Minimax Subset Selection with Applications to Unequal Variance Problems*

by

Roger L. Berger and Shanti S. Gupta
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #495

July 1977

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University.

MINIMAX SUBSET SELECTION WITH APPLICATIONS
TO UNEQUAL VARIANCE PROBLEMS

1. INTRODUCTION

Let $X_1,\ldots,X_k$ be observations from populations whose distributions are determined by unknown real parameters $\theta_1,\ldots,\theta_k$. In a subset selection problem, the goal is to select a subset of the populations which includes the population associated with the largest parameter with "high" probability and includes the other populations with "low" probability. In this paper, rules are found which are minimax in the class of non-randomized, just, and translation invariant rules when risk is measured by the maximum probability of including a non-best population. These rules are of the form proposed and studied by Gupta (1965) in location and scale parameter problems. In many cases, these rules are the unique minimax rule in the class and, hence are also admissible in this class. These results are applied to the normal means problem with known unequal variances (or unequal sample sizes). Comparison of several proposed rules is made. A rule proposed by Gupta and Huang (1976) is found to be minimax. A generalization of the rule, proposed by Gupta and Wong (1976), is likewise minimax. Other rules, proposed by Chen and Dudewicz (1973) and Gupta and Huang (1974), are shown to be not minimax.

2. NOTATION AND FORMULATION

Let $\underline{X} = (X_1,\ldots,X_k)$ be a random vector with cumulative distribution function (c.d.f.) $F(\underline{x} - \underline{\theta})$ and density $f(\underline{x} - \underline{\theta})$ with respect to Lebesque measure on $\mathbb{R}^k$. $\underline{\theta} \in \Theta = \mathbb{R}^k$ is the unknown location parameter. Let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ denote the ordered coordinates of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$. This induces the partition $\Theta = \Theta_1 \cup \ldots \cup \Theta_k$ of the parameter space where $\Theta_i = \{\underline{\theta}: \theta_i = \theta_{[k]}\}$. Let $\pi_1,\ldots,\pi_k$ denote the k populations which give rise to observations $X_1,\ldots,X_k$, respectively, and let $\pi_{(i)}$ denote the (unknown) population associated with $\theta_{[i]}$. The goal in a subset selection problem is to find rules which select a subset of the populations which includes the "best" population $\pi_{(k)}$ with "high" probability and includes the "non-best" populations $\pi_{(1)},\ldots,\pi_{(k-1)}$ with "low" probability, regardless of the true parameter value. In general, a selection rule will be denoted by $\varphi(\underline{x}) = (\varphi_1(\underline{x}),\ldots,\varphi_k(\underline{x}))$ where $\varphi_i(\underline{x}): \mathcal{X} \to [0,1]$ is the probability that π_i is included in the selected subset when $\underline{X} = \underline{x}$ is observed.

Selecting a subset which contains the best population is called a correct selection, CS. To insure a "high" probability of making a CS, we will consider only those rules which satisfy the P*-condition, viz.,

$$\inf_{\Theta} P_{\underline{\theta}}(CS|\varphi) \geq P^*, \tag{2.1}$$

where $\frac{1}{k} < P^* < 1$ is a pre-assigned fixed number. The risk function we will use, to reflect the fact that we want the non-best populations to be included with "low" probability, is

$$M(\underline{\theta},\varphi) = \max_{1 \leq i \leq j-1} P_{\underline{\theta}}(\text{select } \pi_{(i)}|\varphi). \tag{2.2}$$

3. JUST AND INVARIANT RULES AND AN ORDERING OF DISTRIBUTIONS

In this section, two classes of selection rules are defined. An ordering of distributions is also introduced. Some preliminary lemmas are presented.

Definition 3.1. A selection rule is just if for every $i = 1,\ldots,k$, $\varphi_i(x_1,\ldots,x_k)$ is a non-decreasing function of x_i and a non-increasing function of x_j, $j \neq i$.

The concept of justness is appealing if the best population is the one associated with the largest parameter value and an increase of a parameter values causes the corresponding observation to be stochastically larger. Location parameters are common examples of this monotonic behavior. Just rules were defined and investigated in more generality by Nagel (1970) and Gupta and Nagel (1971).

Definition 3.2. A selection rule is translation invariant if for every $\underline{x} \in \mathbb{R}^k$, for every $c \in \mathbb{R}$ and for every $i = 1,\ldots,k$, $\varphi_i(x_1+c,\ldots,x_k+c) = \varphi_i(x_1,\ldots,x_k)$.

Since the sets $\Theta_1,\ldots,\Theta_k$ are translation invariant, restriction to translation invariant rules is reasonable. Lemma 3.1 provides a useful characterization of selection rules which are both just and translation invariant.

Lemma 3.1. A selection rule, $\varphi(\underline{x}) = (\varphi_1(\underline{x}),\ldots,\varphi_k(\underline{x}))$, is just and translation invariant if and only if the following two conditions hold:

(i) for every $i = 1,\ldots,k$, φ_i is a function only of the set of differences $\{x_j - x_i:\ j = 1,\ldots,k,\ j \neq i\}$,

(ii) if $\underline{x}$ and $\underline{y}$ satisfy $x_j - x_i \leq y_j - y_i$ for every $j \neq i$, then $\varphi_i(\underline{x}) \geq \varphi_i(\underline{y})$.

Proof. φ is translation invariant if and only if (i) holds because the differences are a maximal invariant for the translation group (see Lehmann (1959) p. 216). Suppose φ is just and translation invariant. Let $\underline{x}$ and $\underline{y}$ be as in (ii). Using first invariance and then justness yields

$$\varphi_i(\underline{x}) = \varphi_i(x_1-x_i+y_i,\ldots,x_k-x_i+y_i)$$
$$= \varphi_i(x_1-x_i+y_i,\ldots,y_i,\ldots,x_k-x_i+y_i)$$
$$\geq \varphi_i(y_1,\ldots,y_i,\ldots,y_k) = \varphi_i(\underline{y})$$

so (ii) is true.

Now suppose (ii) is true. Fix $\underline{x} \in \mathbb{R}^k$, $\varepsilon \geq 0$ and $i \neq j$. Then $x_j + \varepsilon - x_i \geq x_j - x_i$ and all other differences are equal so by (ii), $\varphi_i(x_1,\ldots,x_k) \geq \varphi_i(x_1,\ldots,x_j+\varepsilon,\ldots,x_k)$, i.e., φ_i is non-increasing in x_j, $j \neq i$. Also, $x_j-(x_i+\varepsilon) \leq x_j-x_i$ for every $j \neq i$ so by (ii), $\varphi_i(x_1,\ldots,x_i+\varepsilon,\ldots,x_k) \geq \varphi_i(x_1,\ldots,x_i,\ldots,x_k)$, i.e., φ_i is non-decreasing in x_i. Hence φ is just.||

The following ordering of distributions was introduced by Lehmann (1952) and further investigated by Lehmann (1955). See also Alam (1973).

Definition 3.3. A subset $A \subset \mathbb{R}^k$ is monotone if $\underline{x} \in A$ and $\underline{y}$ satisfies $y_i \leq x_i$ for all $i = 1,\ldots,k$ implies $\underline{y} \in A$.

Definition 3.4. A family of probability distributions on $\mathbb{R}^k$, $\{F_{\underline{\theta}}: \underline{\theta} \in \Theta \subset \mathbb{R}^k\}$, has the stochastic increasing property (SIP) if $\underline{\theta} \in \Theta$, $\underline{\theta}' \in \Theta$, and $\theta_i \leq \theta_i'$ for all $i = 1,\ldots,k$ implies

$$P_{\underline{\theta}}(A) = \int_A dF_{\underline{\theta}} \geq \int_A dF_{\underline{\theta}'} = P_{\underline{\theta}'}(A)$$

for all monotone sets A.

In Lemma 3.1 it was shown that if a selection rule is translation invariant, then the rule is a function only of the differences of the observations. Thus the distribution of this random vector of differences is of interest. Lemma 3.2 shows that this vector has the SIP. Lehmann (1955) showed that if $\underline{\theta}$ is a location parameter, then the family $\{F_{\underline{\theta}}:\ \underline{\theta} \in \Theta\}$ has the SIP. This fact is used in the proof of Lemma 3.2.

Lemma 3.2. Suppose $\underline{\theta} \in \mathbb{R}^k$ is a location parameter in the distribution of $\underline{X} = (X_1,\ldots,X_k)$. Then the distribution of $\underline{X}^* = (X_1-X_i,\ldots,X_{i-1}-X_i, X_{i+1}-X_i,\ldots,X_k-X_i)$ depends on $\underline{\theta}$ only through the parameter $\underline{\theta}^* = (\theta_1-\theta_i,\ldots,\theta_{i-1}-\theta_i,\theta_{i+1}-\theta_i,\ldots,\theta_k-\theta_i)$ and the family of distributions of $\underline{X}^*$ has the SIP in terms of $\underline{\theta}^*$.

Proof. Let $\underline{Y}$ be a random vector with the same distribution as $\underline{X}$ has if $\underline{\theta} = (0,\ldots,0)$. Let G be the c.d.f. of $(Y_1-Y_i,\ldots,Y_{i-1}-Y_i,Y_{i+1}-Y_i,\ldots, Y_k-Y_i)$ and $G_{\underline{\theta}}$ be the c.d.f. of $\underline{X}^*$. Then $(Y_1+\theta_1,\ldots,Y_k+\theta_k)$ has the same distribution as $\underline{X}$ so, for any constants $c_1,\ldots,c_{k-1}$,

$$\begin{aligned}G_{\underline{\theta}}(c_1,\ldots,c_{k-1}) &= P_{\underline{\theta}}(X_1-X_i \le c_1,\ldots,X_k-X_i \le c_{k-1})\\ &= P(Y_1+\theta_1-Y_i-\theta_i \le c_1,\ldots,Y_k+\theta_k-Y_i-\theta_i \le c_{k-1})\\ &= G(c_1-(\theta_1-\theta_i),\ldots,c_{k-1}-(\theta_k-\theta_i)).\end{aligned}$$

So the distribution of X^* depends only on $\underline{\theta}^*$ and in fact $\underline{\theta}^*$ is a location parameter. By Lehmann's result, the family of distributions has the SIP.||

4. MINIMAXITY AND ADMISSIBILITY OF SELECTION RULES

A non-randomized selection rule is one for which $\varphi_i(\underline{x}) \in \{0,1\}$ for all $\underline{x} \in \mathbb{R}^k$ and all i. Thus a non-randomized rule is completely determined by k sets $A_1,\ldots,A_k$ where $A_i = \{\underline{x}:\ \varphi_i(\underline{x}) = 1\}$ is the set of observations for which π_i is included in the selected subset. By Lemma 3.1, a non-randomized rule is just and translation invariant if and only if $\underline{x} \in A_i$

or $\underline{x} \in A_i^c$ (A^c denotes the complement of A) can be determined from only the differences $\{x_j - x_i: \ j \neq i\}$ and A_i is monotone in these differences. In determining a rule which is minimax with respect to M, the quantity to be minimized is

$$\begin{aligned} \sup_{\Theta} M(\underline{\theta},\varphi) &= \sup_{\Theta} \max_{1 \le i \le k-1} P_{\underline{\theta}}(\text{select } \pi_{(i)} \mid \varphi) \\ (4.1) \qquad &= \max_{1 \le i \le k} \sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i \mid \varphi) \\ &= \max_{1 \le i \le k} \sup_{\Theta_i^c} P_{\underline{\theta}}(A_i). \end{aligned}$$

This can be minimized by choosing sets A_i to minimize each of the terms $\sup_{\Theta_i^c} P_{\underline{\theta}}(A_i)$ separately with the restriction

$\inf_{\Theta_i} P_{\underline{\theta}}(A_i) \ge P^*$ so the P*-condition is satisfied. The form of the set which does this minimizing is given by Theorem 4.1 which is an extension of Lehmann's (1952) Theorem 4.1.

Theorem 4.1. Let the joint distribution of $(Y_1,\ldots,Y_k)$ be $F_{\underline{\gamma}}(y_1,\ldots,y_k)$ where the parameter space is the finite or infinite open rectangle $\underline{\gamma}_i < \gamma_i < \bar{\gamma}_i$ and the sample space is the finite or infinite open rectangle $\underline{y}_i < y_i < \bar{y}_i$, independent of the $\underline{\gamma}$. Suppose $P_{\underline{\gamma}}(S)$ is a continuous function of $\underline{\gamma}$ for any set of the form (4.5). Suppose the family $\{F_{\underline{\gamma}}\}$ has the SIP, that the marginal distribution of Y_i depends only on γ_i and that Y_i converges in probability to $\underline{y}_i$ as $\gamma_i \to \underline{\gamma}_i$. Let $\underline{\gamma}^* = (\gamma_1^*,\ldots,\gamma_k^*)$ be a fixed parameter point and define

$$(4.2) \qquad \Gamma = \{\underline{\gamma}: \ \gamma_i \le \gamma_i^*, \ i = 1,\ldots,k\}.$$

Let $\mathcal{S}$ be the collection of all monotone sets which satisfy

$$(4.3) \qquad \inf P\,(S) > P^*.$$

Then a region $S^* \in \mathcal{S}$ which satisfies

(4.4) $$\sup_{\Gamma^c} P_{\underline{\gamma}}(S^*) = \inf_{\mathcal{S}} \sup_{\Gamma^c} P_{\underline{\gamma}}(S)$$

is given by

(4.5) $$S^* = \{\underline{y}: \ y_i \leq a_i, \ i = 1,\dots,k\},$$

where the constants a_i are determined by

(4.6) $$P_{\underline{\gamma}^*}(S^*) = P^*$$

and

(4.7) $$P_{\gamma_1^*}(Y_1 \leq a_1) = P_{\gamma_2^*}(Y_2 \leq a_2) = \dots = P_{\gamma_k^*}(Y_k \leq a_k).$$

Furthermore, if for every i, the distribution of Y_i given γ_i^* has the entire interval $(\underline{y}_i, \bar{y}_i)$ as its support, the region S^* is the essentially unique element of $\mathcal{S}$ which is minimax, i.e., satisfies (4.4).

Proof. For any set of constants $y_j > \underline{y}_j$ and any $i = 1,2,\dots,k$

(4.8) $$\lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j \neq i}} P_{\underline{\gamma}}(Y_1 \leq y_1,\dots,Y_k \leq y_k) = P_{\gamma_i}(Y_i \leq y_i)$$

because

$$P(Y_1 \le y_1, \ldots, Y_k \le y_k) = P(Y_i \le y_i) - P(Y_i \le y_i, Y_j > Y_j \text{ for at least one } j \ne i)$$
$$\ge P(Y_i \le y_i) - \sum_{j \ne i} P(Y_j > y_j)$$

and every term $P(Y_j > y_j)$ converges to zero in the limit of (4.8) because of the convergence in probability. The $\le$ inequality is immediate.

For an $S \in \mathcal{S}$, the SIP implies that $\lim_{\substack{\gamma_j \to \underline{\gamma}_j \\ j \ne i}} P_{\underline{\gamma}}(S)$ exists and the limit will be denoted by $\beta_i(S|\gamma_i)$. The SIP also implies that

(4.9) $$\sup_{\Gamma^c} P_{\underline{\gamma}}(S) \ge \max_{1 \le i \le k} \beta_i(S|\gamma_i^*)$$

and because of the continuity, for sets of the form (4.5)

$$\sup_{\Gamma^c} P_{\underline{\gamma}}(S^*) = \max_{1 \le i \le k} B_i(S^*|\gamma_i^*).$$

Since for the region S^* given by (4.5), (4.7) and (4.8) imply that $\beta_1(S^*|\gamma_1^*) = \beta_2(S^*|\gamma_2^*) = \ldots = \beta_k(S^*|\gamma_k^*)$, if the theorem were false, an $S \in \mathcal{S}$ could be found which simultaneously decreases all k quantities. But this can not happen. For let $S \in \mathcal{S}$. Let $\underline{y} \in S \cap S^{*c}$. (Such a $\underline{y}$ exists unless S is essentially the same as S^* because of (4.3) and (4.6).) For some $i = 1,2,\ldots,k$, $y_i > a_i$ since $\underline{y} \in S^{*c}$

(4.10) $$P(S^* \cap S^c) \le P\Big(\bigcup_{j \ne i} \{Y_i \le a_i, Y_j > y_j\}\Big)$$

$$\leq \sum_{j\neq i} P(Y_i \leq a_i,\ Y_j > y_j)$$

$$\leq \sum_{j\neq i} P(Y_j > y_j).$$

As $\gamma_j \to \underline{\gamma}_j$, all the terms $P_{\gamma_j}(Y_j > y_j) \to 0$.

$$\beta_i(S|\gamma_i^*) = \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S|\gamma_1,\ldots,\gamma_i^*,\ldots,\gamma_k)$$

(4.11)

$$= \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S^*) + \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S \cap S^{*c}) - \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S^* \cap S^c)$$

$$= \beta_i(S^*|\gamma_i^*) + \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S \cap S^{*c}) - \lim_{\substack{\gamma_j\to\underline{\gamma}_j\\ j\neq i}} P(S^* \cap S^c).$$

From (4.10) the last limit is zero, so $\beta_i(S|\gamma_i^*) \geq \beta_i(S^*|\gamma_i^*)$ and the first part of the theorem is proven.

Furthermore,

$$P(S \cap S^{*c}) \geq P(Y_1 \leq y_1,\ldots,Y_i \leq y_i,\ldots,Y_k \leq y_k)$$
$$- P(Y_1 \leq y_1,\ldots,Y_i \leq a_i,\ldots,Y_k \leq y_k).$$

As $\gamma_j \to \underline{\gamma}_j$, $j \neq i$, by (4.8) the right hand side converges to

(4.12) $$P_{\gamma_i^*}(Y_i \leq y_i) - P_{\gamma_i^*}(Y_i \leq a_i).$$

So if the support of the distribution of Y_i given γ_i^* is the entire interval

$(\underline{y}_i, \bar{y}_i)$, (4.12) is greater than zero and by (4.11), $\beta_i(S|\gamma_i^*) > \beta_i(S^*|\gamma_i^*)$. Hence by (4.9)

$$\sup_{\Gamma^c} P_{\underline{\gamma}}(S) > \sup_{\Gamma^c} P_{\underline{\gamma}}(S^*) \tag{4.13}$$

and S^* is the essentially unique element of $\mathcal{S}$ which is minimax.||

The form of the region S^* in Theorem 4.1 becomes particularly simple if the joint distribution of $(Y_1,\ldots,Y_k)$ is symmetric (i.e., the random variables are exchangeable) given $\underline{\gamma}^*$. Then (4.7) implies $a_1=\ldots=a_k=a$ where a is determined by (4.6) and the minimax region is

$$S^* = \{\underline{y}: \max_{1\leq i\leq k} y_i \leq a\}. \tag{4.14}$$

The following selection rule has been proposed and studied by Gupta (1965).

<u>Definition 4.1</u>. Define the selection rule R_1 by

$$R_1: \text{ select } \pi_i \text{ if } x_i \geq \max_{1\leq j\leq k} x_k - d$$

where d is chosen to be the smallest positive constant such that the P*-condition (2.1) is satisfied.

Theorem 4.1 can be used to show that R_1 is minimax and admissible with respect to M in a restricted class of rules.

<u>Theorem 4.2</u>. Let $\underline{X} = (X_1,\ldots,X_k)$ have a density $f(\underline{x} - \underline{\theta})$, $\underline{\theta} \in \mathbb{R}^k$ with respect to Lebesque measure μ on $\mathbb{R}^k$. Suppose the support of f is $\mathbb{R}^k$ and

f is symmetric (i.e., the random variables are exchangeable if $\theta_1 = \ldots = \theta_k$). Then R_1 is minimax with respect to M in the class of non-randomized, just, and translation invariant rules which satisfy the P*-condition. Furthermore R_1 is the unique minimax rules in this class so R_1 is admissible in this class.

<u>Proof</u>. Fix $i = 1,\ldots,k$. Let $Y_1 = X_1 - X_i,\ldots,Y_{k-1} = X_k - X_i$ (omitting $X_i - X_i$) and $\gamma_1 = \theta_1 - \theta_i,\ldots, \gamma_{k-1} = \theta_k - \theta_i$ (omitting $\theta_i - \theta_i$). As explained at the beginning of this section, by Lemma 3.1 a rule is non-randomized, just, and translation invariant if and only if it is of the form

$$\varphi_i(\underline{x}) = \begin{cases} 1 & \text{if } \underline{y} \in S_i \\ 0 & \underline{y} \in S_i^c \end{cases}$$

where S_i is a monotone subset of $\mathbb{R}^{k-1}$. By Lemma 3.2, the distribution of $\underline{Y}$ depends only on $\underline{\gamma}$ and since $\underline{X}$ has a density with respect to Lebesque measure on $\mathbb{R}^k$, $\underline{Y}$ has a density with respect to Lebesque measure on $\mathbb{R}^{k-1}$. This implies $P_{\underline{\gamma}}(S^*)$ is a continuous function of $\underline{\gamma}$ for sets S^* of the form (4.5) since for such sets, $\overline{S^*}/S^{*\circ}$ (closure minus interior) has Lebesque measure zero. Lemma 3.2 establishes the SIP of $\{F_{\underline{\gamma}}(\underline{y}): \underline{\gamma} \in \mathbb{R}^{k-1}\}$. Also γ_j is a location parameter in the marginal distribution of Y_j so the convergence in probability assumption of Theorem 4.1 is true.

Let $\underline{\gamma}^* = (0,\ldots,0)$ so the set Γ of Theorem 4.1 is equivalent to

$$\overline{\Theta_i} = \{\underline{\theta}: \ \theta_i \geq \theta_j, \ j = 1,\ldots,k\}$$

($\bar{A}$ denotes the closure of A). Because of the continuity of $P_{\underline{\gamma}}(S)$ in terms of $\underline{\gamma}$, the fact that Γ is $\overline{\Theta_i}$ rather than Θ_i is unimportant since the sup's and inf's are all the same taken over a set or its closure. (4.3) simply insures the P*-condition.

Because f is symmetric, the distribution of $\underline{Y}$ given $\underline{\gamma}^*$ is also symmetric so the remark following Theorem 4.1 is relevant and the monotone set S_i^* which minimizes

$$\sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i) = \sup_{\Gamma^c} P_{\underline{\gamma}}(S) \tag{4.15}$$

is given by Theorem 4.1 as

$$\begin{aligned} S_i^* &= \{\underline{y}:\ \ y_j \le d_i \quad j = 1,\ldots,k-1\} \\ &= \{x:\ \ x_j - x_i \le d_i \quad j \ne i\} \end{aligned} \tag{4.16}$$

which is the acceptance region for π_i of R_1.

Since the support of f is $\mathbb{R}^k$, the support of the distribution of Y_j given γ_j^* is $\mathbb{R}$. So S_i^* is the unique acceptance region for π_i which minimizes (4.15). Because of the exchangeability, $d_1 = \ldots = d_k = d$ and

$$\sup_{\Theta_1^c} P_{\underline{\theta}}(S_1^*) = \ldots = \sup_{\Theta_k^c} P_{\underline{\theta}}(S_k^*) = P_{\theta_1=\theta_2}(X_2 - X_1 \le d).$$

So none of the S_i^* can be changed without increasing (4.1). Thus $\{S_1^*,\ldots,S_k^*\}$ is the unique set of acceptance regions which minimizes (4.1) (i.e., R_1 is the unique minimax selection rule). Any unique minimax rule is admissible.||

5. SELECTION OF NORMAL MEANS WHEN THE VARIANCES ARE UNEQUAL

In this section we assume that $X_1,\ldots,X_k$ are independent random variables and X_i is normally distributed with unknown mean θ_i and known variances σ_i^2. If the σ_i^2's were assumed to be equal, we would have exchangeable random variables and by Theorem 4.2 R_1 would be a "good" selection rule. But here, no assumptions about the equality of the variances are made. The variances may be of the form $\sigma_i^2 = \frac{\gamma_i^2}{n_i}$, e.g., X_i is the mean of a random sample of size n_i from π_i. So this formulation includes unequal sample size problems.

The following five rules have been proposed. In each case d is chosen to be the smallest positive constant such that the P*-condition is satisfied. If γ rather than γ_i appears, it is assumed that $\gamma_1=\ldots=\gamma_k=\gamma$. Chen and Dudewicz (1973) proposed rules R_2 and R_3.

$$R_2:\ \text{select } \pi_i \text{ iff } x_i \geq \max_{1\leq i\leq k} x_j - d\gamma\sqrt{\frac{1}{n_i}+\frac{1}{n_{[k]}}}$$

where $n_{[k]} = \max_{1\leq j\leq k} n_j$.

$$R_3:\ \text{select } \pi_i \text{ iff } x_i \geq \max_{1\leq j\leq k} x_j - d\gamma\sqrt{\frac{1}{n_i}+\frac{1}{n_{[1]}}}$$

where $n_{[1]} = \min_{1\leq j\leq k} n_j$.

Gupta and Huang (1974) proposed R_4.

$$R_4:\ \text{select } \pi_i \text{ iff } x_i \geq \max_{1\leq j\leq k} x_j - \frac{d\gamma}{\sqrt{n_i}}\ .$$

Gupta and Huang (1976) proposed R_5.

$$R_5\text{: select } \pi_i \text{ iff } x_i \geq \max_{1\leq j\leq k} \left(x_j - d\gamma\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right)$$

Gupta and Wong (1976) proposed R_6.

$$R_6\text{: select } \pi_i \text{ iff } x_i \geq \max_{1\leq j\leq k} \left(x_j - d\sqrt{\frac{\gamma_i}{n_i} + \frac{\gamma_j}{n_j}}\right)$$

In Chen and Dudewicz (1973), γ was not assumed known so an estimate was used in place of γ in the definitions of R_2 and R_3. R_5 is easily seen to be a specialization of R_6 for the case $\gamma_1 = \ldots = \gamma_k$. It is easy to see that all the rules are just and translation invariant. All reduce to R_1 when the variances and sample sizes are all equal. The following theorem provides a minimax result for R_6.

<u>Theorem 5.1</u>. For the normal means problem, R_6 is minimax with respect to M in the class of non-randomized, just, and translation invariant rules which satisfy the P*-condition.

<u>Proof</u>. By Theorem 4.1, applied as in Theorem 4.2, the just and translation invariant acceptance region S_i for π_i which minimizes $\sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i)$ is given by

$$(5.1) \qquad S_i = \{\underline{x}\text{: } x_j - x_i \leq d_{ij} \quad j = 1,\ldots,k,\ j \neq i\}$$

where (4.7) implies that the d_{ij} satisfy

$$(5.2) \qquad P_{\theta_1=\theta_i}(X_1 - X_i \leq d_{i1}) = \ldots = P_{\theta_k=\theta_i}(X_k - X_i \leq d_{ik}).$$

This implies

$$(5.3)\qquad \Phi(d_{i1}/\sqrt{\sigma_1^2+\sigma_i^2}\,) = \ldots = \Phi(d_{ik}/\sqrt{\sigma_k^2+\sigma_i^2}\,)$$

where Φ is the standard normal c.d.f. Thus, the d_{ij} must satisfy

$$(5.4)\qquad d_{i1}/\sqrt{\sigma_1^2+\sigma_i^2} = \ldots = d_{ik}/\sqrt{\sigma_k^2+\sigma_i^2}\,.$$

Letting $d_i^* = d_{i1}/\sqrt{\sigma_1^2+\sigma_i^2}$, we obtain

$$(5.5)\qquad d_{ij} = d_i^*\sqrt{\sigma_j^2+\sigma_i^2} \qquad j = 1,\ldots,k,\ j \neq i$$

for the minimax region and from the proof of Theorem 4.1 it can be seen that for this minimax region

$$(5.6)\qquad \sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i) = \Phi(d_{i1}/\sqrt{\sigma_1^2+\sigma_i^2}) = \Phi(d_i^*).$$

Recall that, from (4.6), d_i^* is determined by

$$(5.7)\qquad P_{\theta_1=\ldots=\theta_k}(X_j - X_i \leq d_i^*\sqrt{\sigma_j^2+\sigma_i^2}\,,\ j = 1,\ldots,k,\ j \neq i) = P^*$$

Comparing (5.1) (inserting (5.5)) and the definition of R_6, we see that they are the same except that in (5.1), d_i^* depends on i, whereas d in R_6 does not. To minimize (4.1), we must minimize the maximum of the k quantities in (5.6) which, of course, is $\Phi(d^*_{[k]})$ where $d^*_{[k]} = \max_{1\leq j\leq k} d_j^*$. If $d^*_{[k]}$ is used in place of d_i^* for all i, then

$$\sup_{\Theta_i^c} P_{\underline{\theta}}(\text{select } \pi_i) = \Phi(d^*_{[k]}) \qquad \text{for all } i = 1,\ldots,k$$

so (4.1) is unchanged. The rule usingthe d_i^*'s obviously is minimax since it minimizes each of the terms in the max of (4.1). The rule using $d^*_{[k]}$ in place of all d_i^*'s has the same value of (4.1) so it too is minimax. But this rule is R_6 (set $d = d^*_{[k]}$). It should be noted that in going from the rule with the d_i^*'s to R_6, the P*-condition has not been violated. Those acceptance regions which have changed have been increased in size so P(CS) has, if anything, increased.||

The rules R_2, R_3 and R_4 are not minimax. The uniqueness part of Theorem 4.1 is applicable since $X_i - X_j$, $j \neq i$, has support R. They all have regions of the shape (4.5) but do not satisfy (4.7). For example, for R_4, (4.7) implies

$$\Phi\left(d\sqrt{\frac{1}{n_i}} \Big/ \sqrt{\frac{1}{n_i} + \frac{1}{n_1}}\right) = \ldots = \Phi\left(d\sqrt{\frac{1}{n_i}} \Big/ \sqrt{\frac{1}{n_i} + \frac{1}{n_k}}\right).$$

Unless $n_1 = \ldots = n_k$, this is not true. Similar reasoning holds for R_2 and R_3. Of course R_5 is minimax in the $\gamma_1 = \ldots = \gamma_k$ case since it is the same as R_6.

6. SCALE PARAMETERS

Results analogous to those of Section 4 may be obtained when $\underline{\theta}$ is a scale parameter. We assume $\underline{X}$ has c.d.f. $F(x_1/\theta_1,\ldots,x_k/\theta_k)$ and density $f(x_1/\theta_1,\ldots,x_k/\theta_k)/\prod_{i=1}^{k}\theta_i$ with respect to Lebesque measure on $(0,\infty)^k$, $\underline{\theta} \in \Theta = (0,\infty)^k$. In scale problems, it is natural to restrict attention to scale invariant rules, i.e., rules satisfying $\varphi(x_1,\ldots,x_k) = \varphi(cx_1,\ldots,cx_k)$. Replacing translation by scale invariance, lemmas like Lemmas 3.1 and 3.2 can be obtained with the differences (of both observations and parameters) replaced by quotients, e.g., $X_1 - X_i$ becomes X_1/X_i.

Theorem 4.1 is applicable exactly as stated to obtain the following. Define the rule

$$R_7: \text{ select } \pi_i \text{ iff } x_i \geq c \cdot \max_{1 \leq j \leq k} x_j$$

where $0 < c < 1$ is the largest constant such that the P*-condition is satisfied.

Theorem 6.1. Let $\underline{X}$ have density $f(x_1/\theta_1, \ldots, x_k/\theta_k) / \prod_{i=1}^{k} \theta_i$ with respect to Lebesque measure on $(0,\infty)^k$. Suppose that the support of f is $(0,\infty)^k$ and f is symmetric. Then R_7 is minimax with respect to M in the class of non-randomized, just, and scale invariant rules which satisfy the P*-condition. Furthermore R_7 is the unique minimax rule in this class so R_7 is admissible in this class.

Analogous results can also be obtained if the best population is the one associated with the smallest parameter value. In this case the rules

$$R_8: \text{ select } \pi_i \text{ iff } x_i \leq \min_{1 \leq j \leq k} x_j + d$$

and

$$R_9: \text{ select } \pi_i \text{ iff } x_i \leq c \cdot \min_{1 \leq j \leq k} x_j$$

are the minimax, admissible rules in the location and scale problems, respectively.

REFERENCES

Alam, K. (1973): On a multiple decision rule. Ann. Statist. 1, 750-755.

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

Gupta, S. S. (1965): On some multiple decision (selection and ranking) rules. Technometrics, 7, 225-245.

Gupta, S. S. and Huang, D. Y. (1976): Subset selection procedures for the means and variances of normal popualtions: Unequal sample sizes case. Sankhya. Ser. B, 38.

Gupta, S. S. and Huang, W. T. (1974): A note on selecting a subset of normal populations with unequal sample sizes. Sankhya, Ser. A. 36. 389-396.

Gupta, S. S. and Nagel, K. (1971): On some contributions to multiple decision theory. Statistical Decision Theory and Related Topics (ed. Gupta, S. S. and Yackel, J.). 79-102. Academic Press, New York.

Gupta, S. S. and Wong, W. Y. (1976): Subset selection procedures for the means of normal populations with unequal variances: Unequal sample sizes case. Dept. of Statistics, Purdue University, Tech. Report. #473.

Lehmann, E. L. (1952): Testing multiparameter hypotheses. Ann. Math. Statist., 23, 541-552.

_____ (1955): Ordered families of distributions. Ann. Math. Statist., 26, 399-419.

_____ (1959): Testing Statistical Hypotheses. John Wiley and Sons, New York.

Nagel, K. (1970): On subset selection rules with certain optimality properties. Ph.D. Thesis, Dept. of Statistics, Purdue University, West Lafayette, IN

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #495	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Minimax Subset Selection with Applications to Unequal Variance Problems		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #495
7. AUTHOR(s) Roger L. Berger and Shanti S. Gupta		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Purdue University West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE July 1977
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distriubtion unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Subset selection rules, minimaxity, admissibility, probability of selecting non-best populations, just, translation invariant, unequal variances.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $X_1,\ldots,X_k$ be observations from populations whose distributions are determined by unknown real parameters $\theta_1,\ldots,\theta_k$. In a subset selection problem, the goal is to select a subset of the populations which includes the population associated with the largest parameter with "high" probability and includes the other populations with "low" probability. In this paper, rules are found which are minimax in the class of non-randomized, just, and translation invariant

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassifed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

rules when risk is measured by the maximum probability of including a non-best population. These rules are of the form proposed and studied by Gupta (1956) in location and scale parameter problems. In many cases, these rules are the unique minimax rule in the class and, hence are also admissible in this class. These results are applied to the normal mean problem with known unequal variances (or unequal sample sizes). Comparison of several proposed rules is made. A rule proposed by Gupta and Huang (1976) is found to be minimax. A generalization of the rule, proposed by Gupta and Wong (1976), is likewise minimax. Other rules, proposed by Chen and Dudewicz (1973) and Gupta and Huang (1974), are shown to be not minimax.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)