative correlation coefficient ρ, where the random vector $(Y_{1l}, ..., Y_{kl})$ also represents the lth observation in the random sample of size n, $l = 1, ..., n$. After the transformation, let $\bar{Y}_i = -\bar{X}_i$, $\bar{Y}_i$ is associated with a mean of $-\mu_i$ and $\bar{Y}_{[k]} = -\bar{X}_{[1]}$ is associated with the largest mean $\tilde{\mu}_{[k]} = -\mu_{[1]}$. Then, the above probability lower bound (3) of the confidence region (2) can be rewritten as

$$\begin{aligned}
& P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n}) + P(\bar{Y}_{[k]} - d_4 S/ \\
& \sqrt{n} < \tilde{\mu}_{[k]} < \bar{Y}_{[k]} + d_3 S/\sqrt{n}) - 1.
\end{aligned} \tag{4}$$

The two probability statements in (4) are actually the individual probability coverage for the largest and the smallest population means, respectively. One may assign different coverage probabilities to the individual intervals I_1 and I_2, respectively; but, it is reasonable to assign equal coverage probability to each of the two individual intervals due to symmetry of a marginal normal distribution with a common variance. Under such consideration, the optimal choice of the constants d_3 and d_4 in the coverage probability for the smallest mean P($\mu_{[1]} \in I_2$) will be equivalent to those of d_1 and d_2 in the coverage probability for the largest mean P($\mu_{[k]} \in I_1$), i.e., $d_4 = d_1$ and $d_3 = d_2$. Without loss of generality, one may write the joint coverage probability in (3) as

$$\begin{aligned} &P(\mu_{[k]} \in I_1, \mu_{[1]} \in I_2) \\ &\geq 2P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n}) - 1. \\ &= 2\beta_{\boldsymbol{\delta}}(d_1, d_2) - 1, \end{aligned} \tag{5}$$

where $\beta_{\boldsymbol{\delta}}(d_1, d_2) = P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n})$ is a function of $\boldsymbol{\delta}$, and $\boldsymbol{\delta} = (\delta_1, ..., \delta_k)$ with $\delta_i = \sqrt{n}(\mu_{[k]} - \mu_i)/\sigma \geq 0, i = 1, ..., k$. Observe in (5) that

$$\begin{aligned} &\beta_{\boldsymbol{\delta}}(d_1, d_2) \\ &= P(\bar{X}_{[k]} - d_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + d_2 S/\sqrt{n}) \\ &= P(\bar{X}_i \leq \mu_{[k]} + d_1 S/\sqrt{n}, i = 1, \ldots, k) - P(\bar{X}_i \leq \mu_{[k]} - d_2 S/\sqrt{n}, i = 1, \ldots, k) \\ &= P(Z_i \leq \delta_i + d_1 Y, i = 1, \ldots, k) - P(Z_i \leq \delta_i - d_2 Y, i = 1, \ldots, k), \end{aligned} \tag{6}$$

where ($Z_i = \sqrt{n}(\bar{X}_i - \mu_i)/\sigma, i = 1, \ldots, k$) follows a k-variate normal distribution with a mean vector of zero, a common variance of 1 and a common non-negative correlation coefficient of ρ denoted by $N_k(0, 1, \rho)$, and the r.v. $Y = S/\sigma$ is distributed as the root of Chi-square divided by $\sqrt{\nu}$ with $\nu = k(n-1)$ d.f., independent of the random vector ($Z_1, ..., Z_k$). To further reduce the dimension of the k-variate normal distribution in (6) for ease of calculation, let $W, W_1, \ldots, W_k$ be independent and identically distributed ($i.i.d.$) r.v.'s each having a standard normal distribution with a mean of 0 and a variance of 1 such that the component Z_i can be expressed as $Z_i = \sqrt{1-\rho}W_i - \sqrt{\rho}W, i = 1, \cdots, k$. By such transformation the lower bound of the coverage probability (5) of the confidence region for $\mu_{[k]}$ in I_1 and $\mu_{[1]}$ in I_2 simultaneously in Expression (2) can be can be expressed as

$$\begin{aligned} &2\beta_{\boldsymbol{\delta}}(d_1, d_2) - 1 \\ &= 2\Big[P(\sqrt{1-\rho}W_i - \sqrt{\rho}W \leq \delta_i + d_1 Y, i = 1, \ldots, k) \\ &\quad - P(\sqrt{1-\rho}W_i - \sqrt{\rho}W \leq \delta_i - d_2 Y, i = 1, \ldots, k)\Big] - 1 \\ &= 2\left[\int_0^\infty \int_{-\infty}^\infty \left\{\prod_{i=1}^k P(W_i \leq (\delta_i + d_1 y + \sqrt{\rho}w)/\sqrt{1-\rho}) \right.\right. \\ &\quad \left.\left. - \prod_{i=1}^k P(W_i \leq (\delta_i - d_2 y + \sqrt{\rho}w)/\sqrt{1-\rho})\right\} \phi(w) g_\nu(y) dw dy\right] - 1 \\ &= 2\left[\int_0^\infty \int_{-\infty}^\infty \left\{\prod_{i=1}^k \Phi((\delta_i + d_1 y + \sqrt{\rho}w)/\sqrt{1-\rho}) \right.\right. \\ &\quad \left.\left. - \prod_{i=1}^k \Phi((\delta_i - d_2 y + \sqrt{\rho}w)/\sqrt{1-\rho})\right\} \phi(w) g_\nu(y) dw dy\right] - 1, \end{aligned} \tag{7}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function (c.d.f.) and the probability density function (p.d.f.) of a standard normal r.v., respectively and $g_\nu(\cdot)$ denotes the p.d.f. of a $\sqrt{\chi_\nu^2/\nu}$ r.v.. Since the r.v.'s $W, W_1, \ldots, W_k$ are independent of S^2, the second equality in (7) holds. Further, by definition, one of the δ_i's is zero, without loss of generality, we may assume $\delta_k = 0$, i.e.,$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{k-1}, 0)$. For a given y and fixed arbitrary constants d_1 and d_2 satisfying $L = d_1 + d_2 > 0$, we wish to find a least favorable configuration (LFC) of means, denoted by $\boldsymbol{\delta}_0$, over the parameter space Ω of all possible μ_i's and σ^2 satisfying

$$\beta_{\boldsymbol{\delta}_0}(d_1, d_2) = \inf_{\Omega} \beta_{\boldsymbol{\delta}}(d_1, d_2).$$

Then, the probability lower bound in (5) can be expressed as

$$\begin{aligned} &2\beta_{\boldsymbol{\delta}_0}(d_1, d_2) - 1 \\ &= 2\inf_{\Omega} \beta_{\boldsymbol{\delta}}(d_1, d_2) - 1. \end{aligned} \tag{8}$$

By applying a Theorem in Chen et al. (1993), the LFC of means in $\beta_{\boldsymbol{\delta}}(d_1, d_2)$ (8) occurs at $\delta_j = 0$ or at $\delta_j = \infty$ for $j = 1, \ldots, k-1$ and hence the $\inf_\Omega \beta_{\boldsymbol{\delta}}(d_1, d_2)$ is obtained by

$$\beta_{\boldsymbol{\delta}_0}(d_1, d_2) = \min_{1\le r\le k} \{f(r; d_2, L)\}, \tag{9}$$

where

$$\begin{aligned} f(r; d_2, L) = \int_0^\infty \int_{-\infty}^\infty &\Big\{\Phi^r((d_1 y + \sqrt{\rho} w)/\sqrt{1-\rho}) \\ &- \Phi^r((-d_2 y + \sqrt{\rho} w)/\sqrt{1-\rho})\Big\} \phi(w) g_\nu(y) dw dy, \end{aligned}$$

where $d_1 = L - d_2$. Following Theorem 2 in Chen and Chen (2004) for a fixed interval width $L = d_1 + d_2 > 0$ and for every $k \ge 2$, there exists a constant $d_2^* = d_2^*(k, L, \nu)$ (and hence $d_1^* = L - d_2^*$) such that

$$\beta_{\boldsymbol{\delta}_0}(d_1, d_2) = \begin{cases} f(1; d_2, L) = \int_0^\infty \int_{-\infty}^\infty \Big\{\Phi((d_1 y + \sqrt{\rho} w)/\sqrt{1-\rho}) \\ - \Phi((-d_2 y + \sqrt{\rho} w)/\sqrt{1-\rho})\Big\} \phi(w) g_\nu(y) dw dy \text{ if } d_2 < d_2^*, \\ \\ f(k; d_2, L) = \int_0^\infty \int_{-\infty}^\infty \Big\{\Phi^k((d_1 y + \sqrt{\rho} w)/\sqrt{1-\rho}) \\ - \Phi^k((-d_2 y + \sqrt{\rho} w)/\sqrt{1-\rho})\Big\} \phi(w) g_\nu(y) dw dy \text{ if } d_2 > d_2^*, \end{cases} \tag{10}$$

and the constant for (10) is $d_2^* = L/2$ for $k = 2$ and $d_2^* < L/2$ for $k \ge 3$. (11)

The optimal choice of the critical values d_1 and d_2 can be determined by $d_2^* = d_2(k, L, \nu)$ (hence $d_1^* = L - d_2^*$) satisfying

$$\Gamma(d_1^*, d_2^*) = 2\sup_{d_2} \beta_{\boldsymbol{\delta}_0}(d_1, d_2) - 1 = P^*, \tag{12}$$

then d_2^* is the choice of d_2 which maximizes (takes supremum over d_2 the infimum of the coverage probability (12) of the confidence region (2) for $\mu_{[k]}$ in I_1 and $\mu_{[1]}$ in I_2 simultaneously. Once the confidence region (2) is so determined by (12), it is usually called a $100P^*\%$ confidence region for $\mu_{[k]}$ and $\mu_{[1]}$. Therefore, if $\beta_{\boldsymbol{\delta}_0}(d_1^*, d_2^*) = 0.95$, then the confidence region (2) for both $\mu_{[k]}$ and $\mu_{[1]}$ simultaneously has a joint confidence $\Gamma(d_1^*, d_2^*)$ of 0.90 $(= 2 \times 0.95 - 1)$ and $P^* = 0.90$ as seen in (12). The results in Eq. (10) and Expression (11) guarantee that a symmetric interval in the confidence region is optimal with $d_2^* = L/2$ for $k = 2$; and an asymmetric interval in the confidence region with $d_2^* < L/2$ is optimal for $k \geq 3$, which means that the individual interval I_1 in (2) should be shifted more to the left of $\bar{X}_{[k]}$ and the individual interval I_2 in (2) should be shifted more to the right of $\bar{X}_{[1]}$. This is because the largest point estimator $\bar{X}_{[k]}$ overestimates $\mu_{[k]}$ and the smallest estimator $\bar{X}_{[1]}$ underestimates $\mu_{[1]}$ for finite sample and the bias increases as k increases as argued by (Dudewicz (1972)). It can be seen that the coverage probability $\beta_{\boldsymbol{\delta}_0}(d_1, d_2)$ reported in (9) and (10), is monotonically increasing in L and is bounded below by zero and above by one. For calculating Expression (10) we used a 64-point quadrature on each of six subintervals for normal integral and two subintervals for chi/root(df) integral and we find by numerical calculation that for any fixed L, $\beta_{\boldsymbol{\delta}_0}(d_1, d_2)$ is increasing first in d_2 and then decreasing after it reaches its maximum at $d_2 = d_2^*$ with $d_2^* = L/2$ for $k = 2$ and $d_2^* < L/2$ for $k > 2$ for any given non-negative correlation coefficient of ρ, which confirms the theoretical result in (11).

For any given k, ρ, P^* and ν, the optimal choice of d_2^* for a $100P^*\%$ confidence region for the largest and the smallest normal means is calculated via the following algorithm:

(i) Choose a pair of values (L_0, d_2) with L_0 being a low initial value of L (say $L_0 = 3$) and a beginning value of d_2 (say $L_0/3$).

(ii) Calculate the integrals in (10) over the grids $(L_0, .001), (L_0, .002), \ldots, (L_0, L_0/2)$. Find the pair (L_0, d_2) among the grids which gives the maximum coverage probability, say P, in (10).

(iii) (a). If $P < (P^* + 1)/2$, then replace L_0 by $L_0 + .001$ and go to step (ii).
(b). If $P > (P^* + 1)/2$ and $P - (P^* + 1)/2 > 10^{-4}$, then replace L_0 by $L_0 - .001$ and go to step (ii).
(c). If $P > (P^* + 1)/2$ and $P - (P^* + 1)/2 < 10^{-4}$, then stop. The solution is found.

The critical values of d_1 and d_2 for $P^* = 0.8, 0.90, 0.95, 0.975, 0.99$; $\rho = 0.0, 0.5$; $k = 3$, 4, 8, 12, 15 and various degrees of freedom, ν are reported in Tables 1 and 2. For example, let $k = 4$, $\rho = 0.5$, $P^* = .90$, and $n = 10 (df = 36)$, then the solution of $d_1 = 2.26$ and $d_2 = 1.87$ can be obtained from Table 2. Therefore, a 90% confidence region for $\mu_{[4]}$ being included in $I_1 = (\bar{X}_{[4]} - 2.26S/\sqrt{10},\ \bar{X}_{[4]} + 1.87S/\sqrt{10})$ and $\mu_{[1]}$ being included in $I_2 = (\bar{X}_{[1]} - 1.87S/\sqrt{10},\ \bar{X}_{[1]} + 2.26S/\sqrt{10})$ can be so constructed. Other desired critical values can be obtained by a Fortran software program (CONF-REGION-2012.FOR) for any combinations of k, ρ, P^* and ν, available from the authors. Note that the theory discussed in Section 2 works for independent populations of a random variable, the variance estimate in (1) becomes the usual pooled variance estimate with correlation coefficient $\rho = 0$ in Eqs (7)–(10).

3. A Two-stage Confidence Region for Correlated Normal Populations

The confidence region proposed in Section 2 is good for an arbitrary sample size and it has a random width in each individual interval in the confidence region. If one wishes to control the width in each individual interval under a specified confidence, one needs a two-stage procedure. In this section a design-oriented two-stage procedure is proposed to find a confidence region with a fixed width on each individual interval under this setting.

Table 1
Critical values of d_1 (left) and d_2 (right) for various k, P^*, and ν at ρ=0.0

pop	*df*	$P^* = 0.8$		$P^* = 0.9$		$P^* = 0.95$		$P^* = 0.975$		$P^* = 0.99$	
k	ν	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
3	3	2.68	2.12	3.58	2.90	4.68	3.82	6.01	4.96	8.29	6.87
3	6	2.18	1.77	2.71	2.26	3.25	2.77	3.82	3.31	4.67	4.07
3	9	2.05	1.67	2.49	2.10	2.93	2.51	3.36	2.94	3.95	3.51
3	15	1.95	1.60	2.34	1.98	2.70	2.34	3.03	2.70	3.48	3.15
3	30	1.88	1.56	2.24	1.90	2.54	2.23	2.82	2.54	3.19	2.92
3	60	1.85	1.53	2.18	1.87	2.48	2.17	2.74	2.46	3.08	2.80
3	210	1.83	1.51	2.15	1.84	2.42	2.14	2.67	2.41	2.98	2.74
4	4	2.63	1.83	3.35	2.44	4.16	3.11	5.10	3.87	6.57	5.05
4	8	2.26	1.61	2.72	2.06	3.18	2.50	3.65	2.95	4.31	3.56
4	12	2.15	1.55	2.54	1.96	2.92	2.35	3.29	2.73	3.81	3.21
4	20	2.06	1.51	2.42	1.88	2.73	2.24	3.05	2.56	3.46	2.98
4	36	2.01	1.48	2.33	1.84	2.62	2.17	2.91	2.46	3.25	2.84
4	60	1.98	1.47	2.29	1.82	2.58	2.13	2.85	2.41	3.16	2.77
4	200	1.97	1.45	2.26	1.79	2.52	2.10	2.77	2.37	3.07	2.70
8	8	2.65	1.51	3.12	1.97	3.61	2.40	4.11	2.85	4.81	3.45
8	16	2.43	1.43	2.78	1.83	3.13	2.19	3.47	2.54	3.90	2.99
8	24	2.36	1.41	2.68	1.79	2.98	2.14	3.29	2.45	3.67	2.85
8	32	2.34	1.39	2.63	1.77	2.92	2.10	3.20	2.41	3.55	2.79
8	40	2.32	1.38	2.61	1.75	2.88	2.09	3.15	2.39	3.48	2.76
8	64	2.27	1.38	2.57	1.74	2.83	2.06	3.08	2.35	3.38	2.71
8	240	2.25	1.36	2.51	1.72	2.76	2.03	2.99	2.31	3.28	2.64
12	12	2.70	1.42	3.09	1.85	3.48	2.24	3.88	2.61	4.39	3.11
12	24	2.54	1.38	2.86	1.76	3.16	2.11	3.44	2.44	3.81	2.84
12	36	2.49	1.36	2.78	1.74	3.06	2.07	3.32	2.38	3.65	2.76
12	48	2.47	1.35	2.74	1.73	3.00	2.06	3.26	2.35	3.57	2.72
12	60	2.44	1.35	2.72	1.72	2.98	2.04	3.22	2.34	3.53	2.69
12	120	2.42	1.34	2.69	1.70	2.93	2.02	3.15	2.31	3.44	2.65
12	240	2.40	1.34	2.67	1.69	2.90	2.01	3.11	2.29	3.43	2.63
15	15	2.73	1.40	3.10	1.80	3.45	2.17	3.80	2.53	4.26	2.98
15	30	2.59	1.36	2.90	1.74	3.18	2.08	3.46	2.39	3.79	2.79
15	45	2.55	1.35	2.83	1.72	3.10	2.05	3.35	2.35	3.66	2.72
15	60	2.53	1.34	2.80	1.71	3.05	2.04	3.30	2.33	3.59	2.69
15	75	2.52	1.34	2.78	1.71	3.03	2.02	3.26	2.32	3.56	2.67
15	150	2.49	1.33	2.75	1.69	2.99	2.01	3.21	2.29	3.48	2.63
15	300	2.47	1.33	2.74	1.68	2.96	2.00	3.18	2.28	3.44	2.62

For this purpose, a generalized Stein-type (Stein, 1945) two-stage sampling procedure P_2 similar to Chen and Dudewicz (1976) is developed as stated in the following steps:

P_2: Take an independent initial vector sample $(X_{1l}, ..., X_{kl})$ of size $n_0(\geq 2)$ from a k-variate population $N_k(\boldsymbol{\mu}, \sigma^2, \rho), l = 1, ..., n_0$. Let $(\bar{X}_1^0, \ldots, \bar{X}_k^0)$ be the sample mean vector and $Cov^0 = \{s_{ij}^0, i, j = 1, \ldots, k\}$ the sample covariance matrix based on the initial sample

Table 2
Critical values of d_1 (left) and d_2 (right) for various k, P^*, and ν at $\rho = 0.5$

popu	*df*	$P^* = 0.8$		$P^* = 0.9$		$P^* = 0.95$		$P^* = 0.975$		$P^* = 0.99$	
k	ν	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
3	3	2.57	2.18	3.44	2.98	4.52	3.91	5.81	5.07	8.01	7.02
3	6	2.11	1.81	2.64	2.30	3.18	2.81	3.75	3.35	4.58	4.12
3	9	2.00	1.70	2.42	2.14	2.86	2.55	3.30	2.97	3.88	3.55
3	15	1.91	1.63	2.28	2.02	2.66	2.36	2.99	2.72	3.46	3.16
3	30	1.83	1.59	2.20	1.92	2.50	2.25	2.80	2.55	3.18	2.92
3	60	1.82	1.55	2.14	1.89	2.44	2.19	2.72	2.47	3.04	2.82
3	210	1.79	1.54	2.10	1.87	2.39	2.16	2.64	2.43	2.97	2.74
4	4	2.46	1.90	3.15	2.52	3.94	3.20	4.84	3.97	6.24	5.18
4	8	2.13	1.67	2.60	2.11	3.06	2.55	3.53	3.00	4.20	3.60
4	12	2.04	1.60	2.44	2.00	2.84	2.38	3.22	2.75	3.73	3.24
4	20	1.97	1.55	2.33	1.92	2.67	2.26	3.00	2.58	3.41	2.99
4	36	1.93	1.52	2.26	1.87	2.57	2.19	2.86	2.48	3.23	2.84
4	60	1.90	1.51	2.22	1.85	2.53	2.15	2.80	2.43	3.13	2.78
4	200	1.88	1.49	2.20	1.82	2.47	2.12	2.73	2.39	3.05	2.71
8	8	2.41	1.56	2.88	2.01	3.37	2.45	3.88	2.89	4.57	3.50
8	16	2.23	1.48	2.62	1.86	2.98	2.23	3.33	2.57	3.79	3.01
8	24	2.18	1.45	2.54	1.82	2.87	2.16	3.18	2.47	3.58	2.87
8	32	2.15	1.44	2.51	1.79	2.81	2.13	3.11	2.43	3.48	2.81
8	40	2.14	1.43	2.49	1.78	2.78	2.11	3.07	2.40	3.41	2.77
8	64	2.12	1.42	2.45	1.76	2.74	2.08	3.00	2.37	3.33	2.72
8	240	2.09	1.41	2.41	1.74	2.68	2.05	2.93	2.32	3.23	2.65
12	12	2.43	1.47	2.85	1.88	3.27	2.26	3.67	2.64	4.20	3.13
12	24	2.32	1.41	2.67	1.79	3.00	2.13	3.32	2.45	3.71	2.85
12	36	2.28	1.40	2.62	1.76	2.92	2.09	3.22	2.39	3.56	2.77
12	48	2.27	1.39	2.59	1.75	2.88	2.07	3.17	2.36	3.50	2.73
12	60	2.26	1.38	2.58	1.74	2.86	2.06	3.13	2.35	3.46	2.70
12	120	2.24	1.37	2.55	1.72	2.81	2.04	3.07	2.32	3.38	2.66
12	240	2.21	1.37	2.53	1.72	2.80	2.03	3.04	2.30	3.35	2.63
15	15	2.45	1.44	2.86	1.83	3.23	2.20	3.60	2.55	4.08	3.01
15	30	2.36	1.39	2.71	1.76	3.02	2.10	3.32	2.41	3.69	2.79
15	45	2.33	1.38	2.66	1.74	2.96	2.06	3.23	2.37	3.57	2.73
15	60	2.32	1.37	2.64	1.73	2.92	2.05	3.20	2.34	3.53	2.69
15	75	2.31	1.37	2.63	1.72	2.91	2.04	3.17	2.33	3.49	2.68
15	150	2.30	1.36	2.59	1.72	2.87	2.02	3.13	2.30	3.42	2.64
15	300	2.29	1.35	2.58	1.71	2.86	2.01	3.09	2.30	3.39	2.62

vectors, where $\bar{X}_i^0 = \sum_{l=1}^{n_0} X_{il}/n_0$ and $s_{ij}^0 = \sum_{l=1}^{n_0}(X_{il} - \bar{X}_i^0)(X_{jl} - \bar{X}_j^0)/(n_0 - 1)$. From Johnson and Kotz (1972), the unbiased estimate of σ^2 is given by

$$S_0^2 = \frac{\{1 + (k-2)\rho\}\sum_{i=1}^{k} s_{ii}^0 - 2\rho \sum_{i>j} s_{ij}^0}{k(1-\rho)\{1 + (k-1)\rho\}}. \tag{13}$$

The total sample size drawn from the k-variate normal population is defined as:

$$n = \max\{n_0 + 1, \left[\frac{S_0^2}{c^2}\right] + 1\}, \tag{14}$$

where the value of c in (14) is an arbitrary positive constant (c is called a design constant to be discussed later) which is to be chosen in order to meet some probability requirement to control the interval width on the largest mean $\mu_{[k]}$ (and the smallest mean $\mu_{[1]}$), and $[x]$ denotes the largest integer smaller than or equal to x.

Then, take $n - n_0$ additional vector observations $(X_{1l}, ..., X_{kl})$, $l = n_0 + 1, ..., n$, so we have a total of n observations for the ith component denoted by $X_{i1}, ..., X_{in_0}, ..., X_{in}$. For the ith component of the k-variate population, choose the coefficients $a_{i1}, ..., a_{in_0}, a_{i,n_0+1}, ..., a_{in}$ to satisfy the following three conditions:

$$(i)\ \sum_{j=1}^{n} a_{ij} = 1, (ii)\ a_{i1} = \ldots = a_{in_0}, (iii)\ S_0^2 \sum_{j=1}^{n} a_{ij}^2 = c^2.$$

Solving the equation (iii) subject to conditions (i) to (ii), we obtained the coefficients

$$a_{i1} = ... = a_{in_0} = \frac{1 - (n - n_0)b}{n_0} = a,$$

where b is solved by (i) to (iii) as given by

$$b = a_{i,n_0+1} = ... = a_{in} = \frac{1}{n}\{1 + \sqrt{1 - \frac{n}{(n - n_0)}\left(1 - \frac{n_0 c^2}{S_0^2}\right)}\}.$$

Finally we can compute the weighted sample mean for the sample from the ith component by

$$\tilde{X}_i = a\sum_{l=1}^{n_0} X_{il} + b \sum_{l=n_0+1}^{n} X_{il} \tag{15}$$

which is actually a linear combination of the two sets of samples (the initial and the second samples) with random coefficients a and b. Such choice has the property that if the total sample size n is close to S_0^2/c^2, then a and b are close to $1/n$, thus, $\tilde{X}_i$ converges to the unbiased sample mean. Define the r.v. $T_i = (\tilde{X}_i - \mu_i)/c$, then it has a marginal Student t distribution with $\nu_0 = k(n_0 - 1)$ d.f., $i = 1, \ldots, k$ (Chen and Wen, 2006). By the property of Student's t, it is also a conditional normal with a mean of 0 and a variance of σ^2/S_0^2 conditioning on S_0^2. We are going to show that the correlation coefficient between T_i and T_j is the same as the population one, ρ, $i \neq j$, $i = 1, .., k$, $j = 1, ..., k$. Let σ_{ij} be the covariance of X_{il} and X_{jl} for any $l = 1, 2, ..., n$, $i = 1, .., k$, $j = 1, ..., k$. Then the covariance between T_i and T_j in (13) is given by

$$\begin{aligned} Cov(T_i, T_j) &= E\left(T_i T_j\right) - E(T_i)E(T_j) = EE\left(T_i T_j | S_0\right) - EE\left(T_i | S_0\right) EE\left(T_j | S_0\right) \\ &= EE\left(T_i T_j | S_0\right) \\ &= EE\left(\left(a\sum_{l=1}^{n_0}(X_{il} - \mu_i)/c + b\sum_{l=n_0+1}^{n}(X_{il} - \mu_i)/c\right)\right. \end{aligned}$$

$$\times \left(a \sum_{l=1}^{n_0} (X_{jl} - \mu_j)/c + b \sum_{l=n_0+1}^{n} (X_{jl} - \mu_j)/c \right) |S_0 \Bigg)$$

$$= E\left(n_0 a^2 + (n - n_0) b^2\right) \sigma_{ij}/c^2$$

$$= E\left(n_0 a^2 + \frac{(1 - n_0 a)^2}{(n - n_0)}\right) \sigma_{ij}/c^2, \tag{16}$$

the second equality holds as $EE(T_i|S_0) = 0$, $i = 1, ..., k$. It is also clear that the conditional variance of T_i is σ^2/S_0^2 given S_0^2, where $\sigma^2 = \sigma_{ii}$ is the variance of X_{il}, $l = 1, ..., n$. Replacing T_j by T_i in the covariance structure (16), we obtain the variance of T_i as $Var(T_i) = E(n_0 a^2 + \frac{(1-n_0 a)^2}{(n-n_0)})\sigma^2/c^2$, $i = 1, ..., k$. Furthermore, the correlation coefficient between T_i and T_j can be calculated by $Cov(T_i, T_j)/Var(T_i) = \sigma_{ij}/\sigma^2 = \rho$ which is the original population correlation coefficient as assumed in the beginning of Section 2. It is also clear that the conditional distribution of $T_1, ..., T_k$ has a conditional k-variate normal distribution with means of zero, variance of σ^2/S_0^2 and a correlation coefficient of ρ, which is a k-variate Student t distribution by definition. Therefore, we can claim that the r.v.'s $T_i = (\tilde{X}_i - \mu_i)/c$, $i = 1, \ldots, k$, have a k-variate central Student t distribution with $\nu_0 = k(n_0 - 1)$ d.f. and a common correlation coefficient of ρ, denoted by $T_k(0, \nu_0, \rho)$. The condition (i) is to ensure the unbiasedness of $\tilde{X}_i$ for μ_i, the condition (ii) guarantees that the sample mean $\bar{X}_{i0}$ and pooled sample variance S_0^2 in (13) based on the first stage n_0 observations are independent, and the condition (iii) is the variance estimate of $\tilde{X}_i$ controlled at a fixed width-related value c which makes the choices of a_{ij} possible and guarantees that the r.v.'s $\{T_1, ..., T_k\}$ have a k-variate central t distribution as $T_k(0, \nu_0, \rho)$.

Let $\tilde{X}_{[k]}$ denote the largest value of the sample means $\tilde{X}_i$'s and $\mu_{[k]}$ the largest unknown population mean among μ_i's. It is intuitive that the $\tilde{X}_{[k]}$ ($\tilde{X}_{[1]}$) is a natural estimator of $\mu_{[k]}$ ($\mu_{[1]}$) and furthermore, $\tilde{X}_{[k]}$ ($\tilde{X}_{[1]}$) is strongly consistent and asymptotically unbiased for $\mu_{[k]}$ ($\mu_{[1]}$) by a argument due to Chen (1975). Since $\tilde{X}_{[k]}$ overestimates $\mu_{[k]}$ ($\tilde{X}_{[1]}$ underestimates $\mu_{[1]}$) for finite samples as argued by Dudewicz (1972), an asymmetric interval for the largest mean $\mu_{[k]}$ (smallest $\mu_{[1]}$) by allocating more of the interval to the left of $\tilde{X}_{[k]}$ (to the right of $\tilde{X}_{[1]}$) should be a better allocation. For a prespecified number $P^*(1/k^2 < P^* < 1)$, consider the confidence region for the largest mean $\mu_{[k]}$ and the smallest mean $\mu_{[1]}$ simultaneously by

$$I^* = \{\mu_{[k]} \in I_1^* \text{and} \mu_{[1]} \in I_2^*\}, \tag{17}$$

where $I_1^* = (\tilde{X}_{[k]} - \tilde{d}_1 c, \tilde{X}_{[k]} + \tilde{d}_2 c)$ and $I_2^* = (\tilde{X}_{[1]} - \tilde{d}_2 c, \tilde{X}_{[1]} + \tilde{d}_1 c)$, each with a fixed width W, $W = (\tilde{d}_1 + \tilde{d}_2)c$, where $\tilde{d}_1$ and $\tilde{d}_2$ are constants such that $\tilde{d}_1 > \tilde{d}_2$ and $\tilde{d}_1 + \tilde{d}_2 > 0$, and I_1^* (I_2^*) represents the first (second) interval for $\mu_{[k]}$ ($\mu_{[1]}$). The reason we use the same $\tilde{d}_1$ and $\tilde{d}_2$ in both individual intervals is that the individual intervals in (17) are symmetric.

From Section 2, it can be observed that the random vector $\{Z_1/Y, ..., Z_k/Y\}$ has a k-variate multivariate t distribution as $T_k(0, \nu, \rho)$, where $Z_i = (\bar{X}_i - \mu_i)/\sigma$ and $Y = S/\sigma$, $i = 1, ..., k$. Therefore, the optimal critical values of $\tilde{d}_1$ and $\tilde{d}_2$ are equal to those of d_1 and d_2 by replacing the df, ν by ν_0 which are determined by the algorithm as given in Section 2.

For example, let $k = 8$, $n_0 = 9$, $\nu_0 = k(n_0 - 1) = 64$, ρ=0.5, $P^* = 0.90$, and the width is specified to be $W = 2$, using Table 2, the critical values can be found as $\tilde{d}_1 = 2.45$ and $\tilde{d}_2 = 1.76$. Then the design constant $c = W/(\tilde{d}_1 + \tilde{d}_2) = 2/(2.45 + 1.76) = 0.4751$. Substituting the values of n_0, c, and S_0^2 into (14), one can determine the required total sample size n to be drawn from the k-variate normal population. For independent normal populations (with $\rho = 0$), one can find the critical values from Table 1.

4. Comparisons to Other Confidence Regions

In this section we first propose a class of confidence regions for the largest and the smallest means under the setting of a k-variate normal distribution as stated in Section 2. Then, Bonferroni inequality is employed to generate the lower bound of the coverage probability for the confidence region in the class. After a best confidence region is determined, it is compared to the optimal one derived in Section 2. It has been found by numerical comparison that the optimal one proposed in Section 2 is superior to the class of confidence regions formulated by intercepting a lower and an upper confidence interval similar to that of Chen and Dudewicz (1976). Let's consider a class of confidence regions J for $\mu_{[k]}$ and $\mu_{[1]}$ as given by

$$J = (\bar{X}_{[k]} - c_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + c_2 S/\sqrt{n}, \bar{X}_{[1]} - c_2 S/\sqrt{n} < \mu_{[1]} < \bar{X}_{[1]} + c_1 S/\sqrt{n}), \quad (18)$$

where $c_1 + c_2 = L$ and L is fixed such that the confidence probability for (18) satisfies a prespecified value. Using the Bonferroni inequality and a similar argument as in (3)-(4) we have

$$P(J) > 2P(\bar{X}_{[k]} - c_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + c_2 S/\sqrt{n}) - 1. \quad (19)$$

The probability statement of the interval for $\mu_{[k]}$

$$P(\bar{X}_{[k]} - c_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + c_2 S/\sqrt{n})$$

is obtained by intercepting a lower interval $LI = (\bar{X}_{[k]} - c_1 S/\sqrt{n}, \infty)$ and an upper interval $UI = (-\infty, \bar{X}_{[k]} + c_2 S/\sqrt{n})$ for the largest mean $\mu_{[k]}$. Set the right hand side of (19) to P^* to obtain

$$P(\bar{X}_{[k]} - c_1 S/\sqrt{n} < \mu_{[k]} < \bar{X}_{[k]} + c_2 S/\sqrt{n}) = (P^* + 1)/2. \quad (20)$$

By applying Bonferroni inequality on the left side of Equation (20) we have the lower bound for the probability coverage of the lower interval LI for $\mu_{[k]}$,

$$F_{k,\nu}(c_1, \cdots, c_1) = \int_0^\infty \int_{-\infty}^\infty \Phi^k((c_1 y + \sqrt{\rho} w)/\sqrt{1-\rho})\phi(w) g_\nu(y) dw dy, \quad (21)$$

and the lower bound for the probability coverage of the upper interval UI for $\mu_{[k]}$,

$$F_\nu(c_2) = \int_0^\infty \Phi(c_2 y) g_\nu(y) dy. \quad (22)$$

(It should be noted that the lower bound of the probability coverage for the lower interval LI for $\mu_{[k]}$ is a multivariate t c.d.f. while the upper interval is a univariate Student's t c.d.f..) There are many possible choices of c_1 and c_2 that can satisfy the requirement of Equation (20). Set Equation (21) to γ and Equation (22) to $(3 + P^*)/2 - \gamma$, the best choice can be made by choosing the critical values as $c_1 = F_{k,\nu}^{-1}(\gamma_1)$ and $c_2 = F_\nu^{-1}((3 + P^*)/2 - \gamma_1)$ such that the γ_1 minimizes the interval width function, $c_1 + c_2$, or equivalently, minimizing the function

$$h(\gamma) = F_{k,\nu}^{-1}(\gamma) + F_{\nu}^{-1}((3 + P^*)/2 - \gamma) \tag{23}$$

for $\gamma \in ((P^* + 1)/2, 1)$, where $F_{k,\nu}^{-1}(\gamma)$ is the inverse of the c.d.f. of a k-variate t distribution $F_{k,\nu}(\cdot)$ with ν df and a nonnegative correlation coefficient of ρ, and $F_{\nu}^{-1}(\cdot)$ is the inverse of the c.d.f. of a Student t distribution $F_{\nu}(\cdot)$ with $\nu = k(n-1)$ d.f.. For given $L = c_1 + c_2 > 0$, the function in (23) is decreasing first and then increasing after it reaches a minimum point at certain value of c_2, so a unique solution exists. Define the interval width reduction (IWR) by the proposed optimal confidence region (2) over the intercepting intervals (18) similar to that of Chen and Chen (2004) as $IWR = ((c_1 + c_2) - (d_1 + d_2))/(c_1 + c_2)$. The improvement of IWR is calculated for $\rho = 0.5$ which is reported in Table 3 for joint confidence $P^* = 0.80, 0.90, 0.95, 0.975$ and 0.99; $k = 3, 4, 8, 12, 15$; and various degrees of freedom, ν.

A large and positive ratio indicates that the amount of improvement in IWR in each component of the confidence region is significant. We can see from Table 3 that the IWR of the optimal confidence region (2) over the confidence region J in (18) is between 4.9% at $k = 3$, $\nu = 210$, $P^* = 0.99$ and 15.5% at $k = 3$, $\nu = 3$, $P^* = 0.80$. For given k and ν, the IWR decreases as P^* increases; for example, when k=3 and $\nu = 3$, the IWR ranges from 15.5% at $P^* = 0.80$ to 12.7% at $P^* = 0.99$. Secondly, the IWR increases as k increases for given P^* and ν; for example, when $P^* = 0.8$ and $\nu = 60$, the IWR ranges from 10.7% at $k = 3$ increased to 14.2% at $k = 15$. Finally, the IWR decreases as ν increases for given k and P^*; for example, for $k = 3$ and $P^* = 0.95$, the IWR ranges from 13.4% at $\nu = 3$ reduced to 6.9% at $\nu = 210$. A general pattern of the IWR is similar to Table 3 for any nonnegative correlation coefficient and any combinations of k, P^* and ν. Therefore, the proposed optimal confidence region (2) is uniformly better than the intercepting one (18) for all values of $\rho(\geq 0)$, k, P^*, and ν calculated in Table 3.

5. An Example

To illustrate the confidence region in (2) for independent normal distributions (where ρ is zero), we employed the experimental results of four independent groups of physical therapy patients by different treatments, each produced six independent scores by six patients (Daniel, 1974, p. 195.). The scores measuring treatment effectiveness are given in Table 4 and a summary of statistics based on six observations for each treatment is reported in Table 5. The assumption of normality of the data for each of four independent groups was checked by Shapiro-Wilk test using SAS program and they all yielded a high p-value which supported this assumption. In addition, the modified Levene test (BF option in SAS) yielded an F value of 1.43 with a p-value of 0.2645 which leads to acceptance of the hypothesis of homogeneity of population variances.

Moreover, the traditional ANOVA test with an F value of 6.03 and a p-value of 0.0043 indicated a significant difference among four mean treatment effectiveness. We now can apply the confidence region in (2) for the largest and the smallest treatment effectiveness out of four $k = 4$ independent treatments simultaneously with $df = 20$ and $P^* = 0.90$. The largest sample mean $\bar{X}_{[4]} = 87.5$, the smallest sample mean $\bar{X}_{[1]} = 69.16667$, the pooled sample standard deviation $S_p = 7.6043$ and the critical value of d_1=2.42 and d_2=1.88 were obtained (where d1 and d2 are selected from Table 1), and then, a 90% confidence

Table 3
Interval width reduction (IWR) by the optimal confidence region over the region (18) for $\rho = 0.5$ and various P^*

k	ν	$P^* = 0.80$	$P^* = 0.90$	$P^* = 0.95$	$P^* = 0.975$	$P^* = 0.99$
3	3	0.155	0.142	0.134	0.130	0.127
3	6	0.129	0.110	0.099	0.092	0.086
3	9	0.120	0.101	0.089	0.080	0.073
3	15	0.113	0.093	0.080	0.072	0.062
3	30	0.110	0.088	0.075	0.065	0.055
3	60	0.107	0.086	0.072	0.061	0.053
3	210	0.106	0.084	0.069	0.060	0.049
4	4	0.162	0.144	0.133	0.126	0.121
4	8	0.139	0.116	0.103	0.094	0.086
4	12	0.132	0.110	0.093	0.084	0.074
4	20	0.126	0.103	0.087	0.076	0.066
4	36	0.121	0.099	0.082	0.072	0.060
4	60	0.120	0.097	0.079	0.069	0.058
4	200	0.118	0.093	0.078	0.066	0.053
8	8	0.160	0.137	0.130	0.108	0.098
8	16	0.148	0.121	0.102	0.090	0.078
8	24	0.144	0.115	0.096	0.085	0.072
8	32	0.142	0.112	0.094	0.080	0.068
8	40	0.140	0.109	0.092	0.079	0.067
8	64	0.137	0.109	0.089	0.076	0.064
8	240	0.135	0.105	0.086	0.074	0.062
12	12	0.157	0.130	0.111	0.098	0.088
12	24	0.147	0.119	0.100	0.086	0.074
12	36	0.143	0.114	0.096	0.081	0.070
12	48	0.140	0.112	0.094	0.079	0.068
12	60	0.140	0.110	0.092	0.078	0.066
12	120	0.138	0.108	0.090	0.077	0.064
12	240	0.138	0.106	0.086	0.075	0.061
15	15	0.155	0.125	0.108	0.099	0.082
15	30	0.147	0.117	0.098	0.084	0.072
15	45	0.144	0.114	0.095	0.080	0.069
15	60	0.142	0.112	0.094	0.078	0.066
15	75	0.140	0.111	0.091	0.077	0.064
15	150	0.139	0.109	0.090	0.076	0.063
15	300	0.138	0.108	0.088	0.074	0.062

region for the largest mean treatment effectiveness ($\mu_{[4]}$) and the smallest mean treatment effectiveness ($\mu_{[1]}$) is calculated as:

$$I_1 = (87.5000 - 2.42 \times 7.6043/\sqrt{6}, 87.5000 + 1.88 \times 7.6043/\sqrt{6})$$
$$= (87.5000 - 7.5128, 87.5000 + 5.8364) = (79.9872, 93.3364)$$

Table 4
Scores for physical therapy patients with four treatments

Obs	Treatments 1	2	3	4
1	64	76	58	95
2	88	70	74	90
3	72	90	66	80
4	80	80	60	87
5	79	75	82	88
6	71	82	75	85

and

$$I_2 = (69.16667 - 1.88 \times 7.6043/\sqrt{6}, 69.16667 + 2.42 \times 7.6043/\sqrt{6})$$
$$= (69.16667 - 5.8364, 69.16667 + 7.5128) = (63.3303, 76.6795).$$

The interpretation of the confidence region is explained as follows: By ranking these sample means, the fourth treatment is identified to be the most effective physical therapy and the third treatment is the least effective physical therapy. To be more informative and accurate, with a 90% confidence, the largest mean treatment effectiveness score falls in the individual interval ranging from 79.9872 to 93.3364 in the confidence region and the smallest mean treatment effectiveness score falls in the individual interval ranging from

Table 5
Summary of statistics

	Treatments 1	2	3	4
Sample size	6	6	6	6
Sample mean	75.6667	78.8333	69.1667	87.5000
Standard dev	8.4063	6.8823	9.3897	5.0100
p-value by Shapiro-Wilk normality test	0.9332	0.9097	0.653	0.9843

Error mean square = 57.825 , S_p =7.6043
Modified Levene's test for equal variances
F = 1.43 with p-value = 0.2645
Smallest, largest sample mean = 69.1667, 87.5
$k = 4$, $df = v = 20$, $P^* = 0.90$
Critical values: d_1 = 2.42, $d_2 = 1.88$

63.3303 to 76.6795. It is clear that the two individual intervals do not have any overlap, which indicates that the best and the worst treatments are significantly separated apart.

6. Summary and Conclusion

In ranking and selection procedures (see e.g. Bechhofer, 1954), the goal is often to select the best population among several ones, where the best population is defined to be the one having the largest mean. Sometimes, the experimenters want to select the best population and the worst population at the same time and to tell how good the best one and how bad the worse one are. That is the reason why we propose the confidence region in (2) for the largest and the smallest normal means by maximizing the coverage probability of the confidence region, or equivalently by minimizing the expected width of the individual interval associated with an optimal allocation of the critical values at a given confidence probability.

It is common to use the largest sample mean as a point estimate of the largest population mean since the largest sample mean is both asymptotically unbiased and strongly consistent (see e.g. Chen, 1975). Similarly, the smallest sample mean is used as the point estimate of the smallest population mean and it is also asymptotically unbiased and strongly consistent. However, the largest sample mean overestimates the largest population mean as the number of populations increases (Dudewicz, 1972). Likewise, the smallest sample mean underestimates the smallest population mean as the number of populations increases. Hence, it is necessary to make an adjustment for the allocation of each individual intervals in the confidence region centered about the largest sample mean and the smallest sample mean at the same time. It has been found that when there are only two populations, the individual interval is symmetric about the largest (smallest) sample mean, and when there are three or more populations, the optimal individual interval in the confidence region becomes asymmetric by shifting more of its interval to the left of the largest sample mean and shifting more of its interval to the right of the smallest sample mean. Such a confidence region so constructed is thought to be optimal in the sense of a smallest interval width in each component among a class of confidence regions obtained by Bonferroni inequality.

Our numerical calculations have shown that the proposed optimal confidence region for the correlated normal populations is superior to currently existing one-sample procedure for any sample size, which yields a confidence region for the largest normal mean and the smallest normal mean simultaneously.

We also proposed a two-stage optimal confidence region in (17) for correlated normal populations to control the width of each individual interval in the confidence region. If a fixed confidence width is required or preassigned, a design-oriented two-stage confidence region should be employed.

Selected tables of critical values at various confidence probabilities P^*, nonnegative correlation coefficient, the number of populations and degrees of freedom can be found in Chen et al. (2008) at the level of $(P^* + 1)/2$. Extended tables of the critical values for a large number of populations are only available through the authors using a Fortran code. It should be noted that the optimal confidence region in (2) is based on equal sample size for all samples. In situations where the sample sizes are not all equal, it is suggested that the number $1/n$ in the region (2) be replaced by the average of $1/n_i$, or equivalently, by the harmonic mean of individual sample sizes. At last, we give one example to illustrate the confidence region for independent normal populations.

In conclusion, if practitioners wish to find a single-sample confidence region for the largest and smallest normal means simultaneously, the optimal confidence region (2) is recommended to use. Further, if an interval width is desired to be controlled, a design-oriented two-stage confidence region (17) should be considered.

Funding

Hubert Chen's research was supported by The National Science Council 100-2118-M-006-008 and NSC 101-2118-M-006-004 in Taiwan, ROC. Shu-Fei Wu's research was supported by NSC 99-2118-M-032-001-MY2, NSC 101-2118-M-032-003-, NSC 102-2118-M-032-006-, MOST 103-2118-M-032 -004 -MY2, 2010-2015 in Taiwan, ROC.

References

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics* 25: 16–39.

Chen, H. J. (1975). Strong consistency and asymptotic unbiasedness of a natural estimator for a ranked parameter. *Sankhya* Ser. B 38: 92–94.

Chen, H. J., Chen, S. Y. (1999). A nearly optimal confidence interval for the largest normal mean. *Communications in Statistics: Simulation and Computation* 28(1): 131–146.

Chen, H. J., Chen, S. Y. (2004). Optimal confidence interval for the largest normal mean with unknown variance. *Computational Statistics and Data Analysis* 47: 845–866.

Chen, H. J., Dudewicz, E. J. (1976). Procedures for fixed-width interval estimation of the largest normal mean. *Journal of the American Statistical Association* 71: 752–756.

Chen, H. J., Dudewicz, E. J. (1973). Estimation of ordered parameters and subset selection. Technical Report, Columbus, Ohio: Ohio State University.

Chen, H. J., Li, H. L., Wen, M. J. (2008). Optimal Confidence Interval for the Largest Mean of Correlated Normal Populations and Its Application to Stock Fund Evaluation. *Computational Statistics and Data Analysis* 52: 4801–4813.

Chen, H. J., Wen, M. J. (2006). Optimal confidence interval for the largest normal mean under heteroscedasticity. *Computational Statistics and Data Analysis* 51: 982–1001.

Chen, H. J., Wu, S. F. (2011). A confidence region for the largest and the smallest means under heteroscedasticity. *Computational Statistics and Data Analysis*. Online version.

Daniel, W. W. (1974). *Biostatistics: A Foundation for Analysis in the Health Sciences*. New York: Wiley.

Dudewicz, E. J. (1972). Two-sided confidence intervals for ranked means. *Journal of the American Statistical Association* 67: 462–464.

Dudewicz, E. J., Tong, Y. L. (1971). Optimal confidence interval for the largest location parameter. In: Gupta, S. S., Yackel, J., eds. *Statistical Decision Theory and Related Topics*. New York: Academic Press, Inc., pp. 363–375.

Johnson, N. L., Kotz, S. (1972). *Distribution in statistics: Continuous multivariate distributions*. New York: Wiley.

Johnson, R. A., Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis,* 4th ed. Upper Saddle River, New Jersey: Prentice Hall.

Saxena, K. M. L. (1976). A single-sample procedure for the estimation of the largest mean. *Journal of the American Statistical Association* 71: 147–148.

Saxena, K. M. L., Tong, Y. L. (1969). Interval estimation of the largest mean of k normal populations with known variance. *Journal of the American Statistical Association* 64: 296–299.

Stein, C. M. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics* 16: 243–258.

Tong, Y. L. (1973). An asymptotically optimal sequential procedure for the estimation of the largest mean. *The Annals of Statistics* 1: 175–179.