Journal of Mathematical Sciences & Computer Applications 2 (2): 118–124, 2015
doi: 10.5147/jmsca.2015.0177

On Selecting the Normal Population with the Largest Absolute Mean

Khaled Hussein

Department of Mathematics, University of Wisconsin-Fond du Lac, Fond du Lac, WI 54935, USA
E-mail: khaled.hussein@uwc.edu

Received: January 13, 2015/ Accepted: April 14, 2015

Abstract

The problem of selecting the normal population with largest absolute mean is considered. An alternative procedure based on the absolute values of the sample medians and investigate its efficiency relative to Rizvi's means procedure has been studied.

Key Words: selection and ranking, correct selection, indifference-zone approach.

1. Introduction

Let $\Pi_1,\ldots,\Pi_k$ be $k\ (\geq 2)$ independent normal populations with unknown means $\mu_i, i=1,\ldots,k$, and common known variance σ^2. Let $\theta_i = |\mu_i|, i=1,\ldots,k$, and let $\theta_{[1]} \leq .. \leq \theta_{[k]}$ denote the ordered θ_i. It is assumed that there is no prior information regarding the correspondence between the ordered and the unordered θ_i. The population associated with the largest θ_i is called the best population. Our goal is to select the best population, using the indifference-zone formulation of Bechhofer (1954). Under this approach, a procedure is sought which will select one of the k populations as the best with a guaranteed probability $P^*\left(\frac{1}{k} < P^* < 1\right)$ of correct selection (i.e. selection of the best population) whenever $\theta_{[k]} - \theta_{[k-1]} \geq \delta$, where $\delta > 0$ and P^* are specified in advance. The part of the parameter space $\Omega = \{\underline{\theta} = (\theta_1,\ldots,\theta_k), \theta_i \geq 0, i=1,\ldots,k\}$ where $\theta_{[k]} - \theta_{[k-1]} \geq \delta$ holds is known as the preference-zone and is denoted here by Ω_δ. The complement of Ω_δ with respect to Ω is the indifference-zone, so-called because of no requirement on the PCS (probability of correct selection) when the true $\underline{\theta}$ falls in the region.

For the above selection problem, Rizvi (1971) studied a procedure based on the means of samples of size n drawn from the k populations. He, in fact, considered a more general goal of selecting the t $(1 \leq t \leq k-1)$ populations under the indifference-zone formulation, and also a subset selection procedure in the case of $t=1$ following the formulation of Gupta (1956).

Let $X_{i1},\ldots,X_{in}$ denote n independent observations from Π_i, and let $Y_i = |\overline{X}_i|$, where $\overline{X}_i$ is the sample mean, $i=1,\ldots,k$. Rizvi (1971) proposed the rule:

R_m: Select the population that yields the largest Y_i.

For this rule R_m, Rizvi has tabulated the minimum value of $\lambda = \sqrt{n}\delta$ needed to meet the guaranteed PCS for $k = 2(1)10$ and several selected values of P^*.

In the present paper, we propose an alternative procedure based on the absolute values of the sample medians and investigate its efficiency relative to Rizvi's means procedure.

2. Preliminaries

In this section, we discuss some preliminary results regarding the absolute value of the median of a random sample drawn from a normal population.

2.1 Distribution of the Absolute Value of the Sample Median

Let $X_1, \ldots, X_n$ denote independent observations from a normal population with mean μ and variance σ^2. For convenience, we assume that n is odd so that the rth order statistics $X_{(r)}$ with $r = \frac{n+1}{2}$ becomes the median. Let G and g denote the distribution and the density functions of $X_{(r)}$, respectively, when the sample is drawn from N(0, 1), the standard normal population. It is well-known (see, for example, Arnold, Balakrishnan and Nagaraja (1992), pp. 10 and 13) that

$$G(x) = I_{\Phi(x)}(r, n-r+1) \tag{2.1}$$

and

$$g(x) = r\binom{n}{r}[\Phi(x)\{1-\Phi(x)\}]^s \phi(x) \tag{2.2}$$

where $r = \frac{n+1}{2}, s = \frac{n-1}{2}$, $\Phi(\phi)$ denotes the standard normal distribution (density) function, and

$$I_p(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^p u^{a-1}(1-u)^{b-1} du,\ a > 0, b > 0.$$

Now, let H and h denote the c.d.f. and the density function of $T = | X_{(r)} |$. Then, for $t \geq 0$,

$$H(t) = P\{-t \leq X_{(r)} \leq t\}$$
$$= P\left\{-\frac{t+\mu}{\sigma} \leq Z_{(r)} \leq \frac{t-\mu}{\sigma}\right\}$$

Where $Z_{(r)}$ is the rth order statistics in a sample of size n from an N(0, 1) distribution.

Hence, using (2.1), we get

$$H(t) = \begin{cases} G\left(\frac{t-\mu}{\sigma}\right) - G\left(-\frac{t+\mu}{\sigma}\right), & t \geq 0 \\ 0 & otherwise \end{cases} \tag{2.3}$$

and

$$h(t)=\begin{cases}\frac{1}{\sigma}[g\left(\frac{t-\mu}{\sigma}\right)+g\left(\frac{t+\mu}{\sigma}\right)], & t\geq 0\\ 0 & otherwise\end{cases} \qquad (2.4)$$

2.2. Properties of Distribution H

We first note from (2.2) that the distribution of the sample median from N(0, 1) is symmetric about zero, i.e. $g(-y)=g(y)$ for all y. Consequently, $G(-y)=1-G(y)$ for all y.

Theorem 1. The distribution of T depends on μ only through its absolute value.

Proof. This follows directly from (2.3) using the symmetry of the distribution of the sample median from N(0, 1).

In view of the above theorem, we will now rewrite (2.3) and (2.4) as

$$H_\theta(t)=\begin{cases}G\left(\frac{t-\theta}{\sigma}\right)-G\left(-\frac{t+\theta}{\sigma}\right), & t\geq 0\\ 0 & otherwise\end{cases} \qquad (2.5)$$

$$h_\theta(t)=\begin{cases}\frac{1}{\sigma}[g\left(\frac{t-\theta}{\sigma}\right)+g\left(\frac{t+\theta}{\sigma}\right)], & t\geq 0\\ 0 & otherwise\end{cases} \qquad (2.6)$$

where $\theta=|\mu|$.

In order to stablish theorem 2, we need the following Lemma

Lemma 1. $g(y)$ is increasing in $y\leq 0$ and decreasing in $y\geq 0$.

Proof. Since $g(y)$ is symmetric about zero, it is enough to show that $g(y)$ is decreasing in $y\geq 0$, or equivalently, that $\log g(y)$ is decreasing in $y\geq 0$. From (2.2), we get

$$\frac{d}{dx}\log g(y)=s\left\{\frac{\phi(y)}{\Phi(y)}-\frac{\phi(y)}{1-\Phi(y)}\right\}-y$$

which is negative for $y>0$ because $\Phi(y)>1-\Phi(y)$. This proves the lemma.

Theorem 2. The distribution of T is stochastically decreasing in θ.

Proof. We can take $\sigma=1$ without loss of generality. We need to show that

$\frac{\partial}{\partial\theta}H_\theta(t)=g(t+\theta)-g(t-\theta)<0$ for all t and $\theta>0$. When $t-\theta\leq 0$, we get $g(t+\theta)<g(t-\theta)$ by Lemma 1. On the other hand, when $\theta-t<0$, we get $g(t-\theta)=g(\theta-t)$, by symmetry of g

$> g(t+\theta)$, by Lemma 1, since $0 < \theta - t < \theta + t$.

In either case, $g(t+\theta) - g(t-\theta) < 0$. This completes the proof of Theorem 2.

3. The Proposed Selection Rule and It Infimum

Letting the $T_{(i)}$ denote the sample median from the population associated with $\theta_{[i]}$, $i = 1, \cdots, k$, the PCS is given by

$$\begin{aligned} P(CS \setminus R_M) &= \Pr\{T_{(k)} \geq T_{(j)},\ j = 1, \cdots, k-1\} \\ &= \int_0^\infty \prod_{j=1}^{k-1} H_{\theta_{(j)}}(t) h_{\theta_{(k)}} dt \\ &\geq \int_0^\infty H^{k-1}_{\theta_{[k]}-\delta}(t) h_{\theta_{[k]}}(t) dt \qquad for\, \underline{\theta} \in \Omega_\delta, \end{aligned}$$

since $H_\theta(t)$ is stochastically decreasing in θ. Thus, the infimum of $P(CS \setminus R_M)\, over\, \Omega_\delta$ occurs for a configuration of the type

$$\theta_{[1]} = \cdots = \theta_{[k-1]} = \theta = \theta_{[k]} - \delta$$

We now have to evaluate the infimum of

$$\begin{aligned} I(\theta, \delta) &= \int_0^\infty H^{k-1}_{\theta_{[k]}-\delta}(t) h_{\theta_{[k]}}(t) dt \\ &= \int_0^\infty \left[G(t-\theta) - G(-t-\theta)\right]^{k-1} g(t-\theta-\delta) dt + \int_0^\infty \left[G(t-\theta) - G(-t-\theta)\right]^{k-1} g(t+\theta+\delta) dt. \end{aligned}$$

Theorem 3:

$I(\theta, \delta) = \int_0^\infty \left[G(t-\theta) - G(-t-\theta)\right]^{k-1} g(t-\theta-\delta) dt + \int_0^\infty \left[G(t-\theta) - G(-t-\theta)\right]^{k-1} g(t+\theta+\delta) dt$ is a strictly increasing function of θ for $\theta \geq 0$ and a fix δ.

Proof:

Setting $y = t - \theta$ in the first integral and $y = t + \theta$ in the second integral, we get

$$I(\theta, \delta) = \int_{-\theta}^\infty \left[G(y) - G(-y-2\theta)\right]^{k-1} g(y-\delta) dy + \int_\theta^\infty \left[G(y-2\theta) - G(-y)\right]^{k-1} g(y+\delta) dy$$

Now, by differentiation under the integral sign,

$$\frac{\partial I(\theta,\delta)}{\partial\theta} = 2(k-1)\int_{-\theta}^{\infty}[G(y)-G(-y-2\theta)]^{k-2}\,g(y-\delta)g(y+2\theta)dy$$

$$-2(k-1)\int_{\theta}^{\infty}[G(y-2\theta)-G(-y)]^{k-2}\,g(y+\delta)g(y-2\theta)dy$$

Again, by changing the variables of integration by setting $y = t-\theta$ in the first integral and $y = t+\theta$ in the second integral, and combining both, we get

$$\frac{\partial I(\theta,\delta)}{\partial\theta} = 2(k-1)\int_{0}^{\infty}[G(t-\theta)-G(-t-\theta)]^{k-2}\times[g(t+\theta)g(t-\theta-\delta)-g(t-\theta)g(t+\theta+\delta)]dy\,.$$

But it is easily seen from the strict monotone likelihood ratio property of the normal p.d.f. that $g(t+\theta)g(t-\theta-\delta) > g(t-\theta)g(t+\theta+\delta)$ for $t>0$, $\delta>0$.

Hence $\frac{\partial I(\theta,\delta)}{\partial\theta} > 0$ for all $\theta \geq 0$ and a fix δ.

Consequently, $P(CS \setminus R_M)$ is minimized over Ω_δ by seeting

$$\theta_{[1]} = \cdots = \theta_{[k-1]} = 0 = \theta_{[k]} - \delta$$

and

$$\inf_{\Omega_\delta} P(CS \setminus R_M) = \int_{0}^{\infty}[G(t)-G(-t)]^{k-1}[g(t-\delta)+g(t+\delta)]dt$$

4. Asymptotic Results

It is well-known (see, for example, Arnold, Balakrishnan and Nagaraja (1992), p. 225) that the sample median $X_{(r)}$ of n independent observation from $N(\mu,\sigma^2)$ population has an asymptotic $(n\to\infty)$ normal distribution with mean μ and variance $\frac{\pi\sigma^2}{2n}$. Thus, for large n, PCS for our procedure R_M can be obtained from the expression for PCS of Rizvi's procedure R_m by replacing σ^2 by $\frac{\pi\sigma^2}{2}$. Thus the asymptotic least favorable configuration (LFC), which yields the infimum of PCS, is the same, namely,

$$\theta_{[1]} = \cdots = \theta_{[k-1]} = 0 = \theta_{[k]} - \delta$$

It is now easy to see, from Rizvi's results, that

$$\frac{\sqrt{n_m}\,\delta}{\sigma} = \frac{\sqrt{n_M}\,\delta}{\sigma}\sqrt{\frac{2}{\pi}} \qquad \text{Which gives} \qquad n_M = \frac{\pi}{2}n_m\,,$$

Where n_m and n_M denote the minimum sample sizes required by the rules R_m and R_M, respectively, in order to meet the guaranteed PCS.

5. Relative Efficiency of R_M W. R. T. R_m

As before, let n_m and n_M denote the minimum sample sizes required by the rules R_m (based on the sample means) and R_M (based on the sample medians), respectively, in order to guarantee a minimum PCS of specified level P^*. Then the efficiency of R_M relative to R_m is defined by

$$eff(R_M, R_m) = \frac{n_m}{n_M}$$

so that less than 100% efficiency means a larger sample size required by R_M. The efficiency is computed for specified values of $k, P^*, and\ \delta$. The relative efficiency is given in the table below. The following conclusions are drawn from the table:

1) For any given $k\ and\ P^*$, $eff(R_M, R_n)$ decreases in δ.
2) For given $k\ and\ \delta$, $eff(R_M, R_m)$ is less for $P^* = 0.90$ than for $P^* = 0.95$
3) For any $\delta\ and\ P^*$, $eff(R_M, R_n)$ decreases as k increases.

6. References

1. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N, (1992). A First Course in Order Statistics, John Wiley and Sons, New York.
2. Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances, Annals of Mathematical Statistics, 25, 16-39.
3. Gupta, S.S. (1956). On a decision rule for a problem in ranking means, (Ph.D.Theesis), Mimeo. Ser. No. 105, University of North Carolina, Chapel Hill.
4. Hussein, Khaled; Panchapakesan, S. On selection from normal populations in terms of the absolute values of their means, Advances on theoretical and methodological aspects of probability and statistics (Hamilton, ON, 1998), 371-389, Taylor & Francis, London, 2002.
5. Hussein, Khaled; Panchapakesan, S. Simultaneous selection of extreme populations from a set of two-parameter exponential populations. Advances in reliability, 813-830, Handbook of Statist., 20, North-Holland, Amsterdam, 2001
6. Rizvi, M. H. (1971). Some selection problems involving folded normal distribution, Technometrics, 13, 335-369.