Selection Procedures Based on Ranks: Scale Parameter Case

by Prem S. Puri*

University of California, Berkeley

Purdue University, Lafayette

and Madan L. Puri**

Courant Institute of Mathematical Sciences, New York University

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 105

April 1967

---

*This investigation was supported in part by a research grant (No. GM-10525) from the National Institute of Health, Public Health Service and the Aerospace Research Laboratories Contract AF33(615)67C1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

**This research was supported by U.S. Navy Contract NONR-285(38).

Selection Procedures Based on Ranks: Scale Parameter Case

by Prem S. Puri*

University of California, Berkeley

Purdue University, Lafayette

and Madan L. Puri**

Courant Institute of Mathematical Sciences, New York University

1. Summary. In continuation of the authors' paper [12], selection procedures based on ranks for ranking c populations with regard to scale parameters and for selecting a subset containing populations "better than a standard" are proposed here and their properties studied.

2. Introduction. Multiple decision procedures for ranking c populations according to their means, variances, etc. have been extensively studied by several authors notably by Bahadur [1], Bechhofer [3], Bechhofer and Sobel [4], Gupta and Sobel ([5], [6]), Paulson [9] and Lehmann [7], among others, (See Barr and Rizvi [2] for references). Most of these procedures are based on statistics not dependent on the ranks of the observations but directly on the observed values themselves. Unfortunately not much work has been done so far, in developing procedures based on ranks even though in several practical situations one would prefer such procedures because of their invulnerability to gross errors. Only recently work in this direction has been initiated by Lehmann [8] followed by Puri and Puri [12]. Lehmann [8] considers a class of selection procedures based on ranks for selecting the "best" population, that is, the

*This investigation was supported in part by a research grant (No. GM-10525) from the National Institute of Health, Public Health Service and the Aerospace Research Laboratories Contract AF33(615)67C1244 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

**This research was supported by U.S. Navy Contract NONR-285(38).

one having the largest location parameter. In [12], the authors consider the problems of selecting out of c populations (a) the "t best" ones without regard to order (b) the "t best" ones with regard to order and (c) finding a subset which contains all populations "better" than a standard one. In these procedures the bestness of a population is characterized by its location parameter; the best being the one with the largest location parameter, and so on.

In the present paper, however, the procedures based on ranks are considered for the problems (a), (b) and (c) with the difference that the bestness of a population is characterized instead by its scale parameter. (For the general background and motivation for such problems the reader is referred to interesting papers by Bechhofer and Sobel [4] and Gupta and Sobel ([5], [6]).) The asymptotic efficiencies (cf. [8] and [12]) of these procedures relative to the normal theory procedures developed in [4] and [5] are obtained and discussed.

3. The Mathematical Model. Let X_{ij}, $j = 1,2,\dots,n$; $i = 1,2,\dots,c$ be independent samples from populations $\Pi_1, \Pi_2,\dots, \Pi_c$ with cumulative distribution functions $F(x/\sigma_i)$; $i = 1,2,\dots,c$, which are assumed to be continuous but otherwise unknown. (One might start instead with $F(\frac{x-\mu_i}{\sigma_i})$; $i = 1,2,\dots,c$, and assume that the location parameters μ_i's are known and consider the random variables $X_{ij} - \mu_i$.) Let $\sigma_{[1]} \leq \sigma_{[2]} \leq \cdots \leq \sigma_{[c]}$ be the ranked σ's; of course it is not known which population is associated with $\sigma_{[i]}$. Also, let $\theta_{ij} = \sigma^2_{[i]}/\sigma^2_{[j]}$; $i,j = 1,2,\dots,c$. We assume that a population is characterized by its parameter value. Thus the "best" population is the one which has the smallest value; the second being the one which has the second smallest parameter value, and so on. Our aim is to develop selection procedures based on ranks for the problems (a) - (c) mentioned in the introduction.

We shall denote the sample mean square $\sum_{j=1}^{n} X_{ij}^2/n$ by S_i^2 ; the sample mean square associated with the population having the scale parameter $\sigma_{[i]}$ by $S_{(i)}^2$ and the ranked S_i^2, $i = 1,2,\ldots,c$ by $S_{[1]}^2 < S_{[2]}^2 < \ldots < S_{[c]}^2$.

4. Problem (a).

4.1 Bechhofer - Sobel (B-S) Procedure.

For this problem it is assumed that the experimenter can specify a smallest value of $\theta_{t+1,t}$ say $\theta^* > 1$, that he desires to detect. The experimenter must also specify the smallest acceptable probability γ $(1 > \gamma > 1/\binom{c}{t})$ of correct selection when $\theta_{t+1,t} \geq \theta^*$. When F is normal Bechhofer and Sobel [4] have proposed the following procedure: Select the t populations associated with

(1) $$S_{[1]}^2, \ldots, S_{[t]}^2 \quad .$$

The usual problem is to determine the smallest sample size n required to guarantee the desirable property that

(2) $$P\ (\text{correct selection of } t \text{ best populations}) \geq \gamma \ ,$$

subject to the condition

(3) $$\theta_{t+1,t} \geq \theta^* \ .$$

If we do not wish to rely on the assumption of normality, we can find a large - sample solution which depends only on the kurtosis γ_2 of F . To this end, we reformulate the problem by considering a sequence of situations for increasing n where we replace θ^* by $\theta^{(n)}$ and determine the large sample solution for n required to guarantee (2) subject to the condition

(4) $$\theta_{t+1,t} \geq \theta^{(n)} .$$

It may be remarked (see also [8], [12]) that even though θ^* is shown to depend upon n which would imply that the set of acceptable values of $\theta_{t+1,t}$ are changing with n, yet in practice this will not be the case. To consider $\theta^{(n)}$ as a sequence depending upon n is only a mathematical device to approximate the actual situation and in practice $\theta^{(n)}$ will be identified with the given value of θ^* .

It may be noted that the procedure (1) depends on $|X_{ij}|$, $j = 1,2,\ldots,n$; $i = 1,2,\ldots,c$, with $G(x/\sigma_i)$ as the distribution function of $|X_{ij}|$, where

(5) $$G(x/\sigma_i) = \begin{cases} F(x/\sigma_i) - F(-x/\sigma_i), & \text{for } x \geq 0 \\ 0 & \text{, otherwise.} \end{cases}$$

As such, for all F the left side of (2) subject to (4) takes on its minimum value when

(6) $$\theta_{t,1} = \theta_{t+1,c} = 1 \; ; \; \theta_{t+1,t} = \theta^{(n)} .$$

This follows from the stochastic increasing property of the family of distributions $G(x/\sigma_i)$. We refer to the condition (6) as the least favourable configuration of θ's . The sample size n is determined by

(7) $$P(\max(S^2_{(1)},\ldots, S^2_{(t)}) < \min(S^2_{(t+1)},\ldots,S^2_{(c)})) = \gamma$$

under the assumption that (6) holds. In the rest of this section it is assumed that mean and variance of F are zero and unity respectively, which in turn implies that $S^2_{(i)}$ is an unbiased estimator of $\sigma^2_{[i]}$ for each i .

The following lemma gives the large sample solution of the sample size problem.

Lemma 1. For fixed γ, and under the condition (4), let n be so determined that (6) and (7) hold. Then as $n \to \infty$,

$$\theta^{(n)} = 1 + \Delta \sqrt{\frac{\gamma_2 + 2}{n}} + o(n^{-\frac{1}{2}}) , \tag{8}$$

where Δ is determined by the condition

$$\gamma = t\, Q_{c-1} (\underbrace{0, 0, \ldots, 0}_{(t-1) \text{ times}}, \underbrace{\frac{\Delta}{\sqrt{2}}, \ldots, \frac{\Delta}{\sqrt{2}}}_{(c-t) \text{ times}}) , \tag{9}$$

where Q_{c-1} is the cumulative distribution function of a normally distributed vector $(U_1, \ldots U_{t-1}, W_{t+1}, \ldots, W_c)$ with

$$\begin{cases} E(U_i) = E(W_j) = 0;\ Cov(U_i, U_{i'}) = \frac{1}{2}(\delta_{ii'} + 1);\ Cov(W_j, W_{j'}) = \frac{1}{2}(\delta_{jj'} + 1), \\ Cov(U_i, W_j) = -\frac{1}{2};\ i, i' = 1, 2, \ldots, t-1;\ j, j' = t+1, \ldots, c , \end{cases} \tag{10}$$

where δ's are the kronecker deltas.

The proof follows by making the logarithmic transformation of $S^2_{(i)}$ and proceeding essentially as in lemma 1 of [8]. For a given θ^*, it follows from the above lemma that a large sample size solution for which (2) holds subject to (3) is given by

$$n = \frac{(\gamma_2 + 2)\, \Delta^2}{(\theta^* - 1)^2} . \tag{11}$$

4.2 Procedure based on ranks. Let $Z^{(i)}_{N,j} = 1$ if jth smallest of $N = cn$ absolute values $|X_{ij}|$, $j = 1, 2, \ldots, n$; $i = 1, 2, \ldots, c$ is from the ith sample and otherwise $Z^{(i)}_{N,j} = 0$. Denote

(12) $$T_i = \frac{1}{n} \sum_{j=1}^{N} E_{N,j} \, Z_{N,j}^{(i)} \; ; \; i = 1,2,\ldots,c,$$

where $E_{N,j}$ is the expected value of the square of the jth order statistic of a sample of size N from a given continuous distribution function F_o. Furthermore, denote the statistic associated with the population having the scale parameter $\sigma_{[i]}$ by $T_{(i)}$ and let the ranked T_i's be denoted by $T_{[1]} < T_{[2]} < \cdots < T_{[c]}$. Then the proposed procedure based on ranks is as follows:

Select the t populations associated with

(13) $$T_{[1]}, T_{[2]}, \ldots, T_{[t]} \quad .$$

This procedure will be referred to as F_o - score procedure $T(F_o)$.

In sequel, we shall study the relative performance of the procedures (1) and (13) on the basis of the sample sizes required to guarantee (2), subject to the condition that the deviation of $\theta_{ij}^{(m)}$ from one is of the order $m^{-\frac{1}{2}}$ where m denotes the sample size for the procedure $T(F_o)$. To be precise we assume that

(14) $$\theta_{ij}^{(m)} = 1 + (\xi_{[i]} - \xi_{[j]}) \cdot m^{-\frac{1}{2}} + o(m^{-\frac{1}{2}}) \; ,$$

where ξ's are some constants satisfying

(15) $$\xi_{[1]} \leq \xi_{[2]} \leq \cdots \leq \xi_{[c]} \; .$$

Suppose now that m is the smallest sample size required for the procedure (13) if it is to satisfy (2) subject to the conditions (14) and

(16) $$\theta_{t+1,t}^{(m)} \geq \tilde{\theta}^{(m)} \quad ,$$

where $\tilde{\theta}^{(m)}$ is greater than one and is given by (27). Then the following theorem helps in obtaining the least favourable configuration of $\tilde{\theta}$'s subject to (14) for the $T(F_o)$ procedure.

Theorem 1. For $m = 1,2,\ldots,$ let X_{ij} $(j = 1,2,\ldots,m; i = 1,2,\ldots,c)$ be independently distributed according to $F_i(x) = F(x/\sigma_i^{(m)})$ with the sequence of parameter points $\underset{\sim}{\sigma}^{(m)} = (\sigma_1^{(m)},\ldots, \sigma_c^{(m)})$ satisfying (14) and suppose that the assumptions of theorem 6.1 and lemma 7.2 of [10] are satisfied. Then the limiting distribution of the random vector

$$\left(\frac{m}{2A^2}\right)^{\frac{1}{2}} [T_{(i)} - T_{(j)} - \mu_i(\underset{\sim}{\sigma}^{(m)}) + \mu_j(\underset{\sim}{\sigma}^{(m)}); \ i = 1,2,\ldots,t; \ j = t+1,\ldots,c],$$

where

(17) $$\mu_i(\underset{\sim}{\sigma}^{(m)}) = \int_0^\infty J[H(x)]dG_i(x), \quad H(x) = \sum_{i=1}^{c} G_i(x)/c; \ J = (F_o^{-1})^2$$

is the distribution of a $t(c-t)$ dimensional normal vector $(V_{ij}; \ i = 1,2,\ldots,t; \ j = t+1,\ldots,c)$ with

(18) $$\begin{cases} E(V_{ij}) = 0, \ \mathrm{Var}(V_{ij}) = 1, \ \mathrm{Cov}(V_{ij}, V_{ij'}) = \mathrm{Cov}(V_{ij}, V_{i'j}) = 1/2 \\ \mathrm{Cov}(V_{ij}, V_{i'j'}) = 0; \ i \neq i', \ j \neq j'; \ i,i' = 1,2,\ldots,t; \ j,j' = t+1,\ldots,c \end{cases}$$

where

(19) $$A^2 = \int_0^1 J^2(x)dx - \left(\int_0^1 J(x)dx\right)^2 .$$

The proof of this theorem follows immediately from the results of Puri [11] and is therefore omitted.

The probability of correct selection of t best populations is given by

(20) $$P[\max(T_{(1)},\ldots,T_{(t)}) < \min(T_{(t+1)},\ldots,T_{(c)}]$$

$$= P[T_{(i)} - T_{(j)} < 0;\ i = 1,2,\ldots,t;\ j = t+1,\ldots,c]$$

$$= P[(\frac{m}{2A^2})^{\frac{1}{2}} (T_{(i)} - T_{(j)} - \mu_{(i)}(\underset{\sim}{\sigma}^{(m)}) + \mu_{(j)}(\underset{\sim}{\sigma}^{(m)}))$$

$$< (\frac{m}{2A^2})^{\frac{1}{2}} (\mu_{(j)}(\underset{\sim}{\sigma}^{(m)}) - \mu_{(i)}(\underset{\sim}{\sigma}^{(m)}));$$

$$i = 1,2,\ldots,t;\ \ j = t+1,\ldots,c] .$$

On the other hand, for large m (c.f. Lemma 7.2 [10])

(21) $$\sqrt{m}(\mu_{(j)}(\underset{\sim}{\sigma}^{(m)}) - \mu_{(i)}(\underset{\sim}{\sigma}^{(m)})) \sim \tfrac{1}{2}\sqrt{m}\ \ I(J,G)(\theta_{ji}^{(m)} - 1)$$

where

(22) $$I(J,G) = \int_0^\infty x \frac{d}{dx} \{J[G(x)]\} dG(x) .$$

Hence using theorem 1 the right hand side of (20) is asymptotically equivalent to

(23) $$P[V_{ij} < \tfrac{1}{2}(\frac{m}{2A^2})^{\frac{1}{2}} I(J,G)(\theta_{ji}^{(m)} - 1);\ i = 1,2,\ldots,t;\ j = t+1,\ldots,c]$$

Again, in view of (16), $\underset{\sim}{\sigma}^{(m)}$ satisfies

(24) $$\sigma_{[1]}^{(m)} < \sigma_{[2]}^{(m)} < \ldots < \sigma_{[t]}^{(m)} < \sigma_{[t]}^{(m)} (\tilde{\theta}^{(m)})^{\frac{1}{2}} < \sigma_{[t+1]}^{(m)} < \ldots < \sigma_{[c]}^{(m)} .$$

The least favourable configuration among σ's satisfying (24) and (14), for which (20) and hence (23) takes on the minimum value, is given by

(25) $$\theta_{t,1}^{(m)} = \theta_{t+1,c}^{(m)} = 1; \qquad \theta_{t+1,t}^{(m)} = \tilde{\theta}^{(m)} .$$

Thus for large samples, the sample size m is determined by the condition

(26) $$P[\max(T_{(1)},\ldots,T_{(t)}) < \min(T_{(t+1)},\ldots,T_{(c)}] = \gamma$$

where the left hand side is derived subject to (25). The following lemma is analogue of lemma 1.

Lemma 2. For fixed γ, let m be determined so that (26) holds and suppose that G and $J = (F_o^{-1})^2$ satisfy the regularity conditions of theorem 2.1 of [11] and of lemma 7.2 of [10]. Then as $m \to \infty$

(27) $$\tilde{\theta}^{(m)} = 1 + \frac{\Delta}{\sqrt{m}} \quad \frac{2\,A}{I(J,G)} + o(m^{-\frac{1}{2}}) ,$$

where A, $I(J,G)$ and Δ are given by (19), (22) and (9) respectively.

Outline of the proof. It is easily observable that the left side of (26) depends on $\sigma_{[i]}$'s only through θ's. As such, in view of θ's satisfying (25), let

(28) $$G_i(x) = G(x/\sigma_i^{(m)}) ,$$

where the ordered $\sigma_i^{(m)}$'s are given by

(29) $$\sigma_{[i]}^{(m)} = \begin{cases} \sigma_{[t]};\ i = 1,\ldots,t; \\ (\tilde{\theta}^{(m)})^{\frac{1}{2}} \cdot \sigma_{[t]};\ i = t+1,\ldots,c \end{cases} .$$

Then from [10] it follows that the random variables $m^{\frac{1}{2}}(T_i - \mu_i(\sigma^{(m)}))$, $i = 1,2,\ldots,c$, where the μ_i's are given by (17), have asymptotically a joint normal distribution with zero means and covariance matrix

(30) $$\sigma_{ii'} = (\delta_{ii'} - \frac{1}{c})\, A^2 ,$$

where $\delta_{ii'}$ are the kronecker deltas, A^2 is given by (19) and where $\sigma_1^{(m)} = (\sigma_1^{(m)},\ldots,\sigma_c^{(m)})$. The result now follows by proceeding precisely as in lemma 4B.1 of [2].

As before equating $\tilde{\theta}^{(m)}$ with θ^* and using the above lemma, the large sample solution for m for which (2) holds subject to (16) is given by

$$m = \left[\frac{2\,\Delta\,A}{I(J,G)(\theta^*-1)}\right]^2 . \tag{31}$$

Since the sample sizes n and m as given in (11) and (31) are found for the two procedures so as to satisfy the same requirements for θ's of the form (14), namely (2) subject to (3), the relative efficiency $e_{T(F_o),BS}(F)$ (See Lehmann [8], Puri and Puri [12]) of $T(F_o)$ procedure is given by the ratio

$$\lim_{n\to\infty} \frac{n}{m} = \frac{(\gamma_2+2)\; I^2(J,G)}{4\; A^2} . \tag{32}$$

It may be noted that if F is the standard normal distribution and F_o is the chi-distribution with one degree of freedom the above efficiency is one. In general, for all distributions F which are symmetrical with respect to the origin, the efficiency (32) is one whenever $F_o = G$.

Remark. Proceeding as in [11], one may construct a class of tests based on the statistics $\underset{\sim}{T} = (T_1,\ldots,T_c)$ for testing the hypothesis of the equality of the scale parameters. It can then be shown that the asymptotic Pitman-relative efficiency of these tests relative to the classical Bartlett's test is identical with (32). In particular, for all distributions F with F(0) = 0, the test based on $\underset{\sim}{T}$ coincides with the one proposed by Puri [11]. It would be of interest to make efficiency comparisons of the test based on $\underset{\sim}{T}$ with the one proposed in [11], for the cases where $F(0) > 0$.

Finally, given that the parameter point satisfies (6), we observe that $T(F_o)$ and B-S procedures have asymptotically the same performance characteristic provided $m = g(n)$ is determined so as to satisfy (32). On the other hand, let

$$(33) \qquad \theta^{(n)}_{t+1,i} = \Delta^{(n)}_i = 1 + \frac{\Delta_i \sqrt{\gamma_2+2}}{\sqrt{n}} + o(n^{-\frac{1}{2}}),\ i = 1,2,\ldots,c;\ i \neq t+1 ,$$

where not all $\Delta_i = \Delta$ for $i = 1,2,\ldots,t$ and/or not all $\Delta_i = 0$ for $i = t+2,\ldots,c$, so that the parameter point does not satisfy (6). However, condition (4) still holds so that

$$(34) \qquad \sigma^{(n)}_{[1]} < \sigma^{(n)}_{[2]} < \ldots < \sigma^{(n)}_{[t]} < \sigma^{(n)}_{[t]} (\theta^{(n)})^{\frac{1}{2}} < \sigma^{(n)}_{[t+1]} < \ldots < \sigma^{(n)}_{[c]} .$$

Under these conditions, following Lehmann [7] and Puri and Puri [12] it can be easily shown that under the usual regularity conditions imposed on $G(x/\sigma_i)$, the two procedures have the same asymptotic performance characteristics, provided $m = g(n)$ is determined so as to satisfy (32).

5. <u>Problem</u> (b). In this section we shall consider the problem of selecting out of c populations the "t best" ones <u>with</u> regard to order, only briefly as its discussion runs parallel to that of problem (a).

5.1 <u>B-S Procedure</u>. Under this procedure, the "t best" populations with regards to order are selected to be the ones associated with $s^2_{[1]},\ldots,s^2_{[t]}$ respectively. The problem is to find a large sample size solution for n, such that

(35) P(correct selection of t best populations with regard to order) $\geq \gamma$,

where γ is a preassigned positive number and σ's are subject to the conditions

(36) $$\theta_{i+1,i} \geq \delta^{(n)}; \quad i = 1,2,\ldots,t ,$$

and where $\delta^{(n)}$ is greater than one and is given by (39) below.

Following the argument of section 4.1, it is easy to establish that for all F, the least favourable configuration of θ's turns out to be

(37) $$\theta_{i+1,i} = \delta^{(n)}; \quad i = 1,2,\ldots,t; \ \theta_{t+1,c} = 1 .$$

Thus the sample size n is determined as the solution of the equation

(38) $$P[S^2_{(1)} < S^2_{(2)} < \ldots < S^2_{(t)} < \min(S^2_{(i)}; \ i = 1+1,\ldots,c)] = \gamma ,$$

where the left side of (38) is evaluated subject to (37). The following lemma provides the large sample solution for n. Its proof being elementary is omitted.

Lemma 3. For fixed γ, let n be determined so that (37) and (38) hold. Then as $n \to \infty$

(39) $$\delta^{(n)} = 1 + \frac{\delta\sqrt{\frac{\gamma_2+2}{2}}}{\sqrt{n}} + o(n^{-\frac{1}{2}}) ,$$

where δ is determined by the condition

(40) $$\gamma = (c-t)\, Q_{c-1}\,(\underbrace{\frac{\delta}{\sqrt{2}}, \frac{\delta}{\sqrt{2}},\ldots, \frac{\delta}{\sqrt{2}}}_{t \text{ times}},\ \underbrace{0,0,\ldots,0}_{(c-t-1)\text{times}}); \ 1 \leq t \leq c-1 ,$$

where Q is the cumulative distribution function of a normally distributed vector $(U_1,\ldots,U_t;\ W_{t+1},\ldots,W_{c-1})$ satisfying

$$(41)\quad \begin{cases} E(U_i) = E(W_j) = 0;\ \mathrm{Cov}(W_j, W_{j'}) = \frac{1}{2}(1+\delta_{ii'}) \\ \mathrm{Cov}(U_i, W_j) = -\frac{1}{2} \text{ \underline{if} } i = t, \text{ \underline{and zero otherwise}.} \\ \mathrm{Cov}(U_i, U_{i'}) = 1 \text{ or } -\frac{1}{2} \text{ \underline{according as} } i = i' \text{ or } |i'-i|=1, \text{ \underline{and} } 0 \text{ \underline{otherwise},} \\ i,i' = 1,2,\ldots,t;\ j,j' = t+1,\ldots,c-1. \end{cases}$$

5.2 Procedure Based on Ranks and their performance characteristic. With T_i, $i = 1,2,\ldots,c$ defined as in (12), the $T(F_o)$ procedure for selection of "t best" populations with regard to order is to select populations associated with $T_{[1]},\ldots,T_{[t]}$ in that order. As before we restrict ourselves to the set of θ_{ij}'s satisfying the condition (14). Then arguing as in section 4.2 it can be shown that the left side of (35) subject to the condition

$$(42)\qquad \theta^{(m)}_{i+1,i} \geq \tilde{\delta}^{(m)} \ ; \qquad i = 1,2,\ldots,t$$

when θ_{ij}'s satisfy the least favourable configuration is given by

$$(43)\qquad \theta^{(m)}_{i+1,i} = \tilde{\delta} \ ; \qquad i = 1,2,\ldots,t;\ \theta^{(m)}_{t+1,c} = 1 \ .$$

Thus for large samples, the sample size m is determined by the condition

$$(44)\qquad P[T_{(1)} < T_{(2)} < \ldots < T_{(t)} < \min(T_{(i)};\ i = t+1,\ldots,c)] = \gamma$$

where the left side is calculated subject to (43). Furthermore, as in section 4.2, it can be shown under the conditions of lemma 2 that as $m \to \infty$

$$(45)\qquad \tilde{\delta}^{(m)} = 1 + \frac{\delta}{\sqrt{m}}\ \frac{2A}{I(J,G)} + o(m^{-\frac{1}{2}}) \ .$$

Finally, if the large sample solutions for m and n are determined by (45) and (39) respectively and if the θ's are subject to the same condition (36) for both the procedures, so that $\tilde{\delta}^{(m)} = \delta^{(n)}$, it is easily seen that the asymptotic efficiency of procedure $T(F_o)$ relative to B-S procedure <u>with</u> regard to order, is same as in (32). Furthermore, it can be shown that if the ratio of the sample sizes m and n is equal to the efficiency (32) the two procedures have the same asymptotic performance characteristic for the case where for each n, the parameter point satisfies

$$\theta_{t+1,i}^{(n)} = \delta_i^{(n)} = 1 + \frac{\delta_i \sqrt{\gamma_2 + 2}}{\sqrt{n}} + o(n^{-\frac{1}{2}}) , \tag{46}$$

for $i \neq t + 1$; $i = 1,2,\ldots,c$, and the condition (37) is not necessarily satisfied.

6. <u>Problem</u> (c). Let X_{ij}; $i = 0,1,2,\ldots,c$; $j = 1,2,\ldots,n$, be $c + 1$ independent samples from populations $\Pi_0, \Pi_1, \ldots, \Pi_c$ having continuous cumulative distribution functions $F(x/\sigma_o), F(x/\sigma_1),\ldots,F(x/\sigma_c)$ respectively. Π_0 is assumed here to represent the standard population. We shall call population Π_i as "good or better than the standard" if

$$\theta_{io} \leq \theta^{(n)} , \tag{47}$$

where $\theta_{io} = \sigma_i^2/\sigma_o^2$. Here $\theta^{(n)}$ is known to be less than or equal to one and in actual practice is identified as before with a constant $0 < \theta^* \leq 1$, prescribed by the experimenter. We now consider the problem of selecting a subset of c populations such that the probability that all populations "better than the standard one" are included in this subset is at least a given number γ . The reader is referred to Gupta and Sobel [5] for a similar formulation of this problem.

6.1 Normal Theory Procedure. Consider the following selection procedure: Select Π_i if and only if $S_i^2 \leq S_o^2$, where S_i^2, $i = 1,2,\ldots,c$ are the sample mean squares as defined in section 3, for Π_o, Π_i, $i = 1,2,\ldots,c$ respectively. Suppose without loss of generality that the only better populations are $\Pi_1, \Pi_2, \ldots, \Pi_s$ $(s \leq c)$, where s is unknown. Then $\theta_{io} \leq \theta^{(n)}$ for $i = 1,2,\ldots,s$ and $\theta_{io} > \theta^{(n)}$ for $i = s+1,\ldots,c$. We wish to determine the minimum sample size n such that

$$\text{(48)} \qquad P(\text{selected subset includes } \Pi_1,\ldots,\Pi_s) \geq \gamma .$$

We restrict our attention to the case when $1 \leq s \leq c$, since when $s = 0$ that is when none of the c populations is better than the standard, (48) is trivially satisfied. Since s is unknown, we find the least favourable configuration jointly of both σ's as well as s, for which the left side of (48) is minimum. This we attain in two stages. Firstly, for any fixed s, the least favourable configuration of σ's is

$$\text{(49)} \qquad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_s^2 = \theta^{(n)} \sigma_o^2 ,$$

the left side of (48) being independent of $\sigma_{s+1},\ldots,\sigma_c$. This follows from the stochastic increasing property of the family of distributions $G(x/\sigma_i)$ defined in section 4.1; as such

$$\text{(50)} \qquad \min_{\substack{\theta_{io} \leq \theta^{(n)} \\ i=1,2,\ldots,s}} P[\text{selected subset includes } \Pi_1,\ldots,\Pi_s | \Pi_1,\ldots,\Pi_s \text{ are good}]$$

$$= P[S_i^2 \leq S_o^2;\ i = 1,2,\ldots,s | \theta_{io} = \theta^{(n)};\ i = 1,2,\ldots,s] .$$

Secondly, since the right side of (50) is a decreasing function of s, the least favourable value of s is $s = c$. The sample size n is therefore determined by the condition

$$\gamma = \min_{s} \min_{\substack{\theta_{io} \le \theta^{(n)} \\ i=1,2,\ldots,s}} P[\text{selected subset includes } \Pi_1,\ldots,\Pi_s | \Pi_1,\ldots,\Pi_s \text{ are good}] \tag{51}$$

$$= P[S_i^2 \le S_o^2;\ i = 1,2,\ldots,c | \theta_{io} = \theta^{(n)};\ i = 1,2,\ldots,c] \ .$$

The following lemma now provides a large sample size solution for n .

Lemma 4. *For fixed* γ *and with a "goodness" of a population defined by* (47) *let* n *be determined so that* (51) *holds. Then as* $n \to \infty$,

$$\theta^{(n)} = 1 - \frac{\nabla\sqrt{\gamma_2+2}}{\sqrt{n}} + o(n^{-\frac{1}{2}}) \ . \tag{52}$$

Here ∇ *is determined by the condition*

$$\gamma = Q_c\left(\frac{\nabla}{\sqrt{2}},\ldots,\frac{\nabla}{\sqrt{2}}\right) , \tag{53}$$

where Q_c *is the c.d.f. of a normally distributed vector* $(Y_1,\ldots,Y_c)$ *with*

$$E\ Y_i = 0,\ \mathrm{Cov}(Y_i,Y_{i'}) = \tfrac{1}{2}(1+\delta_{ii'});\ i,i' = 1,2,\ldots,c \ . \tag{54}$$

The proof being straight forward is omitted. Equating $\theta^{(n)}$ of (52) with the preassigned value θ^* , the large sample size solution for n is given by

$$n = \frac{\nabla^2(\gamma_2+2)}{(1-\theta^*)^2} \ . \tag{55}$$

Finally since $0 < \theta^{(n)} \le 1$, it is clear from (52) that $\nabla \ge 0$, so that from (53) we have

$$\gamma \ge Q_c(0,0,\ldots,0) = \frac{1}{c+1} \ . \tag{56}$$

6.2 Procedures based on ranks. Let $Z_{M,r}^{(i)} = 1$ if the rth smallest of $M = m(c+1)$ absolute values $|X_{ij}|$; $i = 0,1,2,\ldots,c$; $j = 1,2,\ldots,m$ is from the ith sample and otherwise let $Z_{M,r}^{(i)} = 0$. Denote

$$T_i = \frac{1}{m} \sum_{r=1}^{M} E[V^{(r)}]^2 \, Z_{M,r}^{(i)}; \; i = 0,1,2,\ldots,c \;; \tag{57}$$

where $V^{(1)} < V^{(2)} < \ldots < V^{(M)}$ is an ordered sample from a given distribution F_o and E denotes the expectation. The following distribution-free procedure is then proposed:

(58) Select Π_i if and only if $T_i \leq T_o$; $i = 1,2,\ldots,c$,

we shall propose later an alternative rank procedure which is asymptotically equi-efficient in Pitman sense to the procedure (59).

As before, the aim is to find the minimum sample size $m = g(n)$ such that by using procedure (59), (48) is satisfied. An argument similar to the one used in section 4.2, leads to the least favourable configuration of σ's (amongst the set of σ's satisfying $\theta_{io}^{(m)} = 1 + O(m^{-\frac{1}{2}})$) and s in the present case as

$$s = c, \; \theta_{io}^{(m)} = \tilde{\theta}^{(m)}, \; i = 1,2,\ldots,c \;, \tag{59}$$

where $\theta^{(n)}$ of (48) is now identified by $\tilde{\theta}^{(m)}$ as given by (61). Thus the sample size $m = g(n)$ is the solution for m of the equation

$$P[T_i \leq T_o; \; i = 1,2,\ldots,c] = \gamma \;, \tag{60}$$

where the left side is derived under the condition that $(\sigma_o^{(m)}, \sigma_1^{(m)}, \ldots, \sigma_c^{(m)})$

satisfies (59). Then under the conditions of lemma 2, it is found in a manner similar to that used in section 4.2, that as $m \to \infty$,

(61) $$\tilde{\theta}^{(m)} = 1 - \frac{\nabla}{\sqrt{m}} \frac{2A}{I(J,G)} + o(m^{-\frac{1}{2}}) \quad ,$$

where ∇ ,A and $I(J,G)$ are given by (53), (19) and (22) respectively. The large sample size solution for m is then given by equating $\tilde{\theta}^{(m)}$ and θ^* , as

(62) $$m = [\frac{2A \ \nabla}{I(J,G)(1-\theta^*)}]^2 \quad .$$

The asymptotic efficiency of the rank procedure (58) relative to the normal theory procedure turns out to be same as given by (23). Finally, if we consider a sequence of parameter points $(\sigma_o^{(m)}, \sigma_1^{(m)}, \ldots, \sigma_c^{(m)})$ not satisfying (59), it can be shown as in the previous sections that if the ratio of the sample sizes n/m is equal to the efficiency (32), the two procedures have the same asymptotic performance. The details are omitted to avoid repitition.

6.3 <u>An Alternative Rank Procedure</u>. Consider an alternative rank procedure, where we combine each of the c samples X_{ij}; $i = 1,2,\ldots,c$; $j = 1,2,\ldots,m$ with the sample X_{oj}; $j = 1,2,\ldots,m$ from the standard population, instead of combining all the $(c+1)$ samples together. Let for $i = 1,2,\ldots,c$, $Z_{m,r}^{(i,o)} = 1$ if the rth smallest of $2m$ absolute values $\{|X_{ij}|, |X_{oj}|; j = 1,2,\ldots,m\}$ is from the ith sample and zero otherwise. Denote for $i = 1,2,\ldots,c$,

(63) $$\tilde{T}_i = \frac{1}{m} \sum_{i=1}^{2m} E[V^{(r)}]^2 \; Z_{m,r}^{(i,o)}$$

and

(64) $$\tilde{T}_{i,o} = \frac{1}{m} \sum_{r=1}^{2m} E[V^{(r)}]^2 \; (1-Z_{m,r}^{(i,o)}) \; ;$$

where $V^{(1)} < \ldots < V^{(2m)}$ is an ordered sample from a given distribution F_o. The alternative procedure is then to

(65) Select Π_i if and only if $\tilde{T}_i \leq \tilde{T}_{i,o}$; $i = 1,2,\ldots,c$,

which is equivalent to $\tilde{T}_i \leq$ constant.

Again following Puri [10], it can be shown that under the normal regularity conditions, the joint limiting distribution of the random variable $(2m)^{\frac{1}{2}}[\tilde{T}_1-\mu_1, \ldots, \tilde{T}_c - \mu_c]/A$ is the distribution of a c-dimensional normal vector $(U_1,\ldots,U_c)$ satisfying (54), where A is given by (19);

(66) $$\mu_i = \int J[H_{i,o}(x)] \; dG(x/\sigma_i)$$

(67) $$H_{i,o} = \tfrac{1}{2} [G(x/\sigma_i) + G(x/\sigma_o)]; \; J = (F_o^{-1})^2 \; ,$$

and G is as defined in (5).

One obtains results parallel to the ranks procedures (58) with regards to asymptotic performance; the details are omitted. It may be added, however, that as in [12] the two rank procedures (58) and (65) can be shown to be equi-efficient, in the sense that their large sample solutions are the same.

References

[1] Bahadur, R.R. (1950). On a problem in the theory of k populations. Ann. Math. Stat. 21, 362-375.

[2] Barr, Capt. D.R. and Rizvi, M.H. (1966). An introduction to ranking and selection procedures. J. Am. Stat. Assn. 61, 640-646.

[3] Bechhofer, R.E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Stat. 25, 273-289.

[4] Bechhofer, R.E. and Sobel, M. (1954). A single-sample multiple decision procedure for ranking variances of normal populations, Ann. Math. Stat. 25, 273-289.

[5] Supta, S.S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard, Ann.Math. Stat. 29, 235-244.

[6] Gupta, S.S. and Sobel, M. (1962). On selecting a subset containing the population with the smallest variance, Biometrika, 49, 495-507.

[7] Lehmann, E.L. (1961). Some model I problems of selection. Ann. Math. Stat. 32, 990-1012.

[8] Lehmann, E.L. (1963). A class of selection procedures based on ranks. Math. Annalen, 150, 268-275.

[9] Paulson, E. (1952). On the comparison of several experimental categories with a control. Ann. Math. Stat. 23, 239-246.

[10] Puri, M.L. (1964). Asymptotic efficiency of a class of c-sample tests. Ann. Math. Stat. 35, 102-121.

[11] Puri, M.L. (1965). On some test of homogeneity of variances, Ann. Inst. Stat. Math., 17, 323-330.

[12] Puri, M.L. and Puri, P.S. (1966). Some distribution free multiple decision procedures for certain problems in analysis of variance, submitted to Ann. Math. Stat..

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
selection procedures ranking and selection scale parameters						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (*Corporate author*) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP
3. REPORT TITLE Selection Procedures based on ranks: Scale parameter case	
4. DESCRIPTIVE NOTES (*Type of report and inclusive dates*) Technical Report	
5. AUTHOR(S) (*Last name, first name, initial*) Puri, Prem S. and Puri, Madan L.	

6. REPORT DATE April 1967	7a. TOTAL NO. OF PAGES 17	7b. NO. OF REFS 12
8a. CONTRACT OR GRANT NO. AF 33(615)67C1244	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series No. 105	
b. PROJECT NO.		
c.	9b. OTHER REPORT NO(S) (*Any other numbers that may be assigned this report*)	
d.		

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES Also supported by the National Institute of Health and the Office of Naval Research.	12. SPONSORING MILITARY ACTIVITY Aerospace Research Laboratories Wright Patterson AFB Dayton, Ohio

13. ABSTRACT

In continuation of the authors' paper [12] here selection procedures based on ranks for ranking c populations with regard to scale parameters and for selecting a subset containing populations 'better than a standard' are proposed and their properties studied.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification