ON SUBSET SELECTION PROCEDURES FOR THE LARGEST MEAN FROM NORMAL POPULATIONS HAVING A COMMON KNOWN COEFFICIENT OF VARIATION

by

Shanti S. Gupta and Ashok K. Singh
Purdue University

and

New Mexico Institute of Mining and Technology

Technical Report #83-6

March 1983

Purdue University
West Lafayette, Ind.

*This research was supported in part ONR Contract N00014-75-C-0455 at Purdue University.

1. INTRODUCTION

The problem of selection of a subset containing the population associated with the largest mean has been extensively studied in the literature for normal populations. References may be found in Gupta and Panchapakesan (1979), Dudewicz and Koo (1981). These procedures are not applicable when the population standard deviations are proportional to the population means, a situation that is quite common in physical and biological applications as pointed out by Gleser and Healy (1976). The problem of estimation of the mean of a normal population with known coefficient of variation has been considered by Khan (1968), and the problem of testing hypotheses concerning the means by Khan (1978). Gleser and Healy (1976) derived the minimum risk scale equivariant estimator $\hat{\theta}_I$ of the mean of a normal population with known coefficient of variation. Joshi and Sathe (1982) derived bounds for the values of the estimator $\hat{\theta}_I$ which are helpful in computing the values of $\hat{\theta}_I$ for a given sample with sufficiently high accuracy.

In the area of ranking and selection, Lehmann (1961) discussed minimax procedures for the selection of normal populations with coefficients of variation smaller than a given constant, assuming a common variance for all the populations. Tamhane (1978) used estimators developed by Gleser and Healy (1976) and proposed ranking and selection rules for normal populations with common known coefficient of variation, and provided tables for implementing the rules in the large sample case. In this paper, we have considered the problem of selection of a subset containing the populations with the largest unknown mean of several normal populations which have a common known coefficient of variation, and have compared Tamhane's selection rule to a rule based on sample variances. It appears that, when the sample size is large, there is not much difference between the rules based on sample variances and Tamhane's procedures, which depend on better estimators of the mean. This comparison could not be made for small sample sizes since the exact tables for Tamhane's rule are not available.

In Section 2, the two rules for selecting the population associated with the largest normal mean are described. In Section 3 some operating characteristics of the two rules are computed for a slippage configuration of the normal means. Section 4 consists of a large sample comparison of the two rules. The asymptotic relative efficiency of the variance procedure with respect to Tamhane's procedure is computed. It turns out that, for large values of the coefficient of variation and the sample size, the two procedures are equivalent. For small sample sizes, tables for implementing Tamhane's rule are not available, and therefore the rule based on sample variances may be used.

2. PROCEDURES FOR SELECTION OF THE POPULATION ASSOCIATED WITH THE LARGEST NORMAL MEAN

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k (≥ 2) independent normal populations with positive means $\theta_1, \ldots, \theta_k$ and a common known coefficient of variation $\sqrt{b}$. The goal is to select, on the basis of an independent random sample

$X_{i1},\ldots,X_{in}$ from each π_i, a subset which contains the population corresponding to the largest θ_i with probability at least P*, where P* is a preassigned constant $(1/K < P^* < 1)$. For each i, let

$$\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n,$$
$$S_i^2 = \sum_{j=1}^{n} (X_{ij} - \bar{X})^2/n. \tag{2.1}$$

Gleser and Healy (1976) have discussed several best asymptotically normal (BAN) estimators of a normal mean θ out of which the following two are almost surely nonnegative:

$$\hat{\theta}^{+}_{LMMS} = \begin{cases} c_n S + d_n \bar{X} & \text{if } c_n S + d_n \bar{X} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

where

$$c_n = a_n b/D_n$$

$$d_n = n(ba_n^2 - 1)/D_n$$

$$a_n = \{(n-1)/2b\}^{\frac{1}{2}} \Gamma\{(n-1)/2\}/\Gamma(n/2)$$

$$D_n = b + (b+n)(ba_n^2 - 1),$$

and

$$\hat{\theta}_I = (b^{-1} \sum_{j=1}^{n} X_j^2)^{\frac{1}{2}} (\int_0^\infty u^n e^{-0.5u^2+Wu} du)/(\int_0^\infty u^{n+1} e^{-0.5u^2+Wu} du) \quad (2.3)$$

where

$$W = (\sum_{j=1}^{n} X_j)/(b\Sigma X_j^2)^{\frac{1}{2}}.$$

Here $\hat{\theta}^+_{LMMS}$ is the positive part of $\hat{\theta}_{LMMS}$, which has the minimum mean squared error (MSE) in the class of all estimators of θ which are linear in $\bar{X}$ and S. The second estimator $\hat{\theta}_I$ is the minimum risk scale equivariant estimator of θ.

Tamhane (1978) proposed the following selection rule for the problem under consideration:

$$R_1: \text{ Select } \pi_i \text{ iff } \hat{\theta}_i \geq \max_{1\leq j\leq K} \hat{\theta}_j/c_1 \quad (2.4)$$

where $\hat{\theta}_i$ is either of the two estimators described above. The constant c_1 satisfying the P*-condition when sample size n is large is given by

$$\int_{-\infty}^{\infty} \Phi^{K-1}[tc_1 + (c_1-1)\lambda]\phi(t)dt = P^* \quad (2.5)$$

where $\lambda = \{n(1+2b)/b\}^{\frac{1}{2}}$.

The values of c_1 satisfying (2.5) are given in Tamhane (1978).

We compare the rule R_1 to the following selection rule R_2:

$$R_2: \text{ select } \pi_i \text{ iff } S_i^2 \geq c_2 \max_{1 \leq j \leq K} S_j^2 \tag{2.6}$$

where $c_2 (0 < c_2 < 1)$ satisfies the basic P*-condition and is given by

$$\int_0^\infty G_{n-1}^{K-1}(u/c_2) g_{n-1}(u) du = P^* \tag{2.7}$$

with $G(\cdot)$ being the cdf and $g_n(\cdot)$ the pdf of a central chi-square random variable with r degrees of freedom.

The values of c_2 satisfying (2.7) are available in Gupta (1963).

3. COMPARISON BETWEEN R_1 and R_2 IN A SLIPPAGE TYPE CONFIGURATION OF THE MEANS

The expected proportion (EP) of populations in the selected subset is used as a criterion of efficiency of a Gupta-type rule [see, for example, Gupta (1965)]. For a selection rule R, let $P_{\underline{\theta}}(i,K,n,P^*|R)$ denote the probability of selecting the population $\pi_{(i)}$ associated with mean $\theta_{[i]}$, where $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[K]}$ are the ordered means. Then

$$EP = \frac{1}{K} \sum_{i=1}^{K} P_{\underline{\theta}}(i,K,n,P^*|R). \tag{3.1}$$

For the rule R_1, we have

$$P_{\underline{\theta}}(i,K,n,P^*|R_1) = (1/\sqrt{\pi}) \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq 1}}^{K} \Phi(c_1 \delta_{ij} u\sqrt{2} + [c_1 \delta_{ij} - 1]\lambda) e^{-u^2} du \tag{3.2}$$

where $\delta_{ij} = \theta_{[i]}/\theta_{[j]}$ and $\lambda = \{\frac{n(1+2b)}{b}\}$,

and for R_2, we have

$$P_{\underline{\theta}}(i,K,n,P^* R_2) = \int_0^\infty \prod_{\substack{j=1 \\ j\neq 1}}^{K} G_{n-1}(u\delta_{ij}^2/c_2)g_{n-1}(u)du. \tag{3.3}$$

Now, assume that the ordered means, $\theta_{[1]},\ldots,\theta_{[k]}$, have the following configuration:

$$\theta_{[i]} = \Delta^{i-1}\theta,\ \Delta > 1,\ \theta > 0,\ i=1,2,\ldots,k.$$

Then

$$\delta_{ij} = \Delta^{i-j}(i \neq j). \tag{3.4}$$

We have computed the values of the expressions (3.2) and (3.3) when $\delta_{ij} = \Delta^{i-j}$, and then evaluated the expected proportions of populations selected by R_1 and R_2, given by (3.1). Hermite integration was used for numerical computation of the integral of (3.2), and Laguerre integration for the integral of (3.3) (see Abramovitz and Stepun (1971) pp. 923-924). Tables I and II show these values for $\Delta = 1.5(\cdot 5)3$, $K = 2(1)5$, and a few selected values of P^*, n and b. It has been pointed out by Tamhane (1978) that for certain values of b, K, n, and P*, the rule R_1 does not exist, i.e., no c_1-value satisfying (2.5) can be found.

While comparing the rules R_1 and R_2 through these tables, it must be kept in mind that Table I is based on large sample results, and hence the numbers in Table 1 for n=15 may not be exact. For sample size n=37, the difference in the two rules in terms of EP is quite small. For example, for n=37, K=3, Δ=1.5, the rule R_1 gives EP=.378 when b=1.5, and the same rule gives EP=.393 when b=3.0. The value of EP for the rule R_2 in this situation is .400. This also indicates that the difference between the two rules in terms of their values of EP decreases with b. Similar result is obtained in the next section in which we compute the asymptotic relative efficiency (ARE) of R_2 with respect to R_1 for the case of k=2 populations.

4. ARE OF R_2 WITH RESPECT TO R_1

Suppose we are given two normal populations $N(\theta, b\theta^2)$ and $N(\Delta\theta, b\Delta^2\theta^2)$ with $\theta > 0$, $b > 0$ and $\Delta > 1$. The population with mean θ will be referred to as the non-best population.

Let S* denote the number of non-best populations selected by a rule. Then small values of S* are desirable, and therefore, consistent with the P*-condition, one would like to keep the expected value of S* as small as possible.

For a given ϵ $(0 < \epsilon < 1)$, let $N_R(\epsilon)$ be the number of observations needed so that

$$E(S^*|R) = \epsilon \tag{4.1}$$

The ARE is defined as follows (see, for example, Barlow and Gupta (1969)).

<u>DEFINITION 4.1:</u> The ARE of a rule R_2 with respect to another rule R_1 is given by

$$ARE(R_2, R_1) = \lim_{\epsilon \downarrow 0} \frac{N_{R_1}(\epsilon)}{N_{R_2}(\epsilon)}$$

Following the method used by Gupta and Singh (1980), we can show that

$$N_{R_2}(\epsilon) = [(y-y')/\ln\Delta]^2 + 2 \tag{4.2}$$

and

$$N_{R_1}(\epsilon) = 2b(y-y')^2/[(1+2b)(\ln\Delta)^2] \tag{4.3}$$

where $y = \Phi^{-1}(P^*)$, $y' = \Phi^{-1}(\epsilon)$. Hence

$$ARE(R_2, R_1) = \lim_{\epsilon \downarrow 0} \frac{\frac{2b}{1+2b}\left(\frac{y-y'}{\ln\Delta}\right)^2}{\left(\frac{y-y'}{\ln\Delta}\right)^2 + 2} = \frac{2b}{1+2b} \tag{4.4}$$

since $y' \to -\infty$ as $\epsilon \to 0$.

The expression (4.4) for the ARE of R_2 with respect to R_1 shows that, for large values of b, the two rules are equivalent, and therefore, for reasons of being simpler to use, R_2 may be preferred over R_1.

TABLE 1

For the rule R_1 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, $i = 1, 2, \ldots, k$, and the expected proportion (EP) of populations in the selected subset; the common coefficient of variation $\sqrt{b}$ is assumed to be known.

$P^* = .90,\ b = 1.5$

		n = 15				n = 37			
Δ		1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k	i								
2	1	.340	.049	.006	.001	.063	.000	.000	.000
	2	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	EP	.669	.525	.503	.500	.531	.500	.500	.500
3	1	.020	.000	.000	.000	.000	.000	.000	.000
	2	.419	.072	.010	.001	.135	.001	.000	.000
	3	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	EP	.479	.357	.337	.334	.378	.334	.333	.333
4	1	.000	.000	.000	.000	.000	.000	.000	.000
	2	.030	.000	.000	.000	.000	.000	.000	.000
	3	.485	.097	.014	.002	.173	.002	.000	.000
	4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	EP	.379	.274	.253	.251	.293	.251	.250	.250
5	1	.000	.000	.000	.000	.000	.000	.000	.000
	2	.000	.000	.000	.000	.000	.000	.000	.000
	3	.037	.000	.000	.000	.000	.000	.000	.000
	4	.522	.113	.017	.003	.197	.003	.000	.000
	5	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	EP	.312	.223	.203	.201	.239	.201	.200	.200

TABLE 1 (cont.)

For the rule R_1 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, i = 1, 2, ... , k, and the expected proportion (EP) of populations in the selected subset; the common coefficient of variation $\sqrt{b}$ is assumed to be known.

$P^* = .90,\ b = 3.0$

	n = 15				n = 37			
Δ	1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k i								
2 1	.371	.068	.011	.002	.101	.001	.000	.000
2	.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.684	.534	.505	.501	.550	.501	.500	.500
3 1	.035	.000	.000	.000	.000	.000	.000	.000
2	.474	.107	.019	.004	.180	.003	.000	.000
3	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.503	.369	.340	.335	.393	.334	.333	.333
4 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.048	.000	.000	.000	.000	.000	.000	.000
3	.531	.134	.025	.005	.218	.005	.000	.000
4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.395	.283	.256	.251	.305	.251	.250	.250
5 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.000	.000	.000	.000	.000	.000	.000	.000
3	.057	.000	.000	.000	.000	.000	.000	.000
4	.562	.150	.029	.006	.241	.006	.000	.000
5	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.324	.230	.206	.201	.248	.201	.200	.200

TABLE 1 (cont.)

For the rule R_1 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, $i = 1, 2, \ldots, k$, and the expected proportion (EP) of populations in the selected subset; the common coefficient of variation $\sqrt{b}$ is assumed to be known.

$P^* = .95,\ b = 1.5$

	n = 15				n = 37			
Δ	1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k i								
2 1	.492	.100	.014	.002	.163	.002	.000	.000
2	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.746	.550	.507	.501	.581	.501	.500	.500
3 1	.048	.000	.000	.000	.000	.000	.000	.000
2	.570	.137	.022	.004	.224	.004	.000	.000
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.539	.379	.341	.335	.408	.335	.333	.333
4 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.071	.000	.000	.000	.000	.000	.000	.000
3	.644	.182	.033	.006	.254	.005	.000	.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.429	.295	.258	.251	.314	.251	.250	.250
5 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.000	.000	.000	.000	.000	.000	.000	.000
3	.072	.000	.000	.000	.000	.000	.000	.000
4	.646	.183	.033	.006	.273	.005	.000	.000
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.343	.237	.207	.201	.255	.201	.200	.200

TABLE 1 (cont.)

For the rule R_1 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, i = 1, 2, ... , k, and the expected proportion (EP) of populations in the selected subset; the common coefficient of variation $\sqrt{b}$ is assumed to be known.

P* = .95, b = 3.0

	n = 15				n = 37			
Δ	1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k i								
2 1	.540	.138	.026	.005	.210	.005	.000	.000
2	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.769	.569	.513	.503	.605	.502	.500	.500
3 1	.070	.000	.000	.000	.000	.000	.000	.000
2	.604	.176	.036	.007	.268	.007	.000	.000
3	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.558	.392	.345	.336	.423	.336	.333	.333
4 1	.001	.000	.000	.000	.000	.000	.000	.000
2	.095	.000	.000	.000	.000	.000	.000	.000
3	.663	.219	.049	.010	.297	.009	.000	.000
4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.440	.305	.262	.253	.324	.252	.250	.250
5 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.001	.000	.000	.000	.000	.000	.000	.000
3	.112	.000	.000	.000	.001	.000	.000	.000
4	.696	.247	.058	.013	.314	.010	.000	.000
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.362	.249	.212	.203	.263	.202	.200	.200

TABLE II

For the rule R_2 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, and the expected proportion of populations in the selected subset; the rule R_2 does not depend upon the common known coefficient of variation $\sqrt{b}$.

P* = .90

	n = 15				n = 37			
Δ	1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k i								
2 1	.423	.107	.022	.004	.130	.003	.000	.000
2	.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.71	.554	.511	.502	.565	.501	.500	.500
3 1	.068	.000	.000	.000	.000	.000	.000	.000
2	.531	.165	.039	.008	.198	.000	.000	.000
3	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.532	.388	.346	.336	.400	.335	.333	.333
4 1	.002	.000	.000	.000	.000	.000	.000	.000
2	.090	.000	.000	.000	.000	.000	.000	.000
3	.588	.203	.052	.012	.242	.009	.000	.000
4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.420	.301	.263	.253	.311	.252	.250	.250
5 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.002	.000	.000	.000	.000	.000	.000	.000
3	.108	.001	.000	.000	.001	.000	.000	.000
4	.624	.230	.062	.015	.274	.011	.000	.000
5	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.347	.246	.212	.203	.255	.202	.200	.200

TABLE II (cont.)

For the rule R_2 and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives the probability of selecting the normal population with mean $\theta_{[i]} = \Delta^{i-1}\theta$, and the expected proportion of populations in the selected subset; the rule R_2 does not depend upon the common known coefficient of variation $\sqrt{b}$.

P* = .95

	n = 15				n = 37			
Δ	1.5	2.0	2.5	3.0	1.5	2.0	2.5	3.0
k i								
2 1	.571	.191	.048	.011	.223	.007	.000	.000
2	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.785	.596	.524	.505	.612	.504	.500	.500
3 1	.134	.000	.000	.000	.002	.000	.000	.000
2	.669	.269	.078	.020	.308	.014	.000	.000
3	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.601	.423	.359	.340	.437	.338	.333	.333
4 1	.005	.000	.000	.000	.000	.000	.000	.000
2	.166	.001	.000	.000	.003	.000	.000	.000
3	.715	.313	.098	.027	.358	.020	.000	.000
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.471	.329	.274	.257	.340	.255	.250	.250
5 1	.000	.000	.000	.000	.000	.000	.000	.000
2	.007	.000	.000	.000	.000	.000	.000	.000
3	.191	.002	.000	.000	.003	.000	.000	.000
4	.745	.346	.114	.032	.393	.024	.000	.000
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EP	.389	.270	.223	.206	.279	.205	.200	.200

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I.A. (1970). Handbook of Mathematical Functions. Dover Publications, Inc., New York.

Barlow, R.E. and Gupta, S.S. (1969). Selection procedures for restricted families of distributions. Ann. Math. Statist. 40: 905-917.

Dudewicz, E.J. and Koo, J.O. (1981). The Complete Categorized Guide to Statistical Selection and Ranking Procedures. American Sciences Press, Inc., Columbus, Ohio.

Gleser, L.J. and Healy, J. (1976). Estimating the mean of a normal distribution with known coefficient of variation. J. Amer. Statist. Assoc. 71: 977-981.

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

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

Gupta, S.S. and Panchapakesan, S. (1979). Multiple Decision Proceedures: Methodology of Selecting and Ranking Populations. John Wiley and Sons, New York.

Gupta, S.S. and Singh, A.K. (1980). On rules based on sample medians for selection of the largest location parameter. Commun. Statist.-Theor. Meth. 9: 1277-1298.

Joshi, S. and Sathe, Y.S. (1982). On approximating the equivariant estimator of the mean of a normal distribution with known coefficient of variation. Commun. Statist.-Theor. Meth. A11: 209-215.

Khan, R. (1978). A note on testing the means of normal distributions with known coefficient of variations. Commun. Statist.-Theor. Math. A7: 867-876.

Khan, R. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. J. Amer. Statist. Assoc. 63: 1039-1041.

Lehmann, E.L. (1961). Some model I problems of selection. Ann. Math. Statist. 32: 990-1012.

Tamhane, A.C. (1978). Ranking and selection problems for normal populations with common known coefficient of variation. Sankhyā B39: 344-361.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #83-6	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ON SUBSET SELECTION PROCEDURES FOR THE LARGEST MEAN FROM NORMAL POPULATIONS HAVING A COMMON KNOWN COEFFICIENT OF VARIATION	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: #83-6
7. AUTHOR(s): Shanti S. Gupta and Ashok K. Singh	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: March 1983
	13. NUMBER OF PAGES: 17
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, distribution 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, Largest Mean, Coefficient of Variation, Normal Populations, Best Asymptotic Estimators.

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

The problem of selecting a subset of k normal populations which includes the population associated with the largest mean is considered for the situation in which the normal populations have a common known coefficient of variation. Subset selection rules based on best asymptotically normal (BAN) estimators of the mean have been studied in the literature and tables based on large sample theory for implementing these rules exist. We have compared these rules to a selection rule based on sample variances, and our limited study suggests that, when n is large, the difference between the rules based on BAN estimates and the variance rule, in

DD FORM 1 JAN 73 1473

UNCLASSIFIED (over)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) terms of the expected proportion of the selected subset, is minimal. Moreover, since the exact distribution theory for BAN estimates is too complicated, and these BAN estimates are much harder to compute than the sample variances, the selection rule based on the sample variances may be preferred.