49 Master Copy

On the Properties of Subset Selection Procedures*

by

John J. Deely, Purdue University and Sandia Laboratory

and

Shanti S. Gupta, Purdue University

Department of Statistics

Division of Mathematical Statistics

Mimeograph Series No. 49

August 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 the Properties of Subset Selection Procedures*

by

John J. Deely, Purdue University and Sandia Laboratory

and

Shanti S. Gupta, Purdue University

1. Introduction and Summary

Suppose $\pi_1, \pi_2, \ldots, \pi_k$ are k normal populations with unknown means $\mu_1, \mu_2, \ldots, \mu_k$ and a common known variance which we assume to be unity. In many situations the experimenter is interested in selecting a subset of the populations which contains the best population where by best we mean the population with the largest value $\mu_{[k]}$ of the unknown means, $\mu_1, \mu_2, \ldots, \mu_k$. For further amplification of this type of selection, reference could be made to Gupta (1965) and to references therein. A selection procedure, denoted by R, for this problem along with some of its properties has been discussed by Gupta (1956, 1965). This selection procedure is a member of the class $\mathcal{C}$ of procedures earlier studied by Seal (1955). The selection rule is such that it selects a non-empty subset of random size and under a specified loss function, the associated risk is always bounded above by $\alpha = 1-P*$ where α is a small preassigned number $(0 < \alpha < 1/k)$.

In this paper we study some desirable properties of the above selection procedure and make some comparisons with the ''approximate'' optimal rule $\bar{D}$ of Seal (1955). In particular it is shown that the rule R is minimax and

*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 purposes of the United States Government.

that under the slippage configuration of means, the expected size of the selected subset using R is smaller than that corresponding to $\bar{D}$; as well as being smallest for a subclass $\mathcal{C}'$ of $\mathcal{C}$. ~~and that the probability of a correct selection using R is strictly greater than that of D.~~ The subset selection problem is also considered from a Bayesian viewpoint, and under a linear loss function, the Bayes rule for selecting a subset is derived. This latter result is given in a more general form by Deely (1965) than that given by Dunnett (1960).

2. The Class $\mathcal{C}$ and $\mathcal{C}'$ of rules and the rule R

Let $x_1, x_2, \ldots, x_k$ be the observed values of the sample means each based on n independent observations. Let the ordered sample means be denoted by

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

Let $\underline{c}' = (c_1, c_2, \ldots, c_{k-1})$ be a vector whose components are arbitrary real numbers such that $c_i \geq 0$, for all i, and $\sum_{i=1}^{k-1} c_i = 1$. Then the class consists of rules $D = D_{\underline{c}} = D(c_1, c_2, \ldots, c_{k-1})$ with D given as follows:

''Select the population $\pi_{[i]}$ corresponding to the observed sample mean $x_{[i]}$ iff

$$x_{[i]} \geq c_1 x_{[1]} + \cdots + c_{i-1} x_{[i-1]} + c_i x_{[i+1]} + \cdots + c_{k-1} x_{[k]} - \frac{t(P^*, \underline{c}')}{\sqrt{n}} \tag{2.2}$$

where $t(P^*, \underline{c}')$ is chosen so as to satisfy the basic probability requirement, namely,

$$\inf_{\Omega} P\{CS|D_{\underline{c}}\} = P^*, \tag{2.3}$$

Ω being the parameter space of $\underline{\mu}' = (\mu_1', \ldots, \mu_k)$ and CS standing for correct selection. A correct selection is the selection of any subset which includes the population associated with $\mu_{[k]}$.

To define the subclass $\mathcal{C}'$ of $\mathcal{C}$, we impose the following restrictions on the vector $\underline{c}$: $c_j = 1$ for some $j = 1,2,\ldots,k-1$.

An important rule in the class $\mathcal{C}$ which has been studied in great detail is a member of the class $\mathcal{C}'$ defined above for $\underline{c}' = (0,0,\ldots,0,1)$; this rule denoted in earlier papers by R is:

Select the population corresponding to the observed sample mean $x_{[i]}$ iff

$$x_{[i]} \geqq x_{[k]} - d/\sqrt{n}, \tag{2.4}$$

where $d = d(k,P^*) = t(P^*,\underline{c}')$ again chosen so as to satisfy the basic probability requirement.

It may be pointed out that for the zero-one loss function

$$L(S_j, \underline{\mu}') = \begin{cases} 0 & \text{if } \pi_{(k)} \text{ with } \mu_{[k]} \in S_j\text{, the selected subset, } j = 1,2,\ldots,2^k - 1. \\ 1 & \text{otherwise,} \end{cases} \tag{2.5}$$

the risk is given by

$$\text{Risk} = r(D_{\underline{c}}, \underline{\mu}') = 1 - P\{CS|D_{\underline{c}}\}. \tag{2.6}$$

Hence, the basic probability requirement (2.3) is equivalent to requiring that

$$\sup_{\Omega} r(D_{\underline{c}}, \underline{\mu}) \leq 1-P^*. \tag{2.7}$$

The rule $\bar{D}$ given by Seal (1955) is also a member of $\mathcal{C}$ and is defined by the vector $\underline{c}'$ with $c_j = \frac{1}{k-1}$, $j=1,2,\ldots,k-1$. This rule is:

Select the population corresponding to the observed sample mean x_i iff

$$x_i \geq \tilde{x} - \frac{t(P^*,\underline{c}')}{\sqrt{n}} \qquad \text{where} \tag{2.8}$$

$\bar{x} = \frac{1}{k-1} \sum_{\substack{j=1 \\ j\neq i}}^{j=k} x_j$ and $t(P^*, \underline{c}')$ as stated earlier.

3. <u>Some Results on Minimaxity and Expected Size</u>

From (2.6) and the definition of a minimax rule we see that R is minimax in $\mathcal{C}$ provided

$$\min_{\underline{\mu}\in\Omega} P\{CS|R\} \geq \min_{\underline{\mu}\in\Omega} P\{CS|D\} \tag{3.1}$$

for every $D \in \mathcal{C}$. But recalling the definition of the class $\mathcal{C}$, it can be observed that the quantity $t(P^*,\underline{c}')$ is chosen so as to guarantee

$$\min_{\underline{\mu}\in\Omega} P\{CS|D\} = P^* \quad \text{for any rule } D \text{ in } \mathcal{C}. \tag{3.2}$$

Hence every rule in $\mathcal{C}$ and in particular R is minimax.

It may be pointed out that if we choose $\underline{c}' = (1,0,0,\ldots,0)$, the rule D_0 defined by this choice of $\underline{c}'$ selects all populations with probability 1 and

$$P\{CS|D_0\} = 1 > P\{CS|R\} . \tag{3.3}$$

But the rule D_0 is quite trivial and not very practical from the experimenters viewpoint for the following reason.

omit

The selection procedures discussed here select a subset whose size is a random variable taking values in the set of integers $1,2,\ldots,k$. Thus in comparing decision procedures of this type it is meaningful to use the expected value of this random variable, to be denoted by $E[S|D_{\underline{c}}]$. For the rule R ~~it is immediate that $E[S|D_0] = k$ whereas~~ Gupta (1965) showed that $E[S|R] \leq kP^*$.

More generally, for any rule $D_{\underline{c}} \in \mathcal{C}'$, we have from Seal (1955),

$$\inf_{\Omega} P\{CS|D_{\underline{c}}\} = \inf_{\underline{\mu}'=(\mu,\mu,\ldots,\mu)} P\{CS|D_{\underline{c}}\} \tag{3.3}$$

and hence the solution for the constants $t(P^*,\underline{c}')$, earlier defined by (2.2) and (2.3), can be obtained from equations of the type

$$P\{Y \leq t(P^*,\underline{c}')\} = P^* \tag{3.4}$$

where the random variable Y traverses the real line.

Thus for any rule $D_{\underline{c}} \in \mathcal{C}'$, $t(P^*,\underline{c}')$ will be non-negative provided P^* is sufficiently large; for example $P^* > (1/k)$ for the rule R. Hence every rule $D_{\underline{c}} \in \mathcal{C}'$ (excluding the rule R) will select a subset consisting of at least 2 populations with probability 1. This implies that for any such $D_{\underline{c}}$

$$E[S|D_{\underline{c}}] \geq 2 \tag{3.5}$$

Gupta (1965) showed that in a subset $\Omega(\delta)$ defined by $\mu_{[k]} = \mu + \delta$ $(\delta > 0)$, $\mu_{[i]} \leq \mu$, $i = 1,2,\ldots,k-1$, $E[S|R]$ assumes its maximum over $\Omega(\delta)$ when $\mu_{[i]} = \mu$, $i=1,2,\ldots,k-1$ and

$$\max_{\Omega(\delta)} E[S|R] = \int_{-\infty}^{\infty} \Phi^{k-1}(z+d+\delta\sqrt{n})\,\varphi(z)dz + (k-1)\int_{-\infty}^{\infty} \Phi^{k-2}(z+d)\,\Phi(z+d-\delta\sqrt{n})\,\varphi(z)dz. \tag{3.6}$$

From (3.6) it follows that there exists a δ_0 such that

$$\max_{\Omega(\delta)} E[S|R] < 2 \quad \text{for } \delta > \delta_0 . \tag{3.7}$$

Therefore, if P^* and δ are sufficiently large, it can be seen from (3.5) and (3.7) that

$$E[S|R] < E[S|D_{\underline{c}}] \tag{3.8}$$

for any $\underline{\mu}' \in \Omega(\delta)$.

So in this case good justification can be made for using the decision procedure R. Namely, the P{CS|R} is larger than P*, a preassigned number; and its E[S|R] is smallest among rules in $\mathcal{C}'$ under the restrictions in (3.8).

A table giving the values of the expected proportion of the populations in the selected subset for the slippage configuration is given below.

Table

This table gives the expected proportion of k normal populations retained in the selected subset by the suggested procedure R under the configuration that one of the populations has its mean greater than all the others by δ times the common known standard deviation of the sample means (selected values of k, P* and $\delta\sqrt{n}$ are considered).

P*	$\delta\sqrt{n}$ \ k	2	3	5	10	25	50
.75	.00	.7500	.7500	.7500	.7500	.7500	.7500
	.10	.7495	.7494	.7495	.7497	.7498	.7497
	.25	.7467	.7463	.7469	.7479	.7489	.7492
	.50	.7370	.7349	.7369	.7410	.7451	.7469
	1.00	.7017	.6904	.6947	.7091	.7260	.7346
	2.00	.6057	.5484	.5365	.5638	.6166	.6519
	3.00	.5357	.4203	.3598	.3576	.4121	.4650
.90	.00	.9000	.9000	.9000	.9000	.9000	.9000
	.10	.8994	.8995	.8996	.8998	.8998	.8998
	.25	.8965	.8968	.8976	.8985	.8992	.8995
	.50	.8861	.8872	.8901	.8937	.8968	.8981
	1.00	.8469	.8480	.8574	.8710	.8840	.8901
	2.00	.7219	.7023	.7151	.7527	.8015	.8302
	3.00	.6001	.5271	.5066	.5370	.6090	.6643
.95	.00	.9500	.9500	.9500	.9500	.9500	.9500
	.10	.9496	.9496	.9497	.9499	.9500	.9500
	.25	.9474	.9478	.9484	.9490	.9496	.9498
	.50	.9394	.9409	.9432	.9459	.9480	.9489
	1.00	.9082	.9120	.9202	.9305	.9396	.9437
	2.00	.7950	.7902	.8079	.8414	.8801	.9014
	3.00	.6584	.6114	.6095	.6490	.7185	.7670

$\delta\sqrt{n}$ = .00(.25) 3.00, 3.50, 4.00, 5.00, 6.00, 8.00, 10.00
d depends on $P^* = 1-\alpha$
n = k - 1

d = √2 (the entry in the published ...)

4. Comparison of $E[S|\bar{D}]$ and $E[S|R]$

It can be seen that the expected size of the selected subset using any rule D is given by:

(4.1) $$E[S|D] = \sum_{j=1}^{k} P\{\text{selecting the population with mean } \mu_{[j]} | D\}$$

~~[illegible]~~

~~Probability of Correct Selection~~

So we shall first derive the expression for the probability of selecting the population $\pi_{(j)}$ with mean $\mu_{[j]}$ for the rule $\bar{D}$. (Recall $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ are the ranked values of the mean vector $\underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_k)$.) Let $X_{(j)}$ be the sample mean (unknown) which is associated with the population $\pi_{(j)}$. Then

(4.2) $$P\{\pi_{(j)} \text{ is selected} | \bar{D}\} = P\{X_{(j)} \geq \frac{1}{k-1} \sum_{\substack{i=1 \\ i \neq j}}^{i=k} X_{(i)} - \frac{t(P^*, \underline{c}')}{\sqrt{n}}\}$$

where $\underline{c}' = (\frac{1}{k-1}, \ldots, \frac{1}{k-1})$ for the rule $\bar{D}$. Equation (4.2) reduces to

(4.3) $$P\{\pi_{(j)} \text{ is selected} | \bar{D}\} = \Phi\left(\frac{t(P^*, \underline{c}') + \sqrt{n}\left(\mu_{[j]} - \frac{1}{k-1} \sum_{\substack{i=1 \\ i \neq j}}^{i=k} \mu_{[i]}\right)}{(\frac{k}{k-1})^{1/2}}\right),$$

φ and Φ being the standard normal density and cumulative distribution functions respectively. In particular from (4.3) we obtain

(4.4) $$P\{CS|\bar{D}\} = P\{\pi_{(k)} \text{ is selected} | \bar{D}\} = \Phi\left(\frac{t(P^*, \underline{c}') + \sqrt{n}\,(\mu_{[k]} - \frac{1}{k-1} \sum_{i=1}^{i=k-1} \mu_{[i]})}{(\frac{k}{k-1})^{1/2}}\right)$$

From (2.3) we see that the constant $t(P^*,\underline{c}')$ to carry out the procedure can be obtained by using (4.4) and solving

(4.5) $$\Phi\left(\frac{t(P^*,\underline{c}')}{(\frac{k}{k-1})^{1/2}}\right) = P^* \quad \text{for} \quad t(P^*,\underline{c}')$$

which gives

(4.6) $$t(P^*,\underline{c}') = (\frac{k}{k-1})^{1/2}\,\gamma; \text{ letting } \gamma = \Phi^{-1}(P^*).$$

(go to top of p. 11)

Then using (4.5) in (4.3) we obtain

(4.6) $$P\{CS|\bar{D}\} = \Phi\left(\gamma+(\frac{n}{k(k-1)})^{1/2}\sum_{i=1}^{k-1}(\mu_{[k]}-\mu_{[i]})\right)$$

From Gupta (1965) we have

(4.7) $$P\{CS|R\} = \int_{-\infty}^{\infty}\prod_{j=1}^{k-1}\Phi(z + d + \sqrt{n}\,(\mu_{[k]}-\mu_{[j]})\,\varphi(z)dz.$$

Let $\Omega' \subset \Omega$ be the space of mean vectors $\underline{\mu}'$ such that $\underline{\mu}'=(\mu,\mu,\ldots,\mu,\mu+\delta)$ where $-\infty < \mu < \infty$ and $\delta \geq 0$. Then for $\underline{\mu}' \in \Omega'$ we have

(4.6)' $$P\{CS|\bar{D}\}= \Phi(\gamma + (\frac{k-1}{k})^{1/2}\sqrt{n}\,\delta)$$

and

(4.7)' $$P\{CS|R\} = \int_{-\infty}^{\infty}\Phi^{k-1}(z + d + \sqrt{n}\,\delta)\,\varphi(z)dz.$$

omit

~~the expected size of the selected subset~~
Then using (4.1), (4.3), and (4.6) we have

$$(4.7)\qquad E[S|\bar{D}] = \sum_{j=1}^{k} \Phi\Big(\gamma + \big(\frac{n}{k(k-1)}\big)^{1/2} \sum_{\substack{i=1\\ i\neq j}}^{i=k} (\mu_{[j]} - \mu_{[i]})\Big),$$

Now from Gupta (1965) we have the expression for the expected size of the selected subset using the rule R

$$(4.8)\qquad E[S|R] = \sum_{j=1}^{k} \int_{-\infty}^{\infty} \prod_{\substack{i=1\\ i=j}}^{i=k} \Phi\big(z+d+(\mu_{[j]} - \mu_{[i]})\sqrt{n}\big)\, \varphi(z)dz.$$

Let $\Omega' \subset \Omega$ be the space of all mean vectors $\underline{\mu}'$ such that $\underline{\mu}' = (\mu,\mu,\ldots,\mu,\mu+\delta)$ where $-\infty < \mu < \infty$ and $\delta \geq 0$. Then for $\underline{\mu}' \in \Omega'$, we can write

$$(4.7)'\qquad E[S|\bar{D}] = \sum_{j=1}^{k-1} \Phi\Big(\gamma - \big(\frac{n}{k(k-1)}\big)^{1/2}\delta\Big) + \Phi\Big(\gamma + \big(\frac{k-1}{k}\big)^{1/2} \delta\sqrt{n}\Big)$$

$$= (k-1)\, \Phi\Big(\gamma - \big(\frac{n}{k(k-1)}\big)^{1/2} \delta\Big) + \Phi\Big(\gamma + \big(\frac{k-1}{k}\big)^{1/2} \delta\sqrt{n}\Big)$$

and

$$(4.8)'\qquad E[S|R] = (k-1) \Big\{\int_{-\infty}^{\infty} \Phi^{k-2}(z+d)\, \Phi(z+d\ \sqrt{n}\ \delta)\varphi(z)dz\Big\}$$

$$+ \int_{-\infty}^{\infty} \Phi^{k-1}(z+d + \sqrt{n}\ \delta)\varphi(z)dz.$$

We desire values of δ for which

$$(4.9)\qquad E[S|R] - E[S|\bar{D}] \leq 0.$$

To do this we will find the common values of δ for which

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+d+\sqrt{n}\,\delta)\,\varphi(z)dz - \Phi(\gamma+(\tfrac{k-1}{k})^{1/2}\sqrt{n}\,\delta) \leq 0 \tag{4.10}$$

and

$$\int_{-\infty}^{\infty} \Phi^{k-2}(z+d)\,\Phi(z+d-\sqrt{n}\,\delta)\varphi(z)dz - \Phi(\gamma-(\tfrac{n}{k(k-1)})^{1/2}\,\delta) \leq 0. \tag{4.11}$$

Now

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+d+\sqrt{n}\,\delta)\varphi(z)dz - \Phi(\gamma+(\tfrac{k-1}{k})^{1/2}\delta\sqrt{n}) \tag{4.12}$$

$$\leq \int_{-\infty}^{\infty} \Phi(z+d+\sqrt{n}\,\delta)\varphi(z)dz - \Phi(\gamma+(\tfrac{k-1}{k})^{1/2}\,\delta\sqrt{n})$$

$$= \Phi(\frac{d+\delta\sqrt{n}}{\sqrt{2}}) - \Phi(\gamma+(\tfrac{k-1}{k})^{1/2}\,\gamma\sqrt{n}) < 0$$

provided

$$\gamma + (\tfrac{k-1}{k})^{1/2}\delta\sqrt{n} > \frac{d-\delta\sqrt{n}}{\sqrt{2}} \tag{4.13}$$

which yields

$$\sqrt{n}\,\delta > \frac{\frac{d}{\sqrt{2}} - \gamma}{(\frac{k-1}{k})^{1/2} - \frac{1}{\sqrt{2}}}, \quad k \geq 3. \tag{4.14}$$

Thus (4.10) is satisfied for all δ as in (4.14).

For the inequaltiy (4.11) we observe that

$$(4.15)\qquad \int_{-\infty}^{\infty}\Phi^{k-2}(z+d)\,\Phi(z+d-\sqrt{n}\,\delta)\varphi(z)dz - \Phi(\gamma-(\tfrac{n}{k(k-1)})^{1/2}\,\delta)$$

$$\leq \int_{-\infty}^{\infty}\Phi(z+d-\sqrt{n}\,\delta)\varphi(z)dz - \Phi(\gamma-(\tfrac{n}{k(k-1)})^{1/2}\,\delta)$$

$$= \Phi(\frac{d-\delta\sqrt{n}}{\sqrt{2}}) - \Phi(\gamma-(\tfrac{n}{k(k-1)})^{1/2}\,\delta) < 0$$

provided

$$(4.16)\qquad \frac{d-\delta\sqrt{n}}{\sqrt{2}} < \gamma - (\frac{n}{k(k-1)})^{1/2}\,\delta,$$

which yields

$$(4.17)\qquad \sqrt{n}\,\delta > \frac{\frac{d}{\sqrt{2}}-\gamma}{\frac{1}{\sqrt{2}}-(\frac{1}{k(k-1)})^{1/2}},\ k \geq 3.$$

Thus (4.11) is satisfied for all δ as in (4.17). For $k \geq 3$ we have

$$(4.18)\qquad (\frac{k-1}{k})^{1/2} - \frac{1}{\sqrt{2}} \leq \frac{1}{\sqrt{2}} - (\frac{1}{k(k-1)})^{1/2}$$

which gives the following result.

$$(4.19)\qquad E[S|R] < E[S|\bar{D}] \quad \text{when} \quad \sqrt{n}\,\delta > \frac{\frac{d}{\sqrt{2}}-\gamma}{(\frac{k-1}{k})^{1/2}-\frac{1}{\sqrt{2}}}\ .$$

~~Thus we see that the rule R yields a smaller subset size on the average than the rule $\bar{D}$ for all mean vectors $\underline{\mu}'$ in Ω' (slippage~~

(Insert next page (13a))

To show the existence of such a δ, it remains to show that the right hand side of (4.14) is a positive number. The denominator is positive if $k \geq 3$. [illegible] Thus we need only show that

$$d - \sqrt{2}\,\gamma > 0 \quad \text{for} \quad k \geq 3. \tag{4.20}$$

Recall that ''d'' is chosen so as to make

$$\inf_{\Omega} P\{CS|R\} = \int_{-\infty}^{\infty} \Phi^{k-1}(z+d)\,\varphi(z)dz = P^* \tag{4.21}$$

and ''γ'' is such that $\Phi(\gamma) = P^*$. Now

$P^* = \int_{-\infty}^{\infty} \Phi(z+d)\varphi(z)dz = \Phi(\frac{d}{\sqrt{2}})$ which implies $d = \sqrt{2}\,\gamma$ when $k=2$. But for $k \geq 3$, $\Phi^{k-1}(z+d) < \Phi(z+d)$ for every z; hence if $d \leq \sqrt{2}\,\gamma$, then

$$\int_{-\infty}^{\infty} \Phi^{k-1}(z+d)\varphi(z)dz < \int_{-\infty}^{\infty} \Phi(z+d)\varphi(z)dz \leq P^*,$$

a contradiction. Therefore (4.20) is true for $k \geq 3$. (Go to top of p. 14)

Thus we see that the rule R yields a smaller subset size on the average than the rule D for all mean vectors $\underline{\mu}'$ in Ω' (slippage configuration) except when δ is relatively small.

The table at the end of Section 3 gives the expected size of the selected subset for the rule R at various values of k and δ assuming the mean vector $\underline{\mu}'$ to be in Ω'.

5. Bayesian Approach

A. Introduction and main theorem

In this section we will make the further assumption that each population mean is itself a random variable with a distribution G_i, $i=1,2,\ldots,k$. The distribution G_i is called an a priori dist^n and $G = \prod_{i=1}^{k} G_i$ is called an a priori distribution on the parameter space Ω. This is the so-called Bayesian approach to the multiple decision problem. And whereas the zero-one loss function (2.5) was used in the previous work we will assume here that the loss function is the so called linear loss:

$$L(S_j,\underline{\mu}') = \sum_{q \in S_j} \alpha_{jq}(\mu_{[k]} - \mu_q) \tag{5.1}$$

where $j=1,2,\ldots,2^k-1$ and $\alpha_{jq} \geq 0$; S_j being a subset of the k populations. This loss function is analogous to the loss function considered by Bahadur and Robbins (1950), Dunnett (1960), and Bland (1961) for the problem of selecting only one population. We then define the Bayes risk (overall expected loss, or average risk) of a decision procedure D with respect to the a priori distribution G as:

$$B(D,G) = \int_{\Omega}\left\{\int_{E^k} L(D,\underline{\mu}')f(\underline{x}|\underline{\mu}')dx\right\} dG(\underline{\mu}), \tag{5.2}$$

where $\underline{x} \epsilon E^k$, Euclidean k-space, and $f(\underline{x}|\underline{\mu}')$ is the product of the k-independent normal densities; with the density corresponding to π_i having mean μ_i, variance 1.

By Fubini's theorem we can write

$$(5.2)' \qquad B(D,G) = \int_{E^k} \left\{ \int_\Omega L(D,\underline{\mu}') f(x|\underline{\mu}') dG(\underline{\mu}) \right\} d\underline{x}$$

and then let

$$\psi_G(D,\underline{x}) = \int_\Omega L(D,\underline{\mu}')\, f(\underline{x}|\underline{\mu}') dG(\underline{\mu}) \ .$$

An optimal rule is the so-called Bayes procedure and is defined as any procedure D^* such that

$$(5.3) \qquad B(D^*,G) \leq B(D,G) \qquad \text{for any } D.$$

It can be seen that D^* is Bayes with respect to G if

$$(5.3)' \qquad \psi_G(D^*,\underline{x}) \leq \psi_G(D,\underline{x})$$

for every D at each $\underline{x} \epsilon E^k$. It is noted that condition (5.3)' is sufficient but not necessary. Now each rule D after observing a vector $\underline{x}$ must select one of the 2^k-1 subsets. Hence if at each $\underline{x}$ we compute the 2^k-1 numbers

$$(5.4) \qquad \psi_G(S_j,\underline{x}) = \int_\Omega L(S_j,\underline{\mu}')\, f(x|\underline{\mu}') dG(\underline{\mu}); j=1,2,\ldots,2^k-1,$$

and select the smallest of these (or anyone of those that may equal the smallest), we then have a Bayes decision procedure with respect to the a priori G. That is, a Bayes decision procedure D* is defined by:

(5.5) $D^* = D^*(\underline{x}) = S_j$ where j is any positive integer $1,2,\ldots,2^k-1$ such that

$$\psi_G(S_j; \underline{x}) = \min\{\psi_G(S_i,\underline{x}): 1 \leq i \leq 2^k-1\}.$$

From looking at the loss function (5.1) one might intuitively feel that the Bayes procedure would select a subset consisting of only one population since only the one element subsets can make the loss function zero. We now state and proovethis result as a theorem. Note that the theorem is stated in a more general framework than for the normal means problem. After proving the theorem and two corollaries, the exact Bayes procedure relative to some specific a priori distributions will be given for the normal means problem.

For convenience we adopt the following notation:

S_j = {one element, population π_j} for $j = 1,2,\ldots,k$ and no explicit knowledge about S_j for $j = k+1,k+2,\ldots,2^k-1$; $f(\underline{x}|\underline{\mu}')$ will be the product of the k-independent densities from which an observation vector (or matrix) $\underline{x}$ is taken; each density being conditional upon a parameter μ_i, $i=1,2,\ldots k$.

(5.6) Theorem: In the loss function (5.1) let $\alpha_{jq} = \alpha > 0$ for $j=1,2,\ldots,k$. Let

$$a_q = \int_\Omega (\mu_{[k]} - \mu_q)\; f(\underline{x}|\underline{\mu})dG(\underline{\mu}),$$

$$a_{[1]} = \min\{a_q\colon 1 \le q \le k\}, \qquad \text{and}$$

$$b_q = a_q - a_{[1]}.$$

If $a_{[1]} \neq 0$, then a necessary and sufficient condition that

$$(5.6.1) \qquad \min\{\psi_G(S_j,\underline{x})\colon 1 \le j \le 2^k-1\} = \min\{\psi_G(S_j,\underline{x}) : 1 \le j \le k\}$$

is that

$$(5.6.2) \qquad \sum_{q\in S_j} \alpha_{jq} \ge \alpha - (1/a_{[1]})\Big(\sum_{q\in S_j} \alpha_{jq} b_q\Big)$$

for every $j=1,2,\ldots,2^k-1$. If $a_{[1]} = 0$, then (5.6.1) is true.

Proof: First observe that using the notation of the theorem, we have from (5.4) that

$$(5.6.3) \qquad \psi_G(S_j,\underline{x}) = \alpha a_j; \; j=1,2,\ldots,k$$

$$(5.6.4) \qquad \psi_G(S_j,\underline{x}) = \sum_{q\in S_j} \alpha_{jq} a_q; \; j+k+1,\ldots,2^k-1.$$

Next it is always true that,

$$\min\{\psi_G(S_j,\underline{x})\colon 1 \le \le 2^k-1\} \le \min\{\psi_G(S_j,\underline{x}) : 1 \le j \le k\} .$$

Hence (5.6.1) is true iff

(5.6.5) $$\min\{\psi_G(S_i,\underline{x}):1 \leq i \leq k\} \leq \psi_G(S_j,\underline{x})$$

for every $j=1,2,\ldots,2^k-1$. Now from (5.6.3) and definition of $a_{[1]}$ we have

(5.6.6) $$\min\{\psi_G(S_i,\underline{x}):\ 1 \leq i \leq l\} = \alpha a_{[1]}\ .$$

Thus (5.6.1) is true iff

(5.6.7) $$\alpha\ a_{[1]} \leq \sum_{q\in S_j} \alpha_{jq} a_q,\ (j=1,2,\ldots,2^k-1).$$

Note that if $a_{[1]} = 0$ then (5.6.7) is clearly true and hence (5.6.1) holds. If $a_{[1]} \neq 0$, we have condition (5.6.7) iff

(5.6.8) $$\alpha\ a_{[1]} \leq \sum_{q\in S_j} \alpha_{jq}(a_{[1]} + b_q)$$

which holds iff

(5.6.9) $$(\alpha - \sum_{q\in S_j} \alpha_{jq}) \leq (1/a_{[1]}) \sum_{q\in S_j} \alpha_{jq} b_q$$

iff

(5.6.10) $$\sum_{q\in S_j} \alpha_{jq} \geq \alpha-(1/a_{[1]}) \sum_{q\in S_j} \alpha_{jq} b_q$$

for $j = 1,2,\ldots,2^k-1$, which is condition (5.6.2) and thus completes the proof.

Remark: In general the quantity a_q is difficult to compute and hence b_q is not available. Thus the necessary and sufficient condition (5.6.2) is not too useful from a practical viewpoint. For this reason the following corollary is significant.

(5.7) Corollary Under the conditions of Theorem (5.6),

$$\min\{\psi_G(S_j,\underline{x}): 1 \le j \le 2^k-1\} = \min\{\psi_G(S_j,\underline{x}): 1 \le j \le k\} \tag{5.7.1}$$

if

$$\sum_{q\in S_j} \alpha_{jq} \ge \alpha \qquad \text{for every } j=1,2,\ldots,2^k-1. \tag{5.7.2}$$

Proof: In the proof of the theorem it was shown that (5.6.7) is necessary and sufficient for (5.7.1). But

$$\sum_{q\in S_j} \alpha_{jq} a_q \ge \sum \alpha_{jq} a_{[1]} = a_{[1]} \sum_{q\in S_j} \alpha_{jq} \ge \alpha\, a_{[1]}$$

using (5.7.2), which completes the proof.

The following corollary is the main result:

(5.8) Corollary: Let (5.1) be the loss function and let $\sum_{q\in S_j} \alpha_{jq} \ge \alpha$ for every $j=1,2,\ldots,2^k-1$. Then the Bayes procedure with respect to an *a priori* distribution G for selecting a subset containing the best of k-populations is given by:

$D^* = D^*(\underline{x}) = S_j$ where j is any positive integer $1,2,\ldots,k$ such that

$$\psi_G(S_j,\underline{x}) = \min\{\psi_G(S_i,\underline{x}): 1 \leq i \leq k\} .$$

Proof: From (5.5) we see that the Bayes procedure is obtained from the $\min\{\psi_G(S_i,\underline{x}): 1 \leq i \leq 2^k-1\}$. But from Corollary (5.7), this minimum occurs among the first k-one element subsets; i.e.

$$\min\{\psi_G(S_i,\underline{x}): 1 \leq i \leq 2^k-1\} = \min\{\psi_G(S_i,\underline{x}): 1 \leq i \leq k\}.$$

Thus the Bayes procedure is to select the one element subset which minimizes the quantity $\psi_G(S_j,\underline{x})$ among the k-such numbers. This completes the proof.

Remark: The following are three examples of loss functions for which Corollary (5.8) is true. Note: $|S_j|$ = number of populations in the subset S_j.

(i) $L(S_j,\underline{\mu}') = \sum_{q\in S_j} (\mu_{[k]} - \mu_q)$, sum of losses.

(ii) $L(S_j,\underline{\mu}') = \frac{1}{|S_j|} \sum_{q\in S_j} (\mu_{[k]} - \mu_q)$, average loss.

(iii) $L(S_j,\underline{\mu}') = (k+1-|S_j|) \sum_{q\in S_j} (\mu_{[k]} - \mu_q)$.

B. Normal means problem with specific a priori distributions.

We will consider the normal means problem and give the Bayes procedure for selecting a subset when the a priori distribution is: (1) normal and (2) uniform. These procedures are the same as those for the problem of selecting the best when the loss function is:

(5.9) $$L(S_j,\underline{\mu}') = \mu_{[k]} - \mu_j; \; j=1,2,\ldots,k.$$

and hence the derivations of these procedures will be given as part of a later publication. A loss function of the form (5.1) satisfying Corollary (5.8) is assumed.

(1) G_i is normal with mean λ_i and variance β_i^2.

The Bayes procedure D* given by Corollary (5.8) says:

Select π_j such that

$$\frac{n\beta_j^2 x_j + \lambda_j}{1 + n\beta_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{n\beta_i^2 x_i + \lambda_i}{1 + n\beta_i^2} \right\},$$

where x_i is the sample mean of n observations.

(2) G_i is uniform on $(\lambda_j - d_j, \lambda_j + d_j)$.

The Bayes procedure D* given by Corollary (5.8) says:

Select π_j such that

$$\frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\Phi(\alpha_j) - \Phi(\beta_j)} + \sqrt{n}\, x_j = \max_{1 \leq i \leq k} \left\{ \frac{\varphi(\beta_i) - \varphi(\alpha_i)}{\Phi(\alpha_i) - \Phi(\beta_j)} + \sqrt{n}\, x_i \right\}$$

where

$\alpha_j = \sqrt{n}\, (\lambda_j + d_j - x_j)$,

$\beta_j = \sqrt{n}\, (\lambda_j - d_j - x_j)$, and

x_j being the sample mean based on n observations.

References

1. Bahadur, R.R. and Robbins, H. (1950). The problem of the greater mean, Ann. Math. Statist. 21, 469-487.

2. Bland, R.P. (1961). A minimum average risk solution for the problem of choosing the largest mean. Mimeograph Series No. 280, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

3. 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.

4. Deely, J.J. (1965). Multiple decision procedures from an empirical Bayes approach. Mimeograph Series No. 45, Department of Statistics, Purdue University.

5. Dunnett, C.W. (1960). On selecting the largest of k normal population means. J. Roy. Stat. Soc. Ser. B 22, 1-40.

6. Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.

7. Gupta, S.S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. Ann. Math. Statist. 28, 957-967.

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

9. Robbins, H. (1964). The empirical Bayes approach to statistical decision problems, Annals of Math. Stat. 35, 1-20.

10. Seal, K.C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 26, 387-298.

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 the Properties of Subset Selection Procedures

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

Technical Report, August 1965

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

Deely, John J. and Gupta, Shanti S.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1965	22	10

8a. CONTRACT OR GRANT NO. NONR-1100(26) b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series No. 49 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Releasable without limitation on dissemination

11. SUPPLEMENTARY NOTES Also supported by the Aerospace Research Laboratories	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Sciences Branch Office of Naval Research Washington, D.C. 20360

13. ABSTRACT

In this paper we study some desirable properties of a selection procedure which selects the normal population π_i with mean μ_i and variance unity $(i=1,2,\ldots,k)$ iff the observed sample mean x_i from $\pi_i \epsilon [x_{Max} - d, x_{max}]$. This rule earlier studied by Gupta (1956, 1965) is compared with the ''approximate" optimal rule $\bar{D}$ of Seal (1955). It is shown that the rule R is minimax. It is also shown that under the slippage configuration of means given by $(\mu,\mu,\ldots,\mu+\delta)$ the expected size of the selected subset using R is smaller than that corresponding to $\bar{D}$ and that the probability of a correct selection using R is strictly greater than that of $\bar{D}$, provided δ satisfies some inequalities. Under a more general linear loss function, the Bayes rule for selecting a subset is also derived.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Selection Ranking Multiple Decision Subset Selection Best Population Bayes Rule						

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.

2*a*. 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.

2*b*. 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.

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

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

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

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

9*a*. 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.

9*b*. 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.