AD-A047 948 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
ON RULES BASED ON SAMPLE MEDIANS FOR SELECTION OF THE LARGEST L--ETC(U)
NOV 77 S S GUPTA, A K SINGH N00014-75-C-0455
UNCLASSIFIED MIMEOGRAPH SER-514 NL

1 OF 1
ADA047948

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

(6)

ON RULES BASED ON SAMPLE MEDIANS FOR SELECTION OF THE LARGEST LOCATION PARAMETER*

by

(10) Shanti S./Gupta, Purdue University
and
Ashok K./Singh, Purdue University and
National Institute of Environmental Health
Sciences, Research Triangle Park, NC

(9) Technical rept.,

Department of Statistics
Division of Mathematical Sciences
(14) Mimeograph Series-514
(11) November 1977
(12) 39p.

(15) *This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ON RULES BASED ON SAMPLE MEDIANS FOR SELECTION OF THE LARGEST LOCATION PARAMETER*

Shanti S. Gupta
Purdue University

and

Ashok K. Singh
Purdue University and National Institute of Environmental Health Sciences, Research Triangle Park, NC

1. Introduction

Many of the classical statistical procedures which are superior to their competitors under the assumed model have one drawback, namely, that their behavior is seriously affected if a few gross errors are present in the sample. For example, consider the problem of estimation of the mean θ of a univariate normal population. It is well known that the sample mean is uniformly minimum variance unbiased estimate of θ, but it is not a very good estimate if there are gross errors in the sample. Hodges and Lehmann (1963) have proposed a class of estimates for the location parameter based on rank test statistics; the estimates belonging to this class are approximately normally distributed, provided the sample size is sufficiently large. Gupta and Huang (1974) have investigated selection procedures based on one-sample Hodges-Lehmann estimates of location for the problem of selecting a subset containing the largest t ($1 \leq t < k$) location parameters from k ($k \geq 2$) given populations, assuming the sample size is large. An important member of the class of Hodges-Lehmann estimates is the sample median. Apart from having a simple analytic form for its distribution, the sample median as an estimate of location has some other properties. Intuitively, a reasonable estimate of location should have a distribution which, in some sense,

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

is centered on the true location parameter value. It is shown by Hodges and Lehmann (1963) that the sample median has a distribution which is symmetric about the true parameter value if the underlying distribution is symmetric, and in case the underlying distribution is not symmetric, the sample median is a median unbiased estimate of location, i.e., the median of its distribution coincides with the true location parameter. In this paper we investigate a procedure based on sample medians for selection of the largest location parameter of k (≥ 2) populations.

In Section 2.0 some notations used in this paper are introduced. In Sections 3 and 4 the problem of selecting a subset containing the largest of k (≥ 2) location parameters is considered, and a selection rule based on sample medians is proposed and investigated.

Section 5 consists of investigation of selection rules, which are slight modifications of the rule proposed in Section 3, for the normal means. In Section 5.3 the proposed rule based on sample medians is compared to Gupta's rule based on sample means [see Gupta (1965)], when the normal means are equally spaced. It appears from the numerical computations that, as expected, Gupta's rule is superior. In Section 5.5 we define and compute the asymptotic relative efficiency (ARE) of rules based on sample medians relative to rules based on sample means. For the normal case the medians procedure is inferior to the means procedure, the ARE being $2/\pi$. For the contaminated normal population, however, the medians procedure fares better than it does in the normal case, as the ARE is found to be an increasing function of the variance of the contaminating normal populations. In Section 5.6 a test of homogeneity based on sample medians is proposed and a relation between the test and the selection rule of Section 5.1 is established. Section 5.7 deals with the distribution of a statistics useful in some selection and ranking problems, and its percentage points are computed.

As mentioned above, the means procedure is better than the medians procedure if the underlying distributions are normal. This may be due to the fact that the normal density has short tails, and hence the probability of getting extreme observations is very small. In case the underlying distributions have longer tails, for example, logistic and double exponential, extreme observations are more frequent and they have a serious effect on the sample means, but not on the sample medians. In these situations the medians procedure should perform better than in the normal case. This heuristic argument is strengthened by the fact that, for logistic populations, the ARE is $\pi^2/12$, and for double exponential populations, the ARE is 2. This is the subject of Section 6.

2.0 Preliminaries and Notations

Let $X_1,\ldots,X_{2m+1}$ $(m \geq 1)$ be $(2m+1)$ independent observations from a population with cumulative distribution function (cdf) $F(x,\theta)$ and probability density function (pdf) $f(x,\theta)$, $x,\theta \in IR$, the real line. Then the sample median $\tilde{X}$ is given by

$$\tilde{X} = X_{[m+1]}$$

where $X_{[1]} \leq \ldots \leq X_{[2m+1]}$ are the ordered X_i.

The pdf of $\tilde{X}$ is

$$g(x,\theta) = \frac{(2m+1)!}{(m!)^2}[F(x,\theta)]^m[1-F(x,\theta)]^m f(x,\theta) \qquad (2.0.1)$$

and the corresponding cdf is

$$G(x,\theta) = c_m \int_{-\infty}^{x} [F(u,\theta)]^m[1-F(u,\theta)]^m f(u,\theta)du \qquad (2.0.2)$$

$$= I_{F(x,\theta)}(m+1,m+1)$$

where $I_y(p,q)$ is the incomplete beta function:

$$I_y(p,q) = \int_0^y u^{p-1}(1-u)^{q-1}du/B(p,q),\ 0 \le y \le 1,\ p,q > 0. \qquad (2.0.3)$$

The following result from Karlin (1968) will be used in later sections:

Lemma 2.0.1: If $f(x,\theta)$ has monotone likelihood ratio (MLR) in x and θ, then $g(x,\theta)$ given by (2.0.1) has MLR in x and θ.

Note, the above result is true for the distribution of any order statistic.

3.0 On Procedures Based on Sample Medians for Selection of the Largest Location Parameter.

Let $\pi_1,\ldots,\pi_k$ be k independent populations with cdf's $F(x-\theta_1),\ldots,F(x-\theta_k)$, respectively. Let $X_{i1},\ldots,X_{in}$ be a sample of size $n = 2m+1$ $(m \ge 1)$ from π_i, $i = 1,\ldots,k$. Then the pdf $g(x-\theta_i)$ and the cdf $G(x-\theta_i)$ of the sample median $\tilde{x}_i$ from π_i can be obtained from (2.0.1) and (2.0.2) by substituting $f(x,\theta_i) = f(x-\theta_i)$ and $F(x,\theta_i) = F(x-\theta_i)$.

For selecting a subset containing the population $\pi_{[k]}$ associated with the largest location parameter $\theta_{[k]}$, so that the probability of selecting $\pi_{[k]}$ in the subset is at least a preassigned constant P^* $(1/k \le P^* \le 1)$, we consider the following procedure:

$$R:\ \text{select } \pi_i \text{ iff } \tilde{X}_i \ge \tilde{X}_{[k]}-d \qquad (3.0.1)$$

where $d \ge 0$ is chosen to satisfy the basic P^*-condition.

Let $\tilde{X}_{(i)}$ be the (unknown) sample median which corresponds to the i-th ordered parameter $\theta_{[i]}$ $(i = 1,\ldots,k)$. Then

$$P(CS|R) = P(\tilde{X}_{(k)} \ge \tilde{X}_{[k]}-d)$$

$$= \frac{(2m+1)!}{(m!)^2}\int_{-\infty}^{\infty}\Big[\prod_{j=1}^{k-1} I_{F(u+d+\theta_{[k]}-\theta_{[j]})}(m+1,m+1)\Big][F(u)]^m[1-F(u)]^m \cdot f(u)du \qquad (3.0.2)$$

It is clear from the expression (3.0.2) that the infimum of the probability of a correct selection occurs when all the location parameters θ_i are equal, and hence the constant d is given by

$$\int_{-\infty}^{\infty} [G(u+d)]^{k-1} g(u)du = P^* \tag{3.0.3}$$

3.1 Expected Size of the Selected Subset

The size of the subset selected by the rule R is a random variable which takes values in the set $\{1,\ldots,k\}$. It is desirable that the size of the selected subset be small, and also the ranks of the selected populations be large, where the population associated with $\theta_{[i]}$ is given rank i ($i = 1,\ldots,k$). The expected size of the selected subset and expected sum of ranks of selected populations have been proposed as criteria of efficiency of selection rules [see for example, Gupta (1965)].

Let S and S_r be random variables denoting the size of the selected subset and the sum of ranks of the selected populations, respectively. Then

$$E_{\underline{\theta}}(S|R) = \sum_{i=1}^{k} P_{\underline{\theta}}(i|R) \tag{3.1.4}$$

and

$$E_{\underline{\theta}}(S_r|R) = \sum_{i=1}^{k} i\, P_{\underline{\theta}}(i|R) \tag{3.1.5}$$

where $P_{\underline{\theta}}(i|R)$ is the probability with which the rule R selects the population associated with $\theta_{[i]}$, $i = 1,2,\ldots,k$, and is given by

$$P_{\underline{\theta}}(i|R) = \frac{(2m+1)!}{(m!)^2} \int_{-\infty}^{\infty} \Big[\prod_{\substack{j=1 \\ j\neq i}}^{k} I_{F(u+d+\theta_{[i]}-\theta_{[j]})}(m+1,m+1) \Big] \cdot [F(u)]^m [1-F(u)]^m f(u)du \tag{3.1.6}$$

3.2. Some Properties of the Rule R

(i) Upper bound on $E_{\underline{\theta}}(S|R)$

We assume that $f(x-\theta)$ has MLR in x and θ. Then it follows from Lemma 2.0.1 that $g(x-\theta)$ has MLR in x and θ. Hence from Theorem 1 of Gupta (1965), we have

$$\underset{\underline{\theta}\in\Theta}{\text{Max}}\ E_{\underline{\theta}}(S|R) = \underset{\underline{\theta}\in\Theta_0}{\text{Max}}\ E_{\underline{\theta}}(S|R) = kP^*$$

where $\Theta_0 = \{(\theta_1,\ldots,\theta_k) \in \Theta: \theta_1 = \ldots = \theta_k\}$.

(ii) Monotonicity

It is clear from the expression (3.1.6) that the rule R is strongly monotone [see Santner (1975)], i.e.

$P_{\underline{\theta}}(i|R)$ is

$\uparrow$ in $\theta_{[i]}$	if remaining component, of $\underline{\theta}$ are kept fixed
$\downarrow$ in $\theta_{[j]}$, $j\neq i$	if remaining component, of $\underline{\theta}$ are kept fixed

The following two properties of R are immediate consequences of strong monotonicity:

(a) R is monotone: $P_{\underline{\theta}}(i|R) \geq P_{\underline{\theta}}(j|R)$, $1 \leq j \leq i \leq k$

(b) R is unbiased: $P_{\underline{\theta}}(k|R) \geq P_{\underline{\theta}}(i|R)$, $1 \leq i \leq k$

(iii) Minimaxity with Respect to the Expected Subset Size Among Rules Based on Medians

A selection rule R* is said to be minimax with respect to S if

$$\sup_{\Theta}\ E_{\underline{\theta}}(S|R^*) = \inf_{R'}\ \sup_{\Theta}\ E_{\underline{\theta}}(S|R')$$

where the infimum is taken over all selection rules R' which satisfy the P*-condition.

The pdf $g(x,\theta)$ is clearly of location type and it has MLR in x and θ. It follows from Theorem 1.4.2 of Berger (1977) that the selection rule R is minimax with respect to S among all rules based on sample medians.

4.0 Selection of a Subset Containing all Location Parameter Populations Better Than a Control

Suppose we have k+1 independent populations $\pi_0,\pi_1,\ldots,\pi_k$ with densities $f(x,\theta_0)$, $f(x,\theta_1),\ldots,f(x,\theta_k)$, respectively. The population π_0 is a standard or control population. The population π_i is said to be better than the control population π_0 if $\theta_i \geq \theta_0$. We are interested in the subset of populations which are better than control. Gupta and Sobel (1958) have considered the above problem for several distributions and have investigated rules based on sufficient statistics which select a subset such that all populations better than the control are included in the subset with probability at least P*, where P* ($0 \leq P^* \leq 1$) is a preassigned constant. We will consider the case of location parameter populations, when $f(x,\theta_i) = f(x-\theta_i)$, $i = 0,1,\ldots,k$, and investigate a rule based on sample medians. The parameter θ_0 may or may not be known. We consider these two cases separately:

(a) θ_0 known:

Here we are given sample medians $\tilde{X}_i$ of $n = 2m+1$ independent observations from π_i ($i = 1,2,\ldots,k$). Consider the rule R_a defined as follows:

R_a: Select π_i iff $\tilde{X}_i \geq \theta_0 - a$ (4.0.1)

where a is the smallest value satisfying the P*-condition.

Let k_1 denote the number of populations that are better than (as good as) the control, and k_2 denote the number of populations for which $\theta_i < \theta_0$. Then

$k_1 + k_2 = k$. Also let primes (') refer to the k_1 populations better than control. Then we have

$$P(CS|R_a) = \prod_{i=1}^{k_1} P(\tilde{X}_i' \geq \theta_0 - a)$$

$$= \prod_{i=1}^{k_1} [1 - I_{F(\theta_0 - a - \theta_i')}(m+1, m+1)].$$

Clearly

$$\inf P(CS|R_a) = [I_{1-F(-a)}(m+1, m+1)]^{k_1} \tag{4.0.2}$$

In case k_1 is not known, a lower bound for $P(CS|R_a)$ can be obtained by substituting $k_1 = k$ is (4.2), and thus a conservative value of the constant $a = a(P^*, m, k)$ can be obtained from the following equation:

$$I_{1-F(-a)}(m+1, m+1) = (P^*)^{1/k} \tag{4.0.3}$$

Expected Size of the Selected Subset.

The size of the subset selected by the rule R_a is a random variable which takes values in the set $\{0, 1, \ldots, k\}$. Letting S denote the size of a subset selected by a selection rule, the expected value of S can be looked upon as a measure of the performance of the rule [see Gupta (1965)]. Now

$$E(S|R_a) = \sum_{i=1}^{k} P\{\pi_i \text{ is selected}\}$$

$$= \sum_{i=1}^{k} P\{\tilde{X}_i \geq \theta_0 - a\}$$

$$= \sum_{i=1}^{k} [I_{1-F(\theta_0 - a - \theta_i)}(m+1, m+1)] \tag{4.0.4}$$

Let $\theta_{[1]} \leq \theta_{[2]} \leq \ldots \leq \theta_{[k]}$ be the ordered parameters and let $p_{[i]}(R_a)$ be the probability that the rule R_a selects the population associated with $\theta_{[i]}$, $i = 1,\ldots,k$. Then

$$E(S|R_a) = \sum_{i=1}^{k} p_{[i]}(R_a) \tag{4.0.5}$$

4.1. Probability that the Selected Subset Contains Only the Populations Better Than the Control.

Assume that k_1 of the given k populations have parameter $\theta+\delta$ and the remaining populations have parameter θ, where θ, δ and θ_0 are such that $\theta+\delta \geq \theta_0 > \theta$. In this situation it is meaningful to ask for the probability that the rule select, exactly k_1 populations [cf. Gupta (1965)]. We will consider the special case $k_1 = 1$. Letting $p_1(R_a)$ denote the probability of selecting exactly one population, we have

$$p_1(R_a) = \sum_{i=1}^{k} P(M_i \geq \theta_0-a,\ M_j < \theta_0-a,\ j \neq i,\ j = 1,\ldots,k)$$

$$= [1 - I_{F(\theta_0-\theta-\delta-a)}(m+1,m+1)]\ [I_{F(\theta_0-\theta-a)}(m+1,m+1)]^{k-1}$$

$$+ [1 - I_{F(\theta_0-\theta-a)}(m+1,m+1)]\ [I_{F(\theta_0-\theta-\delta-a)}(m+1,m+1)][I_{F(\theta_0-\theta-a)}(m+1,m+1)]^{k-2} \tag{4.1.1}$$

In this case, the value of a is to be obtained from the equation

$$I_{1-F(-a)}(m+1,m+1) = P^* \tag{4.1.2}$$

(b) θ_0 Unknown:

In this case (2m+1) independent observations are taken from π_0. Let $\tilde{X}_0$ be the median of the sample from π_0. We consider the following selection rule:

$$R_b: \text{ select } \pi_i \text{ iff } \tilde{X}_i \geq \tilde{X}_0 - b \qquad (4.1.3)$$

where the constant b is chosen to satisfy the basic P*-condition.

We have, as in Case (a),

$$P(CS|R_b) = \frac{(2m+1)!}{(m!)^2} \int_{-\infty}^{\infty} \left[\prod_{i=1}^{k_1} \{1 - I_{F(u+\theta_0-\theta_i'-b)}(m+1,m+1)\}\right][F(u)]^m[1-F(u)]f(u)du$$

$$\geq \frac{(2m+1)!}{(m!)^2} \int_{-\infty}^{\infty} [1 - I_{F(u-b)}(m+1,m+1)]^k [F(u)]^m [1-F(u)]^m f(u)du \qquad (4.1.4)$$

The constant $b = b(k,m,P^*)$ is obtained by equating the right hand side of (4.9) to P*.

The expected subset size for the rule R_b is obtained as in Case (a).

Remarks:

(i) It can be seen from expression for $P(\text{select } \theta_i)$ for rules R_a and R_b that, in either case

$$P(\text{select } \theta_i) \geq P(\text{select } \theta_j) \quad \text{if } \theta_i \geq \theta_j.$$

(ii) If $\theta_i \to \infty$ for all $i = 1,\ldots,k$, and θ_0 is finite, then $E(S) \to k$ in each case.

In the following sections we will consider several specific densities of location type and investigate, in some detail, rules based on sample medians for selection problems connected with them. As pointed out earlier, the behavior of the proposed selection rule seems to depend on the type of tails of the underlying distributions. It is known [see, for example, Hajek (1969)] that among normal, logistic and double exponential distributions, the normal distribution has the shortest or the thinnest tails, and then come logistic and double exponential distributions, in that order. Subset selection procedures based on sample medians for double exponential populations have been investigated

by Gupta and Leong (1976). McDonald (1977) has investigated a medians procedure for logistic populations. We will be mainly concerned with normal and double exponential populations, but we will include some results for the logistic distribution.

5.0 Normal Populations

Gupta (1956, 1965) has considered the problem of selecting a subset containing the largest of several normal means, and has investigated rules based on sufficient statistics, namely the sample means, assuming a common known variance. It is well known that the sample mean is a uniformly minimum variance unbiased estimate of the normal mean, and therefore it should provide a better selection rule than the rule of Section 3. In the next section we study the normal case in order to get an idea of how far off the medians procedure is from the means procedure.

5.1. Normal Populations with Common Known Variance: A Procedure Based on Sample Medians for Selecting a Subset Containing the Largest Normal Mean.

Let $\pi_1,\ldots,\pi_k$ be $k(\geq 2)$ independent normal populations with means $\theta_1,\ldots,\theta_k$ and a common known variance σ^2. Let $\tilde{X}_i$ be the median of $n = 2m+1$ $(m \geq 1)$ observations from π_i $(i = 1,\ldots,k)$. The pdf $g(x,\theta_i)$ and the cdf $G(x,\theta_i)$ are obtained from (2.0.1) and (2.0.2) by the substitution $f(x,\theta_i) = \phi(x-\theta_i)$ and $F(x,\theta_i) = \Phi(x-\theta_i)$ where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf, respectively, of a standard normal distribution.

For the problem of selecting a subset containing the population associated with the largest mean $\theta_{[k]}$, consider the following procedure

$$R_1: \text{ Select } \pi_i \text{ iff } \tilde{X}_i \geq \tilde{X}_{[k]} - d_1\sigma \qquad (5.1.1)$$

where $d_1 (\geq 0)$, the smallest constant to satisfy the basic P*-condition, is given by the following equation:

$$\frac{(2m+1)!}{(m!)^2\sigma} \int_{-\infty}^{\infty} [I_{\Phi(u+d_1)}(m+1,m+1)]^{k-1} \Phi^m(u)[1-\Phi(u)]^m \phi(u)du = P^* \qquad (5.1.2)$$

5.2. Some Properties of the Rule R_1.

The expressions for $E_{\underline{\theta}}(S|R_1)$, the expected subset size, and $E_{\underline{\theta}}(S_r|R_1)$, the expected sum of ranks, in using the rule R_1 can be obtained from (3.1.4), (3.1.5) and (3.1.6) by substituting $F(u) = \Phi(u)$ and $f(u) = \phi(u)$. Since the normal density has the MLR property, the rule R_1 has all the properties mentioned in Section 3.2 for the general rule R.

5.3. Comparison between R_1 and Gupta's Selection Procedure Based on Sample Means when the Normal Means are Equally Spaced.

Let $\pi_1,\ldots,\pi_k$ be k independent normal populations with means θ, $\theta+\delta\sigma,\ldots,$ $\theta + (k-1)\delta\sigma$ and a common known variance σ^2, where $\delta > 0$ is a known constant. Let X_{ij} $(j = 1,\ldots,n)$ be a sample of size $n = 2m+1$ $(m \geq 1)$ from π_i $(i = 1,\ldots,k)$, and let $\tilde{X}_i$, $\bar{X}_i$ be the median and the sample mean of the observations from π_i.

For the problem of selecting a subset containing the largest normal mean, namely, $\theta + (k-1)\delta\sigma$, Gupta (1965) has proposed the following rule

$$R_g: \text{ Select } \pi_i \text{ iff } \bar{X}_i \geq \bar{X}_{[k]} - \frac{d\sigma}{\sqrt{2m+1}} \qquad (5.3.1)$$

where the constant d satisfying the P*-condition is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}(u+d)\phi(u)du = P^* \qquad (5.3.2)$$

It should be observed that, unlike d_1, the constant d does not depend on n. We will compare the rule R_1 defined by (5.1.1) and (5.1.3) to Gupta's rule R_g.

Let $P(i,k,P^*,\delta,n|R)$ denote the probability with which a rule R selects the population associated with the i-th largest mean (i = 1,...,k). Then from Gupta (1965) we have

$$P(i,k,P^*,\delta,n|R_g) = P(i,k,P^*,\delta\sqrt{n}|R_g)$$

$$= \int_{-\infty}^{\infty} \Big[\prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi(u+d-(j-i)\delta\sqrt{n}) \Big] \phi(u)du. \qquad (5.3.3)$$

Also, for the rule R_1, we have

$$P(i,k,P^*,\delta,n|R_1) = \frac{(2m+1)!}{(m!)^2\sigma} \int_{-\infty}^{\infty} H(u|i,k,P^*,\delta,n)\Phi^m(u)[1-\Phi(u)]^m\phi(u)du \qquad (5.3.4)$$

where

$$H(u|i,k,P^*,\delta,n) = \prod_{\substack{j=1 \\ j\neq i}}^{k} \left[I_{\Phi(u+d_1-(j-i)\delta)}(m+1,m+1) \right]$$

Next, let $\Psi(k,P^*,\delta,n|R)$ and $\Psi_1(k,P^*,\delta,n|R)$ denote the expected sum of ranks and expected average rank of the populations in the selected subset, respectively. Then

$$\Psi(k,P^*,\delta,n|R) = \sum_{i=1}^{k} iP(i,k,P^*,\delta,n|R) = k\Psi_1(k,P^*,\delta,n|R). \qquad (5.3.5)$$

Tables of $P(i,k,P^*,\delta\sqrt{n}|R_g)$, Ψ_1 and the expected proportion of the populations retained in the subset (= $E(S|R_g)/k$) are available in Gupta (1965). We have computed the values of these functions for R_1 for k = 2(1)5, n = 3,5, δ = 0.5(0.5)5.0 and P* = .90, .95. The numerical integration was performed on a CDC 6500 using Gauss-Hermite quadrature based on twenty nodes. These tables are given at the end of this section. For example, for P* = .90, k = 5, n = 3, $\delta = 1.5/\sqrt{3}$ the rule R_g based on sample means selects the second best and third best populations with probabilities .781 and .357, respectively. The corresponding probabilities

for the rule R_1 are .822 and .467, in that order. The probability of a correct selection (selecting the best) has to be greater than .90 for both the rules and is actually equal to .998 for the rule R_g and .997 for the rule R_1. The expected average rank of the selected subset and the expected proportion of the populations selected in the subset for the rule R_g are 1.86 and .441, respectively. The corresponding values for the rule R_1 are 1.995 and .489.

It appears from these tables that the rule R_g based on sample means is superior to the rule R_1 based on sample medians, and also, as expected, the performance of R_g relative to R_1 improves as sample size is increased.

Remarks:

(1) For fixed P* and k

$$P(1,k,P^*,\delta,n|R_1) \uparrow \quad \text{in } \delta\sqrt{n}$$
$$P(k,k,P^*,\delta,n|R_1) \uparrow$$

(2) For fixed P*, i, δ and n

$$P(i,k,P^*,\delta,n|R_1) \uparrow \text{ in } k,\ 1 \le i \le k$$

(3) For fixed k, P* and $(i-j)\delta$

$$\lim_{n\to\infty} \psi(k,P^*,\delta,n) = k$$

(ii) It follows from (5.3.4) and (5.3.5) that

$$\text{(A)}\quad \psi(k,P^*,\delta,n|R_1) > \frac{(2m+1)!}{(m!)^2\sigma} \sum_{i=1}^{k} i\int_{-\infty}^{\infty} H(u|1,k,P^*,\delta,n)\Phi^m(u)[1-\Phi(u)]^m\phi(u)du$$

where the function H is as defined in (5.3.4).

$$\text{(B)}\quad \psi(k,P^*,\delta,n|R_1) > \frac{k(k+1)(2m+1)!}{2(m!)^2\sigma} \int_{-\infty}^{\infty} [I_{\Phi(u+d_1-(k-1)\delta)}(m+i,m+1)]^{k-1}\Phi^m(u)[1-\Phi(u)]^m\phi(u)du$$

5.4. A Selection Rule Based on Medians from Large Samples

Let $f(x)$ be the pdf of a continuous random variable X and let θ be its unique median. Then the distribution of $\tilde{X}$, the sample median based on $n = 2m+1$ ($m \geq 1$) independent observations on X, is known to be asymptotically $N(\theta,[4(f(\theta))^2(2m+1)]^{-1})$ provided certain regularity conditions on $f(x)$ are met [see Cramér (1946)]. This result will be used to investigate procedures based on medians from large samples.

Using the notations of Section 2.0, we see that, for large samples from normal population $N(\theta_i,\sigma^2)$ $i = 1,\ldots,k$, $\tilde{X}_i$ is approximately distributed as $N(\theta_i,\pi\sigma^2/2(2m+1))$. For the problem of selecting a subset containing the largest mean $\theta_{[k]}$, we proposed the following rule:

$$R_2: \text{ Select } \pi_i \text{ iff } \tilde{X}_i \geq \tilde{X}_{[k]} - \frac{d_2\sigma\sqrt{\pi}}{\sqrt{2(2m+1)}} \tag{5.4.1}$$

where the constant $d_2 \geq 0$ is given by

$$\int_{-\infty}^{\infty} \Phi^{k-1}(u+d_2)\phi(u)du = P^*. \tag{5.4.2}$$

Tables of values of d_2 satisfying (5.4.2) are available in Bechhofer (1954) for $k = 1(1)10$ and in Gupta (1963) for $k = 1(1)51$ [see Table 1 of Gupta which gives values of $d_2/\sqrt{2}$ for $n = k-1 = 1(1)50$].

The expression for $P(i,k,P^*,\delta,n|R_2)$, the probability of selecting the population associated with $\theta_{[i]}$ for the rule R_2, is

$$P(i,k,P^*,\delta,n|R_2) = \int_{-\infty}^{\infty} h(u|i,k,P^*,\delta,n)\phi(u)du \tag{5.4.3}$$

where

$$h(u|i,k,P^*,\delta,n) = \prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi(u+d_2+[\theta_{[i]}-\theta_{[j]}]\sqrt{\frac{2(2m+1)}{\pi\sigma^2}}) \tag{5.4.4}$$

5.5 Asymptotic Relative Efficiency (ARE) of the Rule R_2 Relative to Gupta's Rule

In this section we compute the ARE of the rule R_2 with respect to Gupta's rule in the following two cases: (i) independent normal populations and (ii) independent contaminated normal population. We consider the two cases separately.

(i) Independent Normal Populations

Let $\pi_1,\ldots,\pi_k$ be k independent normal populations with means $\theta_1,\ldots,\theta_k$, respectively, and a common known variance σ^2. Assume that θ_i $(i = 1,\ldots,k)$ are in the following slippage configuration:

$$\theta_i = \begin{cases} \theta + \sigma\Delta \text{ if } i = i_0; \quad \Delta > 0 \text{ unknown} \\ \theta \qquad \text{if } i \neq i_0 \end{cases} .$$

The index i_0 $(1 \leq i_0 \leq k)$ is not known. The population π_{i_0} is the best population.

Our interest is in the relative performance of the following two selection procedures:

R_2: Select π_i iff $\tilde{X}_i \geq \tilde{X}_{[k]} - \dfrac{d_2\sigma\sqrt{\pi}}{\sqrt{2n}}$

R_g: Select π_i iff $\bar{X}_i \geq \bar{X}_{[k]} - \dfrac{d\sigma}{\sqrt{n}}$.

The constants d and d_2 satisfying the basic P*-condition are both given by (5.4.2) and hence we have

$$d_2 = d \tag{5.5.1}$$

Let S* be the number of non-best populations in the selected subset. Then small values of S* are desirable and therefore, consistent with the basic P*-condition, we would like to keep the expected value of S* as small as possible.

It is intuitively clear that the performance of any reasonable selection rule should improve as the sample size is increased. For a given ϵ $(0 < \epsilon < 1)$, let $N_{R'}(\epsilon)$ be the number of observations needed so that

$$E(S^*|R') = \epsilon \tag{5.5.2}$$

We will use the following definition of ARE [see Barlow and Gupta (1969)]:

Definition 5.5.1: The ARE of the rule R_2 relative to the rule R_g is defined as

$$ARE\ (R_2,\ R;\theta) = \lim_{\epsilon \downarrow 0} \frac{N_{R_g}(\epsilon)}{N_{R_2}(\epsilon)}\ .$$

Now using the definitions of $N_{R_2}(\epsilon)$ and $N_R(\epsilon)$, it can be shown that

$$\int_{-\infty}^{\infty} [\Phi(u-\Delta\sqrt{2N_{R_2}(\epsilon)/\pi}+d)-\Phi(u-\Delta\sqrt{N_{R_g}(\epsilon)}+d)]\Phi^{k-2}(u+d)\phi(u)du = 0 \tag{5.5.5}$$

Using the fact that Φ is strictly increasing, it can be seen from (5.5.5) that

$$\frac{2\ N_{R_2}(\epsilon)}{\pi} = N_R(\epsilon)$$

and hence

$$ARE(R_2,R) = \lim_{\epsilon \downarrow 0} \frac{N_R(\epsilon)}{N_{R_2}(\epsilon)} = \frac{2}{\pi} = .64.$$

(ii) Independent Contaminated Normal Populations

Suppose that in the course of sampling from population π_i $(i = 1,\ldots,k)$ something happens to the system and gives rise to some wild observations. Assume that the pdf of π_i can be written as

$$f(x,\theta_i) = \alpha\, f_1(x,\theta_i) + (1-\alpha) f_2(x,\theta_i),\ 0 < \alpha < 1 \qquad (5.5.6)$$

This means that the experimenter is sampling from a population with pdf $f_1(x,\theta_i)$ $100\alpha\%$ of the time, and from $f_2(x,\theta_i)$ $100(1-\alpha)\%$ of the time. The presence of observations from $f_2(x,\theta_i)$ is termed as <u>contamination</u>. For our discussion we will assume that

$$f_1(x-\theta_i) = \frac{1}{\sigma}\,\phi\left(\frac{x-\theta_i}{\sigma}\right) \qquad i = 1,\ldots,k$$

$$f_2(x-\theta_i) = \frac{1}{\sigma\sqrt{b}}\,\phi\left(\frac{x-\theta_i}{\sigma\sqrt{b}}\right)$$

where b is a positive constant. We will also assume that the means θ_i are in the same slippage configuration as in Case (i).

Now, it is known [see, for example, Rohatgi (1976)] that $\tilde{X}_i$ and $\bar{X}_i$ both are asymptotically normal, each with mean θ_i and variances $\tilde{\sigma}^2$ and $\bar{\sigma}^2$, respectively, where

$$\tilde{\sigma}^2 = \frac{\pi\sigma^2}{2n}\,\frac{1}{\{\alpha + \frac{1-\alpha}{\sqrt{b}}\}^2} \qquad (5.5.7)$$

$$\bar{\sigma}^2 = \frac{\sigma^2}{n}\,[\alpha + (1-\alpha)b] \qquad (5.5.8)$$

For the problem of selecting a subset containing the best population, consider the following two rules:

R_2^*: select π_i iff $\tilde{X}_i \geq \tilde{X}_{[k]} - d_2^*\tilde{\sigma}$

R_g^*: select π_i iff $\bar{X}_i \geq \bar{X}_{[k]} - d^*\bar{\sigma}$.

It is easily seen that the constants $d_2^* \geq 0$ and $d^* \geq 0$ both satisfy the equation (5.4.2) and hence $d_2^* = d^* = d$, say. Then as in (i) above, we have

$$\frac{\pi\sigma^2}{2N_{R_2^*}(\epsilon)} \frac{1}{[\alpha + \frac{1-\alpha}{\sqrt{b}}]^2} = \frac{\sigma^2}{N_{R^*}(\epsilon)} [\alpha + (1-\alpha)b].$$

Hence

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

$$= \frac{2}{\pi} [\alpha + (1-\alpha)b][\alpha + \frac{1-\alpha}{\sqrt{b}}]^2$$

$$\to \infty \quad \text{as} \quad b \to \infty.$$

The above result shows that for $\alpha < 1$ and large values of b the rule R_2^* based on sample medians is much better than the rule R_g^* based on sample means. In fact, it can be seen from a result in Rohatgi (1976) that the $ARE(R_2^*, R_g^*)$ is close to 1 when $b = 9$ and $\alpha = .915$, and as the differences $b-9 > 0$ and/or $.915-\alpha > 0$ increase the rule R_2^* shows a considerable improvement over R_g^* in terms of the ARE.

5.6 A Test of Homogeneity Based on $\tilde{X}_{[k]} - \tilde{X}_{[1]}$

Let $\pi_1, \ldots, \pi_k$ be k independent normal populations with means $\theta_1, \ldots, \theta_k$, respectively, and a common known variance σ^2. As before, let $\tilde{X}_i$ be the sample median of $n = 2m+1$ ($m \geq 1$) independent observations from π_i ($i = 1, \ldots, k$), and $\tilde{X}_{[1]}, \ldots, \tilde{X}_{[k]}$ be their ordered values. For the hypothesis of homogeneity

$$H_0: \ \theta_1 = \ldots = \theta_k$$

we propose the following test procedure:

$$\text{Reject } H_0 \text{ if } R = \tilde{X}_{[k]} - \tilde{X}_{[1]} > \gamma \tag{5.6.1}$$

where the constant γ is obtained from the size-condition:

$$P_{H_0}(R > \gamma) \leq \alpha.$$

Here α is the size of the test.

The following theorem gives the constant γ, and also establishes a relationship between the test given by (5.6.1) and the selection procedure R_1 of Section (5.1).

Theorem 5.6.1

For $0 < \alpha < 1$, let γ satisfy

$$P_{H_0}(\tilde{X}_k \geq \tilde{X}_{[k]} - \gamma) \geq 1 - \frac{\alpha}{k}$$

Then

$$P_{H_0}(R > \gamma) \leq \alpha.$$

Proof: The proof is similar to that of Theorem 6.1 of Gupta and Leong (1977), and hence omitted.

5.7 On the Distribution of the Statistics Associated with R_1 when the Underlying Distributions are Normal

Let $\tilde{X}_i$ $(i = 0,1,\ldots,k)$ be sample medians of $(k+i)$ sets of $n = 2m+1$ $(m \geq 1)$ independent observations from a standard normal distribution. Define

$$Z_i = \tilde{X}_i - \tilde{X}_0 \qquad (i = 1,\ldots,k).$$

The random variables Z_i are correlated and the distribution of $Z = \max_{1 \leq i \leq k} Z_i$ is needed in some ranking and selection problems. For standard double exponential populations the distribution of Z has been computed by Gupta and Leong (1977) for selected values of k, n and α. In this section we give an expression for the distribution function of Z

and also provide a short table for its upper percentage points for $P^* = \alpha = .75, .85, .90, .95, .99$; $k = 2(1)5$, $n = 3(2)11$.

Let $F(\cdot)$ be the cdf of Z. Then

$$F(z) = P(Z \leq z) = P(\tilde{X}_i \leq \tilde{X}_0 + z,\ i = 1,\ldots,k)$$

$$= \frac{(2m+1)!}{(m!)^2} \int_{-\infty}^{\infty} [I_{\Phi(z+x)}(m+1,m+1)]^k \Phi^m(x)[1-\Phi(x)]^m \phi(x)dx. \qquad (5.7.1)$$

Computations for upper percentage points of F were done on a CDC 6500 using Gauss-Hermite quadrature based on 20 nodes to perform the required numerical integration.

6.0 Logistic and Double Exponential Distributions

The logistic distribution is used frequently as a model in economic demographic problems, and also as a growth curve. The logistic curve although very similar in shape to the normal curve, is different in many ways. It has a heavier tail than the normal, and it does not belong to the Pearsonian or Exponential families of distributions [see Patel, Kapadia and Owen (1976)].

The problem of selection of a subset containing the largest location parameter of several logistic populations has been investigated in detail by McDonald (1977). For selected values of k, n and P*, values of the constant d required for the rule R have been computed. McDonald has also compared the medians procedure to the means procedure and has found the ARE in the logistic case to be $\pi^2/12$. In this sense the rule based on sample medians fares a little better in the logistic case, than it does in the normal means problem.

TABLE I

Upper 100(1-P*) percentage points of $Z = \max_{1 \le i \le k} (X_i - X_0)$ where $X_0, X_1, \ldots, X_k$ are iid sample median random variables in samples of size $n = 2m+1$ ($m \ge 1$) from the standard normal distribution.

k \ n	3	5	7	9	11
	.638	.511	.445	.409	.393
	.980	.784	.676	.614	.582
1	1.213	.969	.832	.751	.710
	1.558	1.245	1.065	.956	.900
	2.208	1.766	1.522	1.398	1.491
	.959	.768	.667	.610	.579
	1.276	1.019	.876	.792	.744
2	1.493	1.192	1.019	.915	.855
	1.816	1.452	1.239	1.105	1.030
	2.429	1.943	1.676	1.533	1.606
	1.125	.901	.783	.715	.675
	1.432	1.142	.980	.884	.828
3	1.642	1.310	1.117	1.000	.931
	1.854	1.563	1.333	1.184	1.099
	2.551	2.040	1.761	1.609	1.671
	1.235	.989	.859	.784	.738
	1.536	1.223	1.048	.945	.883
4	1.742	1.389	1.182	1.057	.982
	2.049	1.639	1.396	1.238	1.145
	2.634	2.106	1.819	1.661	1.715

For given k,n and P* = .75 (top), .85 (second), .90 (third), .95 (fourth), .99 (bottom), the entries in this table are the values of d which satisfy

$$\int_{-\infty}^{\infty} G^k(x+d)\, g(x)\, dx = P^*$$

where $G(\cdot)$ is the cdf and $g(\cdot)$ the pdf of the median of a sample of size n from a standard normal population; $n \ge 3$ is an odd integer.

TABLE IIA

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank i when the population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .90, n = 3

$\delta\sqrt{n}$	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k i										
2 1	.836	.749	.643	.524	.404	.291	.196	.123	.072	.039
2	.944	.971	.986	.994	.997	.999	1.000	1.000	1.000	1.000
3 1	.778	.585	.359	.171	.061	.016	.003	.000	.000	.000
2	.882	.828	.745	.640	.521	.400	.288	.194	.121	.070
3	.959	.983	.993	.997	.999	1.000	1.000	1.000	1.000	1.000
4 1	.704	.381	.118	.019	.001	.000	.000	.000	.000	.000
2	.818	.648	.423	.217	.084	.024	.005	.001	.000	.000
3	.907	.865	.793	.697	.583	.462	.344	.240	.156	.094
4	.969	.989	.996	.998	.999	1.000	1.000	1.000	1.000	1.000
5 1	.612	.193	.019	.001	.000	.000	.000	.000	.000	.000
2	.740	.425	.142	.024	.002	.000	.000	.000	.000	.000
3	.845	.688	.467	.251	.102	.031	.007	.001	.000	.000
4	.923	.887	.822	.733	.624	.504	.384	.274	.183	.113
5	.975	.992	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000

TABLE IIIA

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .90, n = 3

	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.361	1.345	1.307	1.256	1.199	1.145	1.098	1.061	1.036	1.019
	.890	.860	.814	.759	.701	.645	.598	.562	.536	.519
3	1.806	1.730	1.610	1.481	1.367	1.272	1.193	1.129	1.081	1.047
	.873	.799	.699	.603	.527	.472	.430	.398	.374	.357
4	2.234	2.057	1.831	1.634	1.479	1.358	1.261	1.180	1.117	1.071
	.849	.721	.582	.483	.417	.372	.337	.310	.289	.274
5	2.639	2.323	1.995	1.745	1.561	1.422	1.311	1.220	1.146	1.091
	.819	.637	.489	.401	.346	.307	.278	.255	.237	.223

TABLE IIB

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank i when the population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .95, n = 3

$\delta\sqrt{n}$		.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k	i										
2	1	.910	.850	.768	.665	.548	.427	.312	.213	.136	.080
	2	.974	.988	.995	.998	.999	1.000	1.000	1.000	1.000	1.000
3	1	.871	.718	.500	.277	.118	.037	.009	.001	.000	.000
	2	.938	.902	.842	.758	.653	.535	.414	.301	.204	.129
	3	.982	.993	.998	.999	1.000	1.000	1.000	1.000	1.000	1.000
4	1	.784	.477	.173	.033	.003	.000	.000	.000	.000	.000
	2	.875	.732	.517	.292	.126	.041	.010	.002	.000	.000
	3	.940	.909	.851	.770	.668	.551	.430	.315	.216	.138
	4	.982	.994	.998	.999	1.000	1.000	1.000	1.000	1.000	1.000
5	1	.742	.306	.043	.002	.000	.000	.000	.000	.000	.000
	2	.842	.565	.235	.052	.006	.000	.000	.000	.000	.000
	3	.914	.798	.602	.369	.176	.063	.016	.003	.000	.000
	4	.961	.938	.894	.828	.739	.631	.512	.391	.281	.188
	5	.990	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE IIIB

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .95, n = 3

$\delta\sqrt{n}$	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.429	1.413	1.379	1.330	1.273	1.213	1.156	1.107	1.068	1.040
	.942	.919	.881	.831	.774	.713	.656	.607	.568	.540
3	1.898	1.834	1.725	1.597	1.474	1.369	1.279	1.201	1.136	1.086
	.930	.871	.780	.678	.590	.524	.474	.434	.401	.376
4	2.321	2.161	1.938	1.731	1.565	1.434	1.327	1.237	1.162	1.103
	.895	.778	.635	.523	.449	.398	.360	.329	.304	.284
5	2.792	2.514	2.178	1.904	1.699	1.543	1.419	1.315	1.225	1.150
	.890	.721	.555	.450	.384	.339	.306	.279	.256	.238

TABLE IIC

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank i when the population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .90, n = 5

$\delta\sqrt{n}$		.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k	i										
2	1	.838	.754	.653	.539	.422	.312	.216	.139	.084	.047
	2	.943	.969	.985	.993	.997	.999	1.000	1.000	1.000	1.000
3	1	.782	.596	.379	.191	.073	.021	.005	.001	.000	.000
	2	.882	.832	.753	.652	.538	.422	.311	.215	.139	.084
	3	.958	.982	.993	.997	.999	1.000	1.000	1.000	1.000	1.000
4	1	.710	.400	.134	.024	.002	.000	.000	.000	.000	.000
	2	.822	.657	.442	.239	.098	.031	.007	.001	.000	.000
	3	.907	.868	.799	.708	.599	.483	.368	.264	.177	.110
	4	.967	.988	.995	.998	.999	1.000	1.000	1.000	1.000	1.000
5	1	.620	.214	.025	.001	.000	.000	.000	.000	.000	.000
	2	.746	.444	.161	.031	.003	.000	.000	.000	.000	.000
	3	.848	.698	.486	.274	.119	.039	.010	.002	.000	.000
	4	.922	.890	.828	.743	.639	.525	408	.299	.205	.131
	5	.974	.991	.996	.999	.999	1.000	1.000	1.000	1.000	1.000

TABLE IIIC

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, \ldots, \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .90, n = 5

$\delta\sqrt{n}$	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.361	1.346	1.311	1.262	1.208	1.155	1.107	1.069	1.042	1.023
	.890	.862	.819	.766	.710	.655	.608	.569	.542	.523
3	1.807	1.735	1.621	1.495	1.382	1.288	1.209	1.144	1.093	1.056
	.874	.803	.708	.613	.537	.481	.439	.405	.380	.361
4	2.236	2.068	1.850	1.654	1.499	1.378	1.280	1.198	1.132	1.083
	.851	.728	.593	.492	.425	.378	.344	.316	.294	.278
5	2.643	2.342	2.019	1.770	1.583	1.443	1.332	1.240	1.164	1.105
	.822	.647	.499	.409	.352	.313	.284	.260	.241	.226

TABLE IID

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, .., \theta + (k-1)\delta\sigma)$ this table gives the probability of selecting the normal population with rank i when the population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .95, n = 5

$\delta\sqrt{n}$		.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k	i										
2	1	.912	.854	.776	.678	.566	.449	.337	.236	.155	.095
	2	.974	.987	.994	.998	.999	1.000	1.000	1.000	1.000	1.000
3	1	.875	.729	.521	.304	.137	.047	.012	.002	.000	.000
	2	.939	.905	.848	.769	.670	.557	.441	.328	.230	.150
	3	.981	.993	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000
4	1	.824	.543	.230	.053	.006	.000	.000	.000	.000	.000
	2	.901	.778	.582	.359	.174	.064	.018	.004	.001	.000
	3	.953	.928	.881	.811	.721	.615	.499	.383	.277	.187
	4	.986	.995	.998	.999	1.000	1.000	1.000	1.000	1.000	1.000
5	1	.753	.335	.054	.003	.000	.000	.000	.000	.000	.000
	2	.849	.585	.264	.065	.009	.001	.000	.000	.000	.000
	3	.917	.809	.622	.399	.202	.078	.023	.005	.001	.000
	4	.962	.941	.899	.837	.754	.652	.539	.422	.311	.215
	5	.989	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE IIID

For the rule R_1 and the configuration $(\theta, \theta + \delta\sigma, ..., \theta + (k-1)\delta\sigma)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the normal population with mean $\theta + (i-1)\delta\sigma$ has rank i, i = 1,2,...,k; the common variance σ^2 is assumed to be known.

P* = .95, n = 5

$\delta\sqrt{n}$	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.429	1.414	1.382	1.336	1.282	1.224	1.168	1.118	1.078	1.047
	.943	.921	.885	.838	.783	.724	.668	.618	.578	.547
3	1.899	1.839	1.736	1.613	1.492	1.387	1.298	1.220	1.153	1.100
	.932	.876	.789	.690	.602	.535	.484	.444	.410	.383
4	2.357	2.216	2.007	1.801	1.629	1.493	1.383	1.289	1.208	1.140
	.916	.811	.673	.556	.475	.420	.379	.347	.319	.297
5	2.799	2.536	2.208	1.935	1.728	1.569	1.445	1.340	1.250	1.172
	.894	.733	.568	.461	.393	.346	.312	.285	.262	.243

The problem of selection of a subset containing the largest location parameter of several double exponential populations has been considered by Gupta and Leong (1976), and the selection rule proposed in Section 3.0 has been investigated using both exact and large sample distributions of the sample median. We include some of the results of Gupta and Leong (1976) for the sake of completeness, and investigate the problem a little further by numerically computing the values of the functions $P(i,k,P^*,\delta,n|R_1)$ and $\Psi(k,P^*,\delta,n|R_1)$ defined in Section 5.3 when the location parameters of the double exponential populations are equally spaced. We also compute the ARE of the rule R_1 relative to a rule based on sample means. It is seen in this case that the rule R_1 based on sample medians in superior to the rule based on sample means in terms of the ARE.

6.1 Selection of the Largest of Location Parameters of Several Double Exponential Populations.

Let $\pi_1,\ldots,\pi_k$ be k independent double exponential populations with location parameters $\theta_1,\ldots,\theta_k$ respectively. For the problem of selecting a subset containing $\theta_{[k]}$, the largest location parameter, the equation for the constant d of the rule R is given in Gupta and Leong (1976), and can also be obtained by substituting

$$f(u) = \frac{1}{\sqrt{2}} e^{-|u|\sqrt{2}}, \quad -\infty < u < \infty$$

$$F(u) = \begin{cases} \frac{1}{2} e^{u\sqrt{2}}, & u < 0 \\ 1 - \frac{1}{2} e^{-u\sqrt{2}}, & u \geq 0 \end{cases}$$

in the equation (3.0.3).

Since the double exponential distribution has the MLR property, it follows from Section 3.2 that the rule R_1 has the properties mentioned in that section. This has also been observed by Gupta and Leong (1976).

6.2. On the Performance of the Rule R when the Location Parameters are Equally Spaced.

Suppose the location parameters $\theta_1,\ldots,\theta_k$ of the k given double exponential populations are equally spaced, i.e., $\theta_i = \theta + (i-1)\delta$, $i = 1,\ldots,k$, where $\delta > 0$ is a known constant. Then $P(i,k,P^*,\delta,n|R)$, the probability with which the rule R_1 selects the population associated with $\theta_{[i]}$, is given by

$$P(i,k,P^*,\delta,n|R) = \frac{(2m+1)!}{2(m!)^2} \int_0^\infty [h_1(u|i,k,P^*,\delta,n) + h_2(u|i,k,P^*,\delta,n)]g(u)du \quad (6.2.1)$$

where

$$h_1(u|i,k,P^*,\delta,n) = \prod_{\substack{j=1 \\ j\neq i}}^{k} I_{\frac{1}{2}e^{-u+(d-(j-i)\delta)\sqrt{2}}}(m+1,m+1)$$

$$h_2(u|i,k,P^*,\delta,n) = \prod_{\substack{j=1 \\ j\neq i}}^{k} I_{1-\frac{1}{2}e^{-u-(d-(j-i)\delta)\sqrt{2}}}(m+1,m+1)$$

and $$g(u) = [(1 - \tfrac{1}{2}e^{-u})(\tfrac{1}{2}e^{-u})]^m e^{-u}.$$

Expressions for the expected sum of ranks and the expected average rank of the populations retained in the subset can be obtained from (5.3.5) and (6.2.1).

For selected values of k, n and P*, tables of the constant d for double exponential populations are given in Gupta and Leong (1976). Using these tables, we have computed the values of the function $P(i,k,P^*,\delta,n|R)$, the expected average rank and the expected proportion of populations in the selected subset for n = 3, 5, P* = .75, .90, .95, .99 k = 2(1)5 and δ = 0.5(0.5)5.0. Computations

were made on a CDC 6500 using Gauss Laguerre quadrature based on fifteen nodes for the numerical integration. Tables are given at the end of this section. For example if $n = 3$, $P^* = .75$, $k = 5$ and $\delta = 1.5$, the probability of selecting the third best, the second best, and the best populations are .108, .794 and 1.000, in that order. The expected average rank in this case is 1.701 and the expected proportion of the selected populations is .381.

6.3 A comparison of rules based on medians and means of large samples

Let $\pi_1,\ldots,\pi_k$ be k independent double exponential populations with means $\theta_1,\ldots,\theta_k$, respectively and common variance unity. Assume that for some (unknown) index i_0 $(1 \leq i_0 \leq k)$, $\theta_{i_0} - \Delta = \theta_i = \theta$, $i = 1,\ldots,k$, $i \neq i_0$, where $\Delta > 0$ is an unknown constant. Let $\bar{X}_i$ and $\tilde{X}_i$ denote the sample mean and sample median of an independent sample of size $n = 2m+1$ $(m \geq 1)$ from $\pi_i (i = 1,\ldots,k)$. For the problem of selecting a subset containing the largest mean θ_{i_0}, the following two rules can be used:

$\bar{R}$: Select π_i iff $\bar{X}_i \geq \bar{X}_{[k]} - \bar{d}/\sqrt{2m+1}$ (6.3.1)

$\tilde{R}$: Select π_i iff $\tilde{X}_i \geq \tilde{X}_{[k]} - \tilde{d}/\sqrt{2(2m+1)}$ (6.3.2)

where the constant, $\bar{d} \geq 0$ and $\tilde{d} \geq 0$ are determined by the basic probability requirement. If sample size n is sufficiently large, $\bar{X}_i$ and $\tilde{X}_i$ are both normally distributed with mean θ_i and variances $1/(2m+1)$ and $1/2(2m+1)$, respectively. It is easy to see, as in Section 5.5, that

$$\bar{d} = \tilde{d} = d, \quad \text{say.} \tag{6.3.3}$$

In the notation of Section 5.5 , let S^* be the number of non-best populations in the selected subset, and let $N_{R'}(\epsilon)$ be the number of observations needed so that

$$E(S^*|R') = \epsilon.$$

Following the method of Section 5.5 we can see that

$$N_{\bar{R}}(\epsilon) = 2N_{\tilde{R}}(\epsilon)$$

and hence we have

$$ARE(\tilde{R},\bar{R}) = \lim_{\epsilon \downarrow 0} \frac{N_{\bar{R}}(\epsilon)}{N_{\tilde{R}}(\epsilon)} = 2.$$

TABLE IV A

For the rule R and the configuration $(\theta,\theta+\delta,\ldots,\theta+(k-1)\delta)$ this table gives the probability of selecting the double exponential population with rank i when the population with mean $\theta+(i-1)\delta$ has rank i $(i=1,\ldots,k)$.

$P^* = .90,\ n = 3$

δ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k i										
2 1	.879	.705	.429	.195	.073	.025	.008	.002	.001	.000
2	.986	.996	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3 1	.814	.335	.050	.005	.000	.000	.000	.000	.000	.000
2	.939	.840	.634	.354	.152	.055	.018	.006	.002	.000
3	.993	.998	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4 1	.687	.067	.002	.000	.000	.000	.000	.000	.000	.000
2	.874	.450	.079	.008	.001	.000	.000	.000	.000	.000
3	.960	.894	.737	.468	.219	.084	.030	.009	.003	.001
4	.995	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5 1	.483	.007	.000	.000	.000	.000	.000	.000	.000	.000
2	.753	.095	.003	.000	.000	.000	.000	.000	.000	.000
3	.906	.538	.108	.012	.001	.000	.000	.000	.000	.000
4	.971	.922	.794	.553	.280	.114	.041	.013	.004	.001
5	.997	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE V A

For the rule R and the configuration $(\theta,\theta+\delta,\ldots,\theta+(k-1)\delta)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the double exponential population with mean $\theta+(i-1)\delta$ has rank i $(i=1,\ldots,k)$.

$P^* = .90,\ n = 3$

δ	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.425	1.348	1.213	1.097	1.036	1.013	1.004	1.001	1.000	1.000
	.932	.851	.714	.598	.537	.513	.504	.501	.500	.500
3	1.890	1.670	1.439	1.237	1.102	1.037	1.012	1.004	1.001	1.000
	.915	.724	.561	.453	.384	.352	.339	.335	.334	.333
4	2.324	1.911	1.592	1.355	1.165	1.063	1.022	1.007	1.002	1.001
	.879	.603	.454	.369	.305	.271	.257	.252	.251	.250
5	2.715	2.099	1.701	1.450	1.225	1.091	1.033	1.010	1.003	1.001
	.822	.512	.381	.313	.256	.223	.208	.203	.201	.200

TABLE IV B

For the rule R and the configuration $(\theta, \theta+\delta, \ldots, \theta+(k-1)\delta)$ this table gives the probability of selecting the double exponential population with rank i when the population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = .95$, n = 3

k	i	δ = .5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.954	.872	.692	.414	.187	.069	.024	.007	.002	.001
	2	.995	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	.927	.604	.139	.016	.001	.000	.000	.000	.000	.000
	2	.978	.939	.832	.618	.339	.144	.052	.017	.005	.002
	3	.998	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	.863	.188	.006	.000	.000	.000	.000	.000	.000	.000
	2	.952	.712	.203	.026	.002	.000	.000	.000	.000	.000
	3	.986	.960	.888	.723	.451	.209	.079	.028	.009	.003
	4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	.739	.027	.000	.000	.000	.000	.000	.000	.000	.000
	2	.898	.246	.009	.000	.000	.000	.000	.000	.000	.000
	3	.965	.774	.261	.037	.004	.000	.000	.000	.000	.000
	4	.990	.971	.917	.784	.536	.267	.107	.038	.012	.004
	5	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE V B

For the rule R and the configuration $(\theta, \theta+\delta, \ldots, \theta+(k-1)\delta)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the double exponential population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = 95$, n = 3

k	δ = 0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.472	1.435	1.346	1.207	1.093	1.035	1.012	1.004	1.001	1.000
	.975	.935	.846	.707	.593	.535	.512	.504	.501	.500
3	1.959	1.827	1.601	1.418	1.226	1.096	1.035	1.011	1.004	1.001
	.968	.847	.657	.545	.447	.381	.351	.339	.335	.334
4	2.430	2.123	1.769	1.556	1.339	1.157	1.059	1.021	1.006	1.002
	.950	.715	.524	.437	.363	.302	.270	.257	.252	.251
5	2.877	2.345	1.894	1.649	1.431	1.214	1.086	1.031	1.010	1.003
	.918	.604	.438	.364	.308	.253	.221	.208	.202	.201

TABLE IV C

For the rule R and the configuration $(\theta, \theta + \delta, \ldots, \theta + (k-1)\delta)$ this table gives the probability of selecting the double exponential population with rank i when the population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = .90$, $n = 5$

k	i	δ = .5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.833	.524	.194	.050	.010	.002	.000	.000	.000	.000
	2	.991	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	.680	.100	.004	.000	.000	.000	.000	.000	.000	.000
	2	.914	.708	.350	.108	.024	.005	.001	.000	.000	.000
	3	.996	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	.412	.006	.000	.000	.000	.000	.000	.000	.000	.000
	2	.773	.148	.007	.000	.000	.000	.000	.000	.000	.000
	3	.945	.794	.458	.157	.038	.008	.001	.000	.000	.000
	4	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	.157	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.495	.008	.000	.000	.000	.000	.000	.000	.000	.000
	3	.824	.194	.010	.000	.000	.000	.000	.000	.000	.000
	4	.960	.841	.538	.203	.053	.011	.002	.000	.000	.000
	5	.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE V C

For the rule R and the configuration $(\theta, \theta + \delta, \ldots, \theta + (k-1)\delta)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the double exponential population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = .90$, $n = 5$

k	δ = 0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.408	1.260	1.097	1.025	1.005	1.001	1.000	1.000	1.000	1.000
	.912	.761	.597	.525	.505	.501	.500	.500	.500	.500
3	1.832	1.505	1.235	1.072	1.016	1.003	1.001	1.000	1.000	1.000
	.863	.602	.451	.369	.341	.335	.334	.333	.333	.333
4	2.196	1.671	1.347	1.118	1.029	1.006	1.001	1.000	1.000	1.000
	.782	.487	.366	.289	.260	.252	.250	.250	.250	.250
5	2.490	1.792	1.436	1.162	1.043	1.009	1.002	1.000	1.000	1.000
	.687	.409	.310	.241	.211	.202	.200	.200	.200	.200

TABLE IV D

For the rule R and the configuration $(\theta, \theta + \delta, \ldots, \theta + (k-1)\delta)$ this table gives the probability of selecting the double exponential population with rank i when the population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = .95$, $n = 5$

	δ	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k	i										
2	1	.933	.756	.405	.132	.031	.006	.001	.000	.000	.000
		.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	.857	.239	.013	.000	.000	.000	.000	.000	.000	.000
	2	.969	.871	.600	.248	.069	.014	.003	.000	.000	.000
	3	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	.668	.018	.000	.000	.000	.000	.000	.000	.000	.000
	2	.902	.333	.022	.001	.000	.000	.000	.000	.000	.000
	3	.981	.912	.699	.341	.104	.023	.005	.001	.000	.000
	4	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	.365	.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.736	.026	.000	.000	.000	.000	.000	.000	.000	.000
	3	.927	.406	.031	.001	.000	.000	.000	.000	.000	.000
	4	.986	.935	.763	.413	.135	.032	.006	.001	.000	.000
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE V D

For the rule R and the configuration $(\theta, \theta + \delta, \ldots, \theta + (k-1)\delta)$ this table gives the expected average rank of the selected subset (top) and the expected proportion of the populations selected in the subset (bottom) when the double exponential population with mean $\theta + (i-1)\delta$ has rank i $(i = 1, \ldots, k)$.

$P^* = .95$, $n = 5$

δ	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
k										
2	1.464	1.378	1.202	1.066	1.015	1.003	1.001	1.000	1.000	1.000
	.965	.878	.702	.566	.515	.503	.501	.500	.500	.500
3	1.931	1.660	1.404	1.165	1.046	1.010	1.002	1.000	1.000	1.000
	.942	.703	.538	.416	.356	.338	.334	.333	.333	.333
4	2.353	1.855	1.535	1.256	1.078	1.017	1.003	1.001	1.000	1.000
	.887	.566	.430	.335	.276	.256	.251	.250	.250	.250
5	2.712	2.002	1.629	1.331	1.108	1.025	1.005	1.001	1.000	1.000
	.803	.474	.359	.283	.227	.206	.201	.200	.200	.200

BIBLIOGRAPHY

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

2. Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25: 16-39.

3. Berger, R. (1977). Minimax, Admissible and Gamma-Minimax Multiple Decision Rules. Ph.D. Thesis, Department of Statistics, Purdue University, W. Lafayette, IN 47907.

4. Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press, p. 369.

5. Gupta, S. S. (1956). On a decision rule for a problem in ranking means. Mimeo. Ser. No. 150. Inst. of Statist. University of North Carolina, Chapel Hill, NC.

6. Gupta, S. S. (1963). Probability integrals for the multivariate normal and the multivariate t. Ann. Math. Statist. 34: 792-828.

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

8. Gupta, S. S. and Huang, D. Y. (1974). Nonparametric subset selection procedures for the t best populations. Bull. Inst. Math. Acad. Sinica 2: 377-386.

9. Gupta, S. S. and Leong. Y. K. (1976). Some results on subset selection procedures for double exponential populations. Proceedings on the Workshop on Decision Information for Tactical Command and Control (ed. R. M. Thraal, C. P. Tsokos, J. C. Turner). Robert M. Thrall and Associates, Houston, TX.

10. Gupta, S. S. and Sobel, M. (1958). On selecting a subset which contains all populations better than a standard. Ann. Math. Statist. 29: 235-244.

11. Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. Ann. Math. Statist. 34: 598-611.

12. Johnson, N. L. and Kotz, S. (1970). Continuous Univariate Distributions-1. Houghton Mifflin Co., Boston.

13. Karlin, S. (1968). Total Positivity. Stanford University Press, p. 155.

14. Lorenzen, T. J. and McDonald, G. C. (1977). A Selection Procedure for Logistic Populations-Location Parameter Case. GM Research Labs. Technical Report.

15. Patel, J. K., Kapadia, C. H. and Owen, D. B. (1976). Handbook of Statistical Distributions. Marcel Dekker, Inc., N.Y.

16. Rohatgi, V. K. (1976). An Introduction to Probability Theory and Mathematical Statistics. John Wiley, p. 581-583.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #514	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Rules Based on Sample Medians for Selection of the Largest Location Parameter		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER Mimeo. Ser. No. 514
7. AUTHOR(s) Shanti S. Gupta and Ashok K. Singh		8. CONTRACT OR GRANT NUMBER(s) ONR 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
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE October 1977 13. NUMBER OF PAGES 35
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)

Correct selection, contamination, asymptotic relative efficiency, monotone likelihood ratio, minimax, monotone, unbiased, homogeneity.

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

The problem of selection of a subset containing the largest of several given location parameters is considered, and ~~Gupta type~~ selection rules based on sample medians are proposed and investigated. Selection of the largest normal mean is considered in detail and some new results for double exponential populations are also obtained. Numerical comparisons in the case of equally spaced normal means are made between the medians procedure and the means procedure. The asymptotic relative efficiency (ARE) of the medians procedure relative to the means procedure is also computed, assuming that the normal

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

means are in a slippage configuration. The means procedure is found to be much better than the medians procedure in the sense of ARE. However, if the normal populations are highly contaminated, the proposed rule based on sample medians is superior to the means procedure.