58

On Some Selection and Ranking Procedures with Applications to Multivariate Populations*

by

Shanti S. Gupta

and

William J. Studden

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 58

October, 1965

*Research supported by Contract NONR-1100(26) with the Office of Naval Research and by Contract AF 33(657)11737 with the Aerospace Research Laboratories. Reproduction in whole or in part permitted for any purpose of the United States Government.

On Some Selection and Ranking Procedures with Applications to Multivariate Populations*

by

Shanti S. Gupta

and

William J. Studden

Purdue University

Introduction and Summary

This paper is concerned with ranking and selection of k multivariate normal populations. The selection and ranking problem is formulated in terms of suitably defined scalar functions. For k multivariate normal populations with mean vectors $\mu_i'(i=1,2,\ldots,k)$ each of which has p components, a function that arises naturally, is the scalar quantity $\lambda_i = \mu_i' \Sigma_i^{-1} \mu_i$ where Σ_i is the covariance matrix of the ith population. With suitably defined statistics the ranking of multivariate normal populations in terms of λ_i can be reduced to the ranking of non-centrality parameters of non-central chi-square or non-central F distributions.

We are interested in selecting the populations with large (small) values of the parameters λ_i. The procedures to be defined select a non-empty subset which is small and yet large enough to guarantee a certain basic probability requirement. This requirement is that the population with the largest value of the parameter is included in the selected subset with probability at least equal to a given number $P^*(1/k < P^* < 1)$. This type of problem has been studied in a number of recent papers. For a

*Research supported by Contract NONR-1100(26) with the Office of Naval Research and by Contract AF 33(657)11737 with the Aerospace Research Laboratories.

rather complete bibliography, reference should be made to Gupta (1965).

In Section 2, a formal statement of the problems is given and procedures are defined for selecting populations with the largest and smallest parameters. Probability of a correct selection and its partial infimum are evaluated.

Section 3 deals with a general result concerning the infimum of the probability of a correct selection. In Section 4 applications to multivariate populations are given.

2. Formal Statement of the Problem

Suppose each of the k population, $\pi_1, \pi_2, \ldots, \pi_k$ has an observable random variable $Y_i (i=1,2,\ldots,k)$ whose density function is $f_{\lambda_i}(y), y \geq 0, \lambda_i \geq 0$. We assume that the density function $f_\lambda(y)$ has a monotone likelihood ratio. This implies that the expected value of Y is a monotone increasing function of λ. In all specific cases to be considered the mean value will be a linear increasing function of λ.

Let the ranked λ's be denoted by

$$\lambda_{[1]} \leq \lambda_{[2]} \leq \cdots \leq \lambda_{[k]} . \tag{2.1}$$

It is assumed that there is no a priori information available about the correct pairing of the ordered $\lambda_{[i]}$ values and the k given populations. Any population associated with $\lambda_{[k]}$ $(\lambda_{[1]})$ will be called a best population. A correct selection is defined as the selection of any subset of the k populations which includes a best population. Our problem is to define a selection procedure which selects a small, non-empty subset of the k populations and guarantees that a best population has been included with probability at least P^* $(1/k < P^* < 1)$. If CS stands for a correct selection then our goal is to define a decision rule R such that

$$\inf_{\Omega} P\{CS|R\} \geq P^* \tag{2.2}$$

where $\Omega = \{(\lambda_1,\lambda_2,\ldots,\lambda_k): \lambda_i \geq 0, \text{ all } i\}$.

Selection Procedures

Let y_i be an observation on $Y_i(i=1,2,\ldots,k)$. Then the procedures for selecting the population with the largest value $\lambda_{[k]}$ is

R: Select π_i iff

$$cy_i \geq y_{max}, \quad c > 1 \tag{2.3}$$

where $c = c(k, P^*)$ is the minimal value for which (2.2) is satisfied. Similarly, the procedure R' for selecting a subset containing the population with the smallest value $\lambda_{[1]}$ is defined to be

R': Select π_i iff

$$y_i \leq b\, y_{min}, \quad b > 1 \tag{2.4}$$

where $b = b(k,P^*)$ is again the minimal value for which (2.2) is satisfied.

Probability of a Correct Selection and Its Infimum

We will now derive an expression for the probability of a correct selection and its infimum. Let $y_{(i)}$ $(i=1,2,\ldots,k)$ be the observation which has come from the population $\pi_{(i)}$ with parameter $\lambda_{[i]}$. It should be noted that $y_{(i)}$ is one of the numbers y_i $(i=1,2,\ldots,k)$ though it is not known to us. For selecting the population associated with $\lambda_{[k]}$, we then have

$$(2.5) \qquad P\{\text{Selecting } \pi_{(k)}|R\} = P\{cy_{(k)} \geq y_{max}\}$$

$$= P\{cy_{(k)} \geq y_{(j)},\ j=1,2,\ldots,k-1\}$$

$$= \int_0^\infty \left[\prod_{j=1}^{k-1} F_{\lambda_{[j]}}(cy)\right] f_{\lambda_{[k]}}(y)\, dy\ .$$

Since $f_\lambda(y)$ is assumed to have a monotone likelihood ratio, it follows that $F_\lambda(y) \leq F_{\lambda'}(y)$ for all $\lambda > \lambda'$ and each y. In this case

$$(2.6) \qquad P\{\text{Selecting } \pi_{(k)}\} \geq \int_0^\infty \left[F_{\lambda_{[k]}}(cy)\right]^{k-1} f_{\lambda_{[k]}}(y)dy.$$

Since $P\{CS|R\} \geq P\{\text{Selecting } \pi_{(k)}|R\}$, we conclude that

$$(2.7) \qquad \inf_\Omega P\{CS|R\} \geq \inf_\lambda \int_0^\infty F_\lambda^{k-1}(cy)\ f_\lambda(y)\ dy.$$

For the problem of selecting the population with the smallest $\lambda_{[1]}$, a similar argument shows that

$$(2.8) \qquad \inf_\Omega P\{CS|R'\} \geq \inf_\lambda \int_0^\infty [1-F_\lambda(\tfrac{y}{b})]^{k-1}\ f_\lambda(y)dy.$$

In the next Section we discuss a general theorem dealing with the infima of the expressions on the right sides of (2.7) and (2.8).

3. A Result Concerning the Infima of Probability of a Correct Selection

Let $g_j(x), j=0,1,2,\ldots$ be a sequence of density functions on the interval $[0,\infty)$ and define

$$f_\lambda(x) = \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} g_j(x), \qquad x \geq 0. \tag{3.1}$$

For a fixed integer $k \geq 2$ and $c > 1$ let

$$I(\lambda) = \int_0^\infty [F_\lambda(cx)]^{k-1} f_\lambda(x)\, dx. \tag{3.2}$$

and

$$J(\lambda) = \int_0^\infty [1-F_\lambda(\tfrac{x}{c})]^{k-1} f_\lambda(x) dx. \tag{3.3}$$

The purpose of this section is to provide sufficient conditions on the sequence $g_j(x)$, $j=0,1,\ldots$ which guarantee that the functions $I(\lambda)$ and $J(\lambda)$ attain their minimum value on $[0,\infty)$ at the point $\lambda = 0$.

Theorem 3.1.

(i) If for each integer $\ell \geq 0$

$$\sum_{i=0}^{\ell} \frac{1}{i!(\ell-i)!} \{(G_{i+1}(cx)-G_i(cx))g_{\ell-i}(x) - cg_i(cx)(G_{\ell-i+1}(x)-G_{\ell-i}(x))\} \geq 0 \tag{3.4}$$

then the functions $I(\lambda)$ and $J(\lambda)$ defined in (3.2) and (3.3) are non-decreasing in λ.

(ii) If strict inequality holds in (3.4) for some integer ℓ then $I(\lambda)$ and $J(\lambda)$ are strictly increasing.

Corollary 3.1.a

Let

$$g_j(x) = \frac{x^{\mu+j-1}e^{-x}}{\Gamma(\mu+j)}, \qquad j = 0,1,\ldots \tag{3.5}$$

where $\mu > 0$. Then the functions $I(\lambda)$ and $J(\lambda)$ defined in (3.2) and (3.3) are strictly increasing.

Proof: For $g_j(x)$ defined by (3.5). Integrating by parts we see that

$$G_i(x) - G_{i+1}(x) = g_{i+1}(x), \quad i = 0,1,\ldots \tag{3.6}$$

For $\ell \geq 1$ we insert the above expression in (3.4) and combine the terms i and $\ell-i$. It may easily be shown that (3.4) reduces to

$$\sum_{i=0}^{[\frac{\ell-1}{2}]} \frac{e^{-x(c+1)}\, x^{2\mu+\ell-1}\, (\ell-2i)\, c^{i+\mu}\, (c^{\ell-2i}-1)}{\Gamma(\mu+i+1)\,\Gamma(\mu+\ell-i+1)\, i!\, (\ell-i)!} .$$

For $\ell \geq 1$ the above expression is strictly positive. For $\ell = 0$ equation (3.4) reduces to zero. The corollary thus follows from Theorem 3.1.

<u>Corollary 3.1.b.</u> Let

$$g_j(x) = \frac{\Gamma(\mu+\nu+j)}{\Gamma(\nu)\,\Gamma(\mu+j)} \frac{x^{\mu+j-1}}{(1+x)^{\mu+\nu+j}},\ x \geq 0, \tag{3.7}$$

where $\mu > 0$, $\nu > 0$ and $j = 0,1,\ldots$. Then the functions $I(\lambda)$ and $J(\lambda)$ defined in (3.2) and (3.3) are strictly increasing in λ.

<u>Proof.</u> The proof in this case proceeds as in Corollary 3.1.a. The expression corresponding to (3.6) is

$$G_j(x) - G_{j+1}(x) = \frac{\Gamma(\mu+\nu+j)}{\Gamma(\nu)\,\Gamma(\mu+j+1)} \frac{x^{\mu+j}}{(1+x)^{\mu+\nu+j}} . \tag{3.8}$$

Combining terms as in Corollary 3.1.a, equation (3.4) can be reduced to

$$\sum_{i=0}^{[\frac{\ell-1}{2}]} \frac{\Gamma(\mu+\nu+\ell-i)(\ell-2i)}{i!(\ell-i)!\,\Gamma(\mu+\ell-i+1)} \frac{x^{2\mu-2i-1}\,c^{\mu+i}}{[(1+x)(1+cx)]^{\mu+\nu+i}} \left[\left(\frac{cx}{1+cx}\right)^{\ell-2i} - \left(\frac{x}{1+x}\right)^{\ell-2i}\right] .$$

For $\ell \geq 1$ the above expression is positive since the function $[x/(1+x)]$ is strictly increasing in x. For $\ell=0$, (3.4) can be checked separately.

In order to prove Theorem 3.1 we first consider a number of elementary lemmas. For each integer $\alpha \geq 0$ we define $\underline{A}(\alpha)$ as the set of k-tuples $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ where $\alpha_i (i=1,\ldots,k)$ are non-negative integers and $\sum_{i=1}^{k} \alpha_i = \alpha$. The multinomial coefficient $\frac{\alpha!}{\alpha_1!\alpha_2!\ldots,\alpha_k!}$ will be denoted by

$\binom{\alpha}{\alpha_1\ \alpha_2 \ldots \alpha_k}$ as usual.

Lemma 3.1: The functions $I(\lambda)$ and $J(\lambda)$ defined in (3.2) and (3.3) can be expressed as

$$I(\lambda) = e^{-\lambda k} \sum_{\alpha=0}^{\infty} a_\alpha \lambda^\alpha \tag{3.9}$$

and

$$J(\lambda) = e^{-\lambda k} \sum_{\alpha=0}^{\infty} b_\alpha \lambda^\alpha \tag{3.10}$$

where

$$\alpha!\, a_\alpha = \sum_{A(\alpha)} \binom{\alpha}{\alpha_1 \ldots \alpha_k} \int_0^\infty \left\{\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\right\} g_{\alpha_k}(x)dx \tag{3.11}$$

and

$$\alpha!\, b_\alpha = \sum_{A(\alpha)} \binom{\alpha}{\alpha_1 \ldots \alpha_k} \int_0^\infty \left\{\prod_{i=1}^{k-1} \left[1-G_{\alpha_i}(\tfrac{x}{c})\right]\right\} g_{\alpha_k}(x)dx. \tag{3.12}$$

Proof. Equation (3.9) follows easily by inserting the expression for $f_\lambda(x)$ from (3.1) in (3.2). Equation (3.10) follows in the same manner after observing that

$$1-F_\lambda(x) = \sum_{j=0}^{\infty} \frac{\lambda^j e^{-\lambda}}{j!} (1-G_j(x)).$$

Lemma 3.2: The functions $I(\lambda)$ is nondecreasing provided

$$(\alpha+1)\, a_{\alpha+1} - k\, a_\alpha \geq 0, \qquad \alpha = 0,1,\ldots \, . \tag{3.13}$$

If strict inequality holds for some α then $I(\lambda)$ is strictly increasing. Similar statements hold for $J(\lambda)$ if a_α is replaced by b_α.

Proof: The above statements follow readily by differentiating the expressions (3.9) and (3.10).

Lemma 3.3:

(i) For each set of integers $\alpha_1, \alpha_2, \ldots, \alpha_k$ we have

$$\int_0^\infty \prod_{i=1}^{k-1} G_{\alpha_i}(cx)\; g_{\alpha_k+1}(x)dx = \int_0^\infty \prod_{i=1}^{k-1} G_{\alpha_i}(cx)\; g_{\alpha_k}(x)dx - \int_0^\infty \frac{d}{dx}\left[\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\right](G_{\alpha_k+1}(x)-G_{\alpha_k}(x))dx \tag{3.14}$$

(ii) Equation (3.14) remains true if the $k-1$ functions G_{α_i}, $i=1,\ldots,k-1$ are replaced by $1-G_{\alpha_i}, i = 1,\ldots,k-1$.

Proof: To proovepart (i) we first integrate the left side of (3.14) by parts to obtain

$$\int_0^\infty \prod_{i=1}^{k-1} G_{\alpha_i}(cx)\; g_{\alpha_k+1}(x)dx = 1 - \int_0^\infty \frac{d}{dx}\left[\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\right] G_{\alpha_k+1}(x)dx. \tag{3.15}$$

The right side of (3.15) is then written as

$$(3.16)\quad 1-\int_0^\infty \frac{d}{dx}\Big[\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\Big] G_{\alpha_k}(x)dx - \int_0^\infty \frac{d}{dx}\Big[\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\Big] (G_{\alpha_k+1}(x) - G_{\alpha_k}(x))\, dx\ .$$

Now applying (3.15), with α_k+1 replaced by α_k, to the first two terms of (3.16), the desired result (3.14) follows. Part (ii) is obtained by a similar argument.

We now proceed with the Proof of Theorem 3.1. We first show that if (3.4) holds then $I(\lambda)$ is nondecreasing. From (3.11) we have

$$(3.17)\qquad (\alpha+1)!\, a_{\alpha+1} = \sum_{A(\alpha+1)} \binom{\alpha+1}{\alpha_1,\ldots,\alpha_k} \int_0^\infty \Big\{\prod_{i=1}^{k-1} G_{\alpha_i}(cx)\Big\}\, g_{\alpha_k}(x)dx.$$

Since $\binom{\alpha+1}{\alpha_1,\ldots,\alpha_k} = \binom{\alpha}{\alpha_1-1,\alpha_2,\ldots,\alpha_k} + \binom{\alpha}{\alpha_1,\alpha_2-1,\alpha_3\ldots\alpha_k} + \ldots + \binom{\alpha}{\alpha_1,\ldots,\alpha_{k-1},\alpha_k-1}$

we rewrite (3.17), after a simple change of variables in the sums, as

$$(\alpha+1)!\, a_{\alpha+1} = \sum_{A(\alpha)} \binom{\alpha}{\alpha_1,\ldots,\alpha_k} \left\{ \sum_{j=1}^{k-1} \int_0^\infty G_{\alpha_j+1}(cx) \left\{ \prod_{i=1}^{k-1} G_{\alpha_i}(cx) \right\} g_{\alpha_k}(x)dx \right.$$

$$\left. + \int_0^\infty \left\{ \prod_{i=1}^{k-1} G_{\alpha_i}(cx) \right\} g_{\alpha_k+1}(x)dx \right].$$

Then

$$(\alpha+1)!\, a_{\alpha+1} = (k-1)\ \alpha!\ a_\alpha$$

$$+ \sum_{A(\alpha)} \binom{\alpha}{\alpha_1,\ldots,\alpha_k} \left[\sum_{j=1}^{k-1} \int_0^\infty (G_{\alpha_j+1}(cx) - G_{\alpha_j}(cx)) \left\{ \prod_{\substack{i=1 \\ i \neq j}}^{k-1} G_{\alpha_i}(cx) \right\} g_{\alpha_k}(x)dx \right.$$

$$\left. + \int_0^\infty \left\{ \prod_{i=1}^{k-1} G_{\alpha_i}(cx) \right\} g_{\alpha_k+1}(x)dx \right].$$

For the last integral in the above expression we insert its value from Lemma 3.1 to obtain

$$(\alpha+1)!\, a_{\alpha+1} = k\ \alpha!\ a_\alpha$$

$$+ \sum_{A(\alpha)} \binom{\alpha}{\alpha_1,\ldots,\alpha_k} \left[\sum_{j=1}^{k-1} \int_0^\infty (G_{\alpha_j+1}(cx) - G_{\alpha_j}(cx)) \left\{ \prod_{\substack{i=1 \\ i \neq j}}^{k-1} G_{\alpha_i}(cx) \right\} g_{\alpha_k}(x)dx \right.$$

$$\left. - \int_0^\infty \frac{d}{dx} \left\{ \prod_{i=1}^{k-1} G_{\alpha_i}(cx) \right\} (G_{\alpha_k+1}(x) - G_{\alpha_k}(x))dx \right]$$

$$= k\,\alpha!\, a_\alpha + \sum_{A(\alpha)} \binom{\alpha}{\alpha_1,\ldots,\alpha_k} \Big[\sum_{j=1}^{k-1} \int_0^\infty \prod_{\substack{i=1\\ i\neq j}}^{k-1} \{G_{\alpha_i}(cx)\} \Big[(G_{\alpha_j+1}(cx) - G_{\alpha_j}(cx)) g_{\alpha_k}(x)$$

$$- c\, g_{\alpha_j}(cx)(G_{\alpha_k+1}(x) - G_{\alpha_k}(x)) \Big] dx \Big] .$$

We now interchange the summations and fix $\alpha_i (i=1,\ldots,k-1, i\neq j)$.

Summing over α_j and α_k with $\alpha_j + \alpha_k = \ell = \alpha - \sum_{\substack{i=1\\ i\neq j}}^{k-1} \alpha_i$ we find that

$$(3.18) \qquad \alpha! \left[(\alpha+1)\, a_{\alpha+1} - k\, a_\alpha \right] \geq 0$$

provided (3.4) holds. Therefore the function $I(\lambda)$ is nondecreasing whenever (3.4) is satisfied for all integers $\ell (\ell \geq 0)$. Moreover if strict inequality holds for some ℓ in (3.4) then strict inequality holds for some α in (3.18) and hence $I(\lambda)$ is strictly increasing.

The proof of the statements concerning the function $J(\lambda)$ are analogous and will be omitted.

4. Selection and Ranking of Multivariate Normal Populations in terms of $\lambda_i = \mu_i' \Sigma_i^{-1} \mu_i$.

Let $\pi_i : N(\mu_i, \Sigma_i)$, $i=1,2,\ldots,k$ be p-variate normal populations with mean vectors μ_i and covariance matrix Σ_i, respectively. Let $\lambda_i = \mu_i' \Sigma_i^{-1} \mu_i$.

Case 1. Σ_i known, $(i=1,2,\ldots,k)$.

We take a sample of n independent observations from each of the k populations. Let $\underline{x}_{ij}$ denote the jth observation of the p-dimensional random vector on the ith population; then for each $j = 1,2,\ldots,n$, we compute

$$(4.1) \qquad y_{ij} = \underline{x}_{ij}' \Sigma_i^{-1} \underline{x}_{ij}, \quad i=1,2,\ldots,k; \; j=1,2,\ldots,n \; .$$

Since $y_{ij} = (1,2,\ldots,n)$ correspond to the n independent observations on a non-central χ^2 for each i, then $Y_i = \sum_{j=1}^{n} Y_{ij}$ is distributed as a non-central χ^2 with non-centrality parameter $\lambda_i' = n\lambda_i = n\underline{\mu}_i' \Sigma_i^{-1} \underline{\mu}_i$ and degrees of freedom $p' = np$. The proposed selection rule for the population with the largest value of λ_i is:

R: Select π_i iff

$$c \sum_{j=1}^{n} y_{ij} \geq \max_i \left\{ \sum_{j=1}^{n} y_{ij}; \; i = 1,2,\ldots,k \right\}$$

where the constant $c = c(k, np, P^*)$ $(c > 1)$, is determined to satisfy

$$(4.2) \qquad \inf_{\lambda'} \int_0^\infty F_{\lambda'}^{k-1}(cy) \, f_{\lambda'}(y) \, dy = P^*$$

where, now, $F_{\lambda'}(\cdot)$ and $f_{\lambda'}(\cdot)$ are the cdf and the density function of a non-central χ^2 with np d.f. Since the infimum of the above integral takes place when $\lambda' = 0$, by Corollary 3.1.a, we have, the equation determining c

$$(4.3) \qquad \int_0^\infty H_{p'}^{k-1}(cy)\, h_{p'}(y)dy = P^*, \quad p' = np$$

where

$$H_{p'}(x) = \int_0^x \frac{e^{-y}}{\Gamma(p')} y^{p'-1}dy \quad \text{and} \quad \frac{d}{dx} H_{p'}(x) = h_{p'}(x).$$

The values of $c' = 1/c$ satisfying (4.3) are given by Gupta (1963) for selected values of p' and P^* (see Table 1 , $p' = \nu/2$). Approximate c' values (obtained by using Wilson-Hilferty cube root transformation) are given by Gupta (1965) where the result concerning the infimum of $P\{CS|R\}$ is proved for the case $k=2$. Armitage and Krishnaiah (1964) have extensive tables for c'.

The rule for selecting the population with the minimum value of λ, is defined by

R': Select π_i iff

$$\sum_{j=1}^n y_{ij} \le b \min_i \left\{ \sum_{j=1}^n y_{ij};\ i = 1,2,\ldots,k \right\}.$$

It follows from the Corollary 3.1.a. that the constant $b=b(k,np,P^*)$ is given by

$$(4.4) \qquad \int_0^\infty \left[1-H_{p'}(\tfrac{y}{b})\right]^{k-1} h_{p'}(y)\, dy = P^*, \qquad p' = np.$$

The values $b' = 1/b$ satisfying (4.4) are tabulated in Gupta and Sobel (1962) for selected values of p' and P^* and more extensively by Krishnaiah and Armitage (1964).

Case 2. Σ_i unknown $(i=1,2,\ldots,k)$.

If Σ_i's are not known, we modify the rules R and R' as follows. Let $z_i = \bar{\underline{x}}_i' S_i^{-1} \bar{\underline{x}}_i$ where $\bar{\underline{x}}_i'$ is the sample mean vector of the ith population and where S_i is the usual sample covariance matrix with $(n-1)$ as the divisor.

R: Select π_i iff

$$c_1 z_i \geq z_{max}$$

where $z_{max} = \max(z_1, z_2, \ldots, z_k)$ and where $c_1 = c_1(k,p,n,P^*)$ is a constant (greater than unity) which satisfies

$$\text{(4.5)} \qquad \int_0^\infty F_{p,n-p}^{k-1}(c_1 x)\, f_{p,n-p}(x)\, dx = P^*$$

where $f_{p,n-p}$ is given by (3.7) with $j=0$, $\mu = p/2$ and $\nu = (n-p)/2$ i.e. it is the density of a random variable which is $\frac{p}{n-p}$-times the central F random variable. $F_{p,n-p}(\cdot)$ is the corresponding cdf. The modified procedure R' is

R': Select π_i iff

$$z_i \leq b_1 z_{min}$$

where $b_1 = b_1(k,p,n,P^*)$ is a constant (greater than unity) determined by

$$\int_0^\infty \left[1-F_{p,n-p}(x/b_1)\right]^{k-1} f_{p,n-p}(x)dx = P^* . \tag{4.6}$$

It should be pointed out that (4.5) and (4.6) are consequences of Corollary 3.1.b and the fact that each $Kz_i (i=1,\ldots,k)$ (K = Constant) has the density (non-central F) given by (3.1) in conjunction with (3.7).

Case 3. $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$ (unknown).

In this case the usual pooled estimator $S = (S_1 + S_2 + \ldots + S_k)/k$ is used in the procedures R and R' of Case 2. In this case the constants c_2 and b_2 are again determined by equations of the type (4.5) and (4.6), respectively, with degrees of freedom p, $k(n-1) - p+1$, respectively.

Remark 1: It should be pointed out that the procedures R and R' discussed under case 1 are not ''strictly analogous'' to those given for cases 2 and 3. If we use procedures based on $\bar{x}_i' \Sigma_i^{-1} \bar{x}_i$ in case 1, the corresponding constants c and b turn out to be independent of the number of observations which is undesirable.

Remark 2: The efficiency of these procedures in terms of expected size or related criteria has not been investigated here. Also the ''indifference zone'' approach, a different type of formulation, due to Bechhofer (1954) has not been discussed here.

REFERENCES

[1] Armitage, J.V. Major, and Krishnaiah, P.R. (1964). Tables for the Studentized largest chi-square distribution and their applications. ARL 64-188, Aerospace Research Laboratories, WP AFB, Ohio.

[2] Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25, 16-29.

[3] Gupta, S.S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. Tokyo 14, 199-216.

[4] Gupta, S.S. and Sobel, M. (1962). On the smallest of several correlated F statistics. Biometrika 49, 509-523.

[5] Gupta, S.S. (1965). On some selection and ranking procedures for multivariate normal populations using distance functions. Mimeograph Series No. 43, Department of Statistics, Purdue University.

[6] Krishnaiah, P.R. and Armitage, J.V. (1964). Distribution of the studentized smallest chi-square with tables and applications. ARL 64-218, Aerospace Research Laboratories, WP AFB, Ohio.

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

On Some Selection and Ranking Procedures with Applications to Multivariate Populations

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, October 1965

5. AUTHOR(S) (Last name, first name, initial)

Gupta, Shanti S. and Studden, William J.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1965	18	6

8a. CONTRACT OR GRANT NO. NONR 1100(26) b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series No. 58 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES

Releasable without limitations on dissemination.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Also supported by Aerospace Research Laboratories Contract AF 33(657)11737	Logistics and Mathematical Statistics Branch, Office of Naval Research Washington, D.C. 20360

13. ABSTRACT

A problem of subset selection for parameters which are not necessarily scale or location parameters is considered. A general theorem dealing with the infimum of the probability of a correct selection for parameters occurring in densities which are Poisson mixtures of arbitrary densitys on $[0,\infty)$ is proved. This theorem is applied to obtain the minimum value of the probability of a correct selection in several cases where multivariate normal populations are ranked according to $\lambda_i = \mu_i' \Sigma_i^{-1} \mu_i$ $(i=1,2,\ldots,k)$ and μ_i is the unknown mean vector and Σ_i (known or unknown) the covariance matrix of the ith p-variate normal population.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Multiple Decisions Selection and Ranking Multivariate Normal Correct Selection Distance Function						

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, rules, and weights is optional.