Journal of Mathematical Extension
Vol. 2, N0. 1, 2 (2007-2008), 45–56

# Generalized Confidence Interval for the Largest Normal Mean

**J. Behboodian**
Islamic Azad University - Shiraz Branch

**A. A. Jafari**
Shiraz University

**Abstract.** In this article we consider inferences on the largest mean of $k( \geqslant 2)$ normal populations with unequal unknown variances. The coverage probability and expected length of the two confidence intervals for the largest mean are compared using Monte Carlo simulation. We found that the coverage probability of the generalized confidence interval and optimal confidence interval are close to nominal level but the expected length of generalized confidence interval is smaller than the optimal confidence interval. Finally an example is provided.

**AMS Subject Classification:** Primary 62F03; Secondary 65U05.
**Keywords and Phrases:** Largest mean, Generalized confidence interval, Generalized p-value, Optimal confidence interval.

## 1. Introduction

Let $\pi_1, \pi_2, ..., \pi_k$ be $k(\geqslant 2)$ normal populations with means $\mu_i$ and variances $\sigma_i^2$, and $n_i$ observations taken from each population, $i = 1, 2, ..., k$. The ordered values of $\mu_i$'s, denoted by $\mu_{[1]} \leqslant \mu_{[2]} \leqslant ... \leqslant \mu_{[k]}$, provide some important and useful information about the populations. For example, when the variances are equal, the ordering of $\mu_i$'s is equivalent to ordering of the populations. In reliability and lifetesting, ranked parameters are applied to study a system.

The problem of ranking of means of several independent normal populations, specially the largest and smallest mean, has been studied by many authors; Dudewicz (1972) showed that the largest sample mean, as a natural estimator of the largest mean, is biased. However, Chen (1976) noted that the largest sample mean is both strongly consistent and asymptotically unbiased. For the largest mean, when the common population variance is known, Saxena and Tong (1969) and Dudewicz (1972) analyzed confidence intervals that were not optimal, but Dudewicz and Tong (1971) proposed an optimal confidence interval, and then Tong (1973) provided percentage points for this kind of interval. When the variances are equal but unknown, Saxena (1976) gave a large sample approximation, Chen and Chen (1999) obtained a nearly optimal, and Chen and Chen (2004) found an optimal confidence interval. When the variances are unknown and possibly unequal, Chen and Dudewicz (1976), and Chen (1977a,b) proposed a class of intersecting intervals, and then Chen and Wen (2006) found an optimal confidence interval.

For the ranked mean, $\mu_{[i]}$ there are two articles, by Dudewicz (1972), who obtained a class of two-sided intervals when variances are known, and by Alam and Saxena (1974), who gave a confidence interval such that the distribution of each population is stochastically increasing in mean.

In this paper, we first review two methods that are applicable for interval estimation ranked means, $\mu_{[i]}$ in Section 2. First method is based on the concepts of generalized confidence interval and generalized p-value, and is given by Chang and Huang (2000). The concepts of generalized p-value and generalized confidence interval that provide exact tests and exact confidence intervals, proposed by Tsui and Weerahandi (1989) and Weerahandi (1993). Second method is optimal confidence interval that is given by Chen and Wen (2006). In Section 3, by a numerical example, we compare these methods with each other. Simulation studies are presented in Section 4, to compare the coverage probabilities and the expected lengths of these methods.

## 2. Inferences for the Mean of the Best Population

Suppose $X_{i1}, X_{i2}, ..., X_{in_i}$, $i = 1, ..., k$ are $k$ random samples with size $n_i$ from normal populations with means $\mu_i$ and unequal variances $\sigma_i^2$. The largest normal mean, $\mu^* = \mu_{[k]} = \max(\mu_1, ..., \mu_k)$ provides some important and useful information about populations. The population that has the largest mean, $\mu^*$ is known as the best population.

The problem of interest is to find confidence interval for the largest mean. In this Section we first study the confidence interval of $\mu^*$ based on a generalized pivotal variable that is given by Chang and Huang (2000). Then briefly we review the method of Chen and Chen (2006) that is an optimal confidence interval for the largest mean.

### 2.1 Generalized Confidence Interval for the Largest Mean

Let, for the $i$th population, $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $S_i^2 = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 /(n_i - 1)$ be the sample mean and sample variance, respectively, and $\bar{x}_i$ and $s_i^2$ denote the observed values of $\bar{X}_i$ and $S_i^2$, $i = 1, ..., k$.

The generalized pivotal variable for $\mu_i$ is given by

$$T_i = \bar{x}_i - \frac{t_{n_i-1} s_i}{\sqrt{n_i}}, \quad i = 1, ..., k, \tag{1}$$

where $t_{n_i-1} = \sqrt{n_i}\frac{\bar{X}_i - \mu_i}{S_i}$ follows a student's t-distribution with $n_i - 1$ degrees of freedom.

The generalized pivotal variable for the largest mean is the largest generalized pivotal variables, $T_i$'s of $\mu_i$'s, i.e.

$$T^* = \max(T_1, T_2, ..., T_k). \tag{2}$$

The exact $(1 - \alpha)$ confidence interval is given by

$$[T^*(\alpha/2), T^*(1 - \alpha/2)], \tag{3}$$

where $T^*(\gamma)$ stands for the $\gamma$th quantile of $T^*$.

## 2.2 Optimal Confidence Interval

Chen and Wen (2006) proposed a two-stage procedure for obtaining an optimal confidence interval for the largest or smallest mean of $k$ independent normal population, where the population variances are unknown and possibly unequal. This procedure requires additional samples and it may not be practicable in real problems. Therefore they employed a one-stage sampling procedure that is proposed by Chen and Lam (1989) for interval estimation. This one-stage sampling procedure for interval estimation is stated as follows:

Let $X_{ij}$ $(j = 1, ..., n_i)$ be an independent random sample from normal population $\pi_i$, $(i = 1, ..., k)$ with unknown and unequal variance $\sigma_i^2$, $i = 1, ..., k$.

Choose $n_i^{'}$ $(n_i^{'} \geqslant 3)$ observations, $X_{i1}^{'}, ...., X_{in_i^{'}}^{'}$ from each random samples, $X_{i1}, ...., X_{in_i}$.

Employ the first $n_0$ observations, $(2 \leqslant n_0 < n_i^{'})$ and calculate the sample mean and sample variance of these observations, respectively by

$$\tilde{X}_i = \frac{1}{n_0} \sum_{j=1}^{n_0} X_{ij}^{'} \quad , \quad \tilde{S}_i^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (X_{ij}^{'} - \tilde{X}_i)^2.$$

It is noted that Chen and Lam (1989) suggested taking $n_0$ to be $\min \left\{ n_i^{'} - 1,\ i = 1, ..., k \right\}$.

Choose the weights for the $i$th sample observations to be

$$U_i = \frac{1}{n_i'} + \frac{1}{n_i'}\sqrt{\frac{n_i' - n_0}{n_0}\left(\frac{n_i' h^*}{\tilde{S}_i^2} - 1\right)} \quad , \qquad V_i = \frac{1}{n_i'} - \frac{1}{n_i'}\sqrt{\frac{n_0}{n_i' - n_0}\left(\frac{n_i' h^*}{\tilde{S}_i^2} - 1\right)},$$

where $h^*$ is the maximum of $\left\{\frac{\tilde{S}_1^2}{n_1'}, ..., \frac{\tilde{S}_k^2}{n_k'}\right\}$.

Let final weighted sample mean using all observations be defined by

$$\tilde{Y}_i = U_i \sum_{j=1}^{n_0} X_{ij}' + V_i \sum_{j=n_0+1}^{n_i'} X_{ij}'.$$

and $\tilde{Y}^*$ represents the largest one of $\tilde{Y}_1, ..., \tilde{Y}_k$. Then the $100\%(1-\alpha)$ confidence interval for largest mean, $\mu^*$ is

$$(\tilde{Y}^* - d_1\sqrt{h^*}, \tilde{Y}^* + d_2\sqrt{h^*}), \tag{4}$$

where $d_1$ and $d_2$ are left entry and right entry percentage points, respectively, that are given by Chen and Wen (2006).

## 3. Numerical Example

The data, taken from Bishop and Dudewicz (1978), is an experiment for studying the bacterial killing. The experiment involved testing four types of solvents for their effects on the ability of a fungicide methyl-2-benzimidazole-carbamate to destroy the fungus Penicillium expansum. The fungicide was diluted in exactly the same manner in the four different types of solvents and sprayed on the fungus, and the percentage of fungus destroyed was measured. Let $\mu_i$ denote the mean percentage of fungus destroyed by solvent $i$. Our interest is in finding an interval estimation for the largest mean percentage, $\mu^* = \max(\mu_1, \mu_2, \mu_3, \mu_4)$.

The summary statistics of all data for different solvents and other needed values for one-stage sampling procedure by $n_0 = 15$, are given in Table 1 (For details see Chen and Wen (2006)).

The 95% proposed generalized confidence interval in [3] and one-stage confidence interval in [4] for the largest mean, $\mu^*$ are (97.038, 97.953) and (96.290, 97.868), respectively.

Table 1: Summary statistics of bacterial killing ability example

| Statistics | $n_i$ | $\bar{x}_i$ | $s_i^2$ | $n_i'$ | $\tilde{x}_i$ | $\tilde{s}_i^2$ |
|---|---|---|---|---|---|---|
| Solvent 1 | 19 | 97.18 | 2.091 | 18 | 96.8420 | 2.1099 |
| Solvent 2 | 28 | 95.41 | 4.112 | 25 | 94.6860 | 3.1708 |
| Solvent 3 | 52 | 95.42 | 4.764 | 49 | 94.3833 | 5.8842 |
| Solvent 4 | 16 | 97.37 | 0.753 | 16 | 97.3327 | 0.7797 |

## 4. Simulation Study

For comparing the coverage probabilities and expected length of the confidence intervals for the largest mean, simulations were studied for the case of $k = 4$ only. The observations are generated with size $n_i$ from four normal populations with means $\mu_i$ and variances $\sigma_i^2$, $i = 1, 2, 3, 4$. Also we considered that $\mu^* = \mu_1 = 10$, that is, the largest mean is 10 and the first population is the best population.

The generalized confidence interval (GC) in (3) that is considered by Chang and Huang (2000), compared by the optimal confidence interval (OC) of Chen and Chen (2006) in (4). The results are given in Tables 2 and 3. We found that the coverage probabilities of generalized confidence interval and optimal interval are close to significant level and in some cases the coverage probabilities of optimal interval is greater than significant level. Also the expected length of generalized confidence interval is smaller than expected length optimal confidence interval.

## References

[1] K. Alam and K. M. L. Saxena, On interval estimation of a ranked parameter. *J. Roy. Statist. Soc*, B 36 (1974), 277-283.

[2] T. A. Bishop and E. J. Dudewicz, Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics,* 20 (1978), 419-430.

[3] Y. P. Chang and W. T. Huang, Generalized confidence intervals for the largest value of some functions of parameters under normality. *Statistics Sinica*, 10 (2000), 1369-1383.

[4] H. J. Chen, Strong consistency and asymptotic unbiasedness of a natural estimator for a ranked parameter. *Sankhya*, Ser. B 38 ( 1976), 92-94.

[5] H. J. Chen, Estimation of ordered parameters from k stochastically increasing distributions. *Naval Res. Logist. Quarterly,* 24 (2) (1977a), 269-280.

[6] H. J. Chen, A class of fixed-width confidence intervals for a ranked normal mean. *Comm. Statist.* B 6 (1977b), 137-156.

[7] S. Y. Chen and H. J. Chen, Single-stage analysis of variance under heteroscedasticity. *Comm. Statist. Simulation Comput.* 27 (3) (1998), 641-666.

[8] H. J. Chen and S. Y. Chen, A nearly optimal confidence interval for the largest normal mean.*Comm. Statist. Simulation Comput.* 28 (1) (1999), 131-146.

[9] H. J. Chen and S. Y. Chen, Optimal confidence interval for the largest normal mean with unknown variance. *Comput. Stat. Data Anal.,* 47 (2004), 845-866.

[10] H. J. Chen and E. J. Dudewicz, Procedures for fixed-width interval estimation of the largest normal mean. *J. Amer. Statist. Assoc.*, 71, (1976), 752-756.

[11] H. J. Chen and K. Lam, Single-stage interval estimation of the largest normal mean under heteroscedasticity. *Comm. Statist. Theory and Methods* 18 (10) (1989), 3703-3718.

[12] H. J. Chen and M. J. Wen, Optimal confidence interval for the largest normal mean under heteroscedasticity. *Comput. Stat. Data Anal.,* 51 ( 2006), 982-1001.

[13] E. J. Dudewicz, Two-sided confidence intervals for ranked means. *J. Amer. Statist. Assoc.,* 67 (1972), 462-464.

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

[15] J. Neter , H. Kutner, C. J. Nachtsheim, and W. Wasserman, Applied *Statistical Linear Models*, 4th Edition. IRWIN, Inc., Harvard, IL, (1996).

[16] K. M. L. Saxena, A single-sample procedure for the estimation of the largest mean. *J. Amer. Statist. Assoc.* 71 (1976), 147-148.

[17] K. M. L. Saxena and Y. L. Tong, Interval estimation of the largest mean of k normal populations with known variance. *J. Amer. Statist. Assoc.,* 64 (1969), 296-299.

[18] Y. L. Tong, An asymptotically optimal sequential procedure for the estimation of the largest mean. *Ann. Statist.,* 1 (1973), 175-179.

[19] K. W. Tsui and S. Weerahandi, Generalized p-values in significance testing of hypothesis in the presence of nuisance parameters, *J. Am. Statist. Assoc.*, 84 (1989), 602-607.

[20] S. Weerahandi, Generalized confidence intervals, *J. Am. Statist. Assoc.*, 88 (1993), 899-905.

[21] S. Weerahandi, *Exact statistical methods for data analysis*, Springer, NewYork, (1995).

Table 2: Simulation of %95 coverage probabilities with unequal variances and $\mu^* = \mu_1 = 10$

| $\mu_2, \mu_3, \mu_4$ | 1, 1, 1 | | 1, 2, 5 | | 2, 5, 7 | |
|---|---|---|---|---|---|---|
| | $OC$ | $GC$ | $OC$ | $GC$ | $OC$ | $GC$ |
| $\sigma^2 = (1, 1, 2, 2)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 0.951 | 0.958 | 0.947 | 0.951 | 0.954 | 0.954 |
| $n = (10, 10, 20, 20)$ | 0.944 | 0.944 | 0.956 | 0.947 | 0.953 | 0.939 |
| $n = (20, 10, 10, 30)$ | 0.937 | 0.969 | 0.959 | 0.948 | 0.956 | 0.944 |
| $n = (10, 30, 30, 30)$ | 0.961 | 0.943 | 0.942 | 0.951 | 0.947 | 0.955 |
| $n = (30, 30, 30, 30)$ | 0.951 | 0.946 | 0.950 | 0.958 | 0.956 | 0.965 |
| $\sigma^2 = (4, 1, 2, 3)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 0.937 | 0.957 | 0.944 | 0.953 | 0.944 | 0.951 |
| $n = (10, 10, 20, 20)$ | 0.941 | 0.943 | 0.947 | 0.959 | 0.955 | 0.942 |
| $n = (20, 10, 10, 30)$ | 0.947 | 0.954 | 0.955 | 0.950 | 0.959 | 0.950 |
| $n = (10, 30, 30, 30)$ | 0.956 | 0.938 | 0.954 | 0.954 | 0.952 | 0.946 |
| $n = (30, 30, 30, 30)$ | 0.944 | 0.941 | 0.941 | 0.961 | 0.955 | 0.947 |
| $\sigma^2 = (1, 10, 2, 1)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 0.947 | 0.942 | 0.961 | 0.965 | 0.961 | 0.953 |
| $n = (10, 10, 20, 20)$ | 0.941 | 0.954 | 0.956 | 0.949 | 0.960 | 0.950 |
| $n = (20, 10, 10, 30)$ | 0.951 | 0.942 | 0.942 | 0.952 | 0.944 | 0.947 |
| $n = (10, 30, 30, 30)$ | 0.950 | 0.959 | 0.941 | 0.947 | 0.942 | 0.945 |
| $n = (30, 30, 30, 30)$ | 0.946 | 0.939 | 0.948 | 0.958 | 0.962 | 0.949 |
| $\sigma^2 = (5, 2, 1, 10)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 0.948 | 0.938 | 0.963 | 0.949 | 0.957 | 0.962 |
| $n = (10, 10, 20, 20)$ | 0.947 | 0.945 | 0.942 | 0.949 | 0.960 | 0.957 |
| $n = (20, 10, 10, 30)$ | 0.953 | 0.946 | 0.954 | 0.954 | 0.955 | 0.954 |
| $n = (10, 30, 30, 30)$ | 0.956 | 0.945 | 0.943 | 0.953 | 0.951 | 0.941 |
| $n = (30, 30, 30, 30)$ | 0.945 | 0.948 | 0.951 | 0.955 | 0.962 | 0.956 |

Table 3: Simulation of expected length of %95 confidence interval with unequal variances and $\mu^* = \mu_1 = 10$

| $\mu_2, \mu_3, \mu_4$ | 1, 1, 1 | | 1, 2, 5 | | 2, 5, 7 | |
|---|---|---|---|---|---|---|
| | *OC* | *GC* | *OC* | *GC* | *OC* | *GC* |
| $\sigma^2 = (1, 1, 2, 2)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 5.202 | 1.387 | 5.162 | 1.401 | 5.317 | 1.401 |
| $n = (10, 10, 20, 20)$ | 4.210 | 1.373 | 4.147 | 1.372 | 4.129 | 1.393 |
| $n = (20, 10, 10, 30)$ | 4.462 | 0.923 | 4.304 | 0.927 | 3.716 | 0.927 |
| $n = (10, 30, 30, 30)$ | 3.441 | 1.397 | 3.431 | 1.395 | 3.324 | 1.391 |
| $n = (30, 30, 30, 30)$ | 2.902 | 0.738 | 2.878 | 0.744 | 2.781 | 0.743 |
| $\sigma^2 = (4, 1, 2, 3)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 6.733 | 2.781 | 6.709 | 2.787 | 6.840 | 2.770 |
| $n = (10, 10, 20, 20)$ | 5.839 | 2.805 | 5.934 | 2.825 | 5.929 | 2.773 |
| $n = (20, 10, 10, 30)$ | 5.324 | 1.854 | 5.903 | 1.842 | 5.298 | 1.854 |
| $n = (10, 30, 30, 30)$ | 5.158 | 2.767 | 5.220 | 2.804 | 5.152 | 2.804 |
| $n = (30, 30, 30, 30)$ | 3.741 | 1.782 | 3.771 | 1.484 | 3.767 | 1.482 |
| $\sigma^2 = (1, 10, 2, 1)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 8.936 | 1.386 | 8.903 | 1.397 | 8.445 | 1.391 |
| $n = (10, 10, 20, 20)$ | 8.469 | 1.384 | 8.081 | 1.381 | 8.026 | 1.389 |
| $n = (20, 10, 10, 30)$ | 8.142 | 0.915 | 8.151 | 0.916 | 8.163 | 0.931 |
| $n = (10, 30, 30, 30)$ | 5.606 | 1.392 | 5.430 | 1.404 | 5.558 | 1.386 |
| $n = (30, 30, 30, 30)$ | 4.788 | 0.739 | 5.004 | 0.741 | 4.908 | 0.743 |
| $\sigma^2 = (5, 2, 1, 10)$ | | | | | | |
| $n = (10, 10, 10, 10)$ | 10.97 | 4.354 | 11.07 | 4.318 | 10.83 | 4.138 |
| $n = (10, 10, 20, 20)$ | 9.477 | 4.429 | 9.305 | 4.416 | 9.266 | 4.174 |
| $n = (20, 10, 10, 30)$ | 8.035 | 2.938 | 7.901 | 2.955 | 7.764 | 2.861 |
| $n = (10, 30, 30, 30)$ | 8.225 | 4.449 | 8.530 | 4.476 | 8.389 | 4.147 |
| $n = (30, 30, 30, 30)$ | 5.967 | 2.346 | 6.229 | 2.360 | 5.921 | 2.329 |

**Javad Behboodian**
Department of Mathematics
Islamic Azad University - Shiraz Branch
Shiraz, Iran.
E-mail: behboodian@susc.ac.ir

**Ali Akbar Jafari**
Department of Mathematics
Shiraz University
Shiraz, Iran