The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1640–1645

SOME ROBUST SELECTION PROCEDURES[1]

BY RONALD H. RANDLES

The University of Iowa

1. Introduction and summary. Let X_{it} $(t = 1, \cdots, n; i = 1, \cdots, k)$ be independent observations from k populations with respective distribution functions $F(x-\theta_i)$, where the translation parameters θ_i are unknown. Consider the problem of selecting one population, the objective being to select the population with largest translation parameter. Procedures based on the joint ranking of all nk observations have been considered by Lehmann [5], Bartlett and Govindarajulu [1], and Puri and Puri [9]. Robust procedures for related problems have been considered by Sobel [11] and McDonald and Gupta [7], among others.

Define the ith population to be good if $\theta_i > \theta_{\max} - \Delta$ where $\theta_{\max} = \max\{\theta_1, \cdots, \theta_k\}$ and where Δ is a specified positive constant. The asymptotic relative efficiency (A.R.E.) of two procedures is then the limiting ratio of the sample sizes required to achieve a preassigned minimum probability of selecting a good population. It was hoped that procedures based on ranks would be more robust in terms of A.R.E. than corresponding parametric procedures. However, it has recently been shown that the slippage configuration used to find the A.R.E. in references [5] and [1] was not least favorable for the selection of a good population (See reference [10].). Puri and Puri [9] avoided this difficulty by restricting consideration to parameter points $\boldsymbol{\theta}^{(n)} = (\theta_1^{(n)}, \cdots, \theta_k^{(n)})$ for which $\theta_{\max}^{(n)} - \theta_i^{(n)} = b_i/n^{\frac{1}{2}} + o(1/n^{\frac{1}{2}})$ for $i = 1, \cdots, k$, where the b_i are nonnegative constants.

In Section 2, selection procedures are defined which are based on two-sample estimates of shift. It is shown in Section 3 that if the underlying distribution $F(x)$ is absolutely continuous then the procedures defined in Section 2 will select a unique population. Conditions are given under which the slippage configuration is the least favorable parameter point for the selection of a good population. This result does not require restrictions on the set of translation parameters comprising the parameter space. The A.R.E. of these procedures is defined in Section 4. If we consider the procedure based on the Hodges–Lehmann estimates of shift corresponding to the two-sample F_0-scores test, it is shown that the A.R.E. of this procedure relative to the normal theory procedure of Bechhofer [2] is simply the Pitman efficiency of the two-sample F_0-scores test relative to the t-test. Hence this approach yields efficiency results which are similar to those in references [5], [1], and [9]. However, the use of estimates in the definition of the selection procedure has the advantage of eliminating the difficulties concerning the least favorable parameter point.

2. Procedures. Define the nk-vector $\mathbf{X} = (X_{11}, \cdots, X_{1n}, X_{21}, \cdots, X_{kn})'$ and the n-vectors $\mathbf{X}_i = (X_{i1}, \cdots, X_{in})'$, $i = 1, \cdots, k$. Let $\Psi(\mathbf{X}_i, \mathbf{X}_j)$ be an estimate of $\theta_i - \theta_j$

Received September 29, 1969; revised April 1, 1970.

[1] This research was supported by the National Science Foundation under Grant GP-8124.

with the following properties for each $\mathbf{x}_i$ and $\mathbf{x}_j$:

(a) If $\mathbf{c} = (c, \cdots, c)'$ is an n-vector, then

$$\Psi(\mathbf{x}_i+\mathbf{c}, \mathbf{x}_j) = c+\Psi(\mathbf{x}_i, \mathbf{x}_j);$$

(2.1) (b) $$\Psi(\mathbf{x}_i, \mathbf{x}_j) = -\Psi(\mathbf{x}_j, \mathbf{x}_i);$$

(c) $$\text{If} \quad x_{it} > x_{jt} \quad \text{for} \quad t = 1, \cdots, n, \quad \Psi(\mathbf{x}_i, \mathbf{x}_j) > 0.$$

Examples of estimates with these properties are:

(a) $$\bar{x}_i - \bar{x}_j,$$

(2.2) (b) $$\text{median}_{\alpha,\beta}(x_{i\alpha} - x_{j\beta}),$$

(c) $$\text{median}_{\alpha}(x_{i\alpha}) - \text{median}_{\beta}(x_{j\beta}).$$

For testing the equality of $F(x-\theta_i)$ and $F(x-\theta_j)$ we might consider the F_0-scores test which is based on the statistic $\sum_{t=1}^{n} E_{F_0}(V^{(r_t)})$ where $V^{(1)} < \cdots < V^{(2n)}$ are the order statistics of a random sample of size $2n$ from a distribution with df F_0 and where r_t is the rank of x_{it} in the joint ranking of $x_{i1}, \cdots, x_{in}, x_{j1}, \cdots, x_{jn}$. Hodges and Lehmann [4] have proposed estimates of $\theta_i - \theta_j$ based on these rank tests.

LEMMA 2.1. *If* $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ *is the Hodges–Lehmann estimate corresponding to the two-sample* F_0*-scores test,* $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ *satisfies properties* (2.1).

PROOF. Parts (a) and (b) of (2.1) are proved by equations (7.1) and (8.3) of reference [4]. Property (c) follows directly from equations (2.2), (3.3), and (7.1) of the same paper.

We now estimate $\theta_i - \theta_j$ by

$$Z_{ij}(\mathbf{x}) = \Psi_{i\cdot}(\mathbf{x}) - \Psi_{j\cdot}(\mathbf{x}),$$

where $\Psi_{i\cdot}(\mathbf{x}) = (1/k)\sum_{s=1}^{k} \Psi(\mathbf{x}_i, \mathbf{x}_s)$. Our selection procedure will select the ith population if $Z_{ij}(\mathbf{x}) > 0$ for all $j \neq i$. That is, our rule selects the population corresponding to

(2.3) $$\max\{\Psi_{1\cdot}(\mathbf{x}), \cdots, \Psi_{k\cdot}(\mathbf{x})\},$$

where $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ is assumed to satisfy properties (2.1). Note that if $\Psi(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)$, procedure (2.3) selects the population corresponding to $\max\{\varphi(\mathbf{x}_1), \cdots, \varphi(\mathbf{x}_k)\}$. In particular, if $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ is defined as in (a) of (2.2), the resulting procedure is one studied by Bechhofer [2].

Lehmann introduced the averaged estimates $Z_{ij}(\mathbf{x})$ in order to form compatible estimates of contrasts (See reference [6].). If we estimate each $\theta_i - \theta_j$ by an estimate having the properties (2.1), we would want to select the ith population if the estimate of $\theta_i - \theta_j$ is positive for each $j \neq i$. This would imply that in terms of our estimates the ith population is best. However, if $\theta_i - \theta_j$ is estimated by $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ as defined in

(b) of (2.2), configurations of x exist for which, given any i, there exists a j such that $\Psi(\mathbf{x}_i, \mathbf{x}_j) < 0$. Hence the averaged estimates are used to insure the existence of a best population.

3. The least favorable parameter point. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)'$ be in $\Theta = R^k$. Assume that the underlying distribution function $F(x)$ is absolutely continuous. Define

$$P_j(\boldsymbol{\theta}) = P[Z_{j1}(\mathbf{X}) > 0, \cdots, Z_{jj-1}(\mathbf{X}) > 0, Z_{jj+1}(\mathbf{X}) > 0, \cdots, Z_{jk}(\mathbf{X}) > 0 \mid \boldsymbol{\theta}]$$

for $j = 1, \cdots, k$.

LEMMA 3.1. *Let s be a positive integer such that $1 \leqq s \leqq k$. If $\boldsymbol{\theta}^*(\boldsymbol{\theta}) = (\theta_1, \cdots, \theta_{s-1}, \theta_s + a, \theta_{s+1}, \cdots, \theta_k)'$ and $v \neq s$, then*

$$P_v(\boldsymbol{\theta}) \geqq (\leqq) P_v(\boldsymbol{\theta}^*(\boldsymbol{\theta})) \tag{3.1}$$

as $a \geqq (\leqq) 0$.

PROOF. Let

$$\begin{aligned} Y_{it} &= X_{it} \quad \text{for} \quad t = 1, \cdots, n; i = 1, \cdots, k; & i \neq s, \\ &= X_{it} + a \quad \text{for} \quad t = 1, \cdots, n; & i = s. \end{aligned} \tag{3.2}$$

Application of (a) and (b) of (2.1) shows that

$$\begin{aligned} \Psi(\mathbf{Y}_i, \mathbf{Y}_j) &= \Psi(\mathbf{X}_i, \mathbf{X}_j) \quad \text{if} \quad i = j \quad \text{or if} \quad i \neq s \quad \text{and} \quad j \neq s, \\ &= \Psi(\mathbf{X}_i, \mathbf{X}_j) + a \quad \text{if} \quad i = s \neq j, \\ &= \Psi(\mathbf{X}_i, \mathbf{X}_j) - a \quad \text{if} \quad j = s \neq i. \end{aligned} \tag{3.3}$$

It follows that

$$\begin{aligned} Z_{ij}(\mathbf{Y}) &= Z_{ij}(\mathbf{X}) \quad \text{if} \quad i = j \quad \text{or if} \quad i \neq s \quad \text{and} \quad j \neq s, \\ &= Z_{ij}(\mathbf{X}) + a \quad \text{if} \quad i = s \neq j, \\ &= Z_{ij}(\mathbf{X}) - a \quad \text{if} \quad j = s \neq i. \end{aligned} \tag{3.4}$$

Thus if $a \geqq (\leqq) 0$ and $v \neq s$,

$$\begin{aligned} P_v(\boldsymbol{\theta}) &= P[Z_{v1}(\mathbf{X}) > 0, \cdots, Z_{vk}(\mathbf{X}) > 0 \mid \boldsymbol{\theta}] \\ &\geqq (\leqq) P[Z_{v1}(\mathbf{Y}) > 0, \cdots, Z_{vk}(\mathbf{Y}) > 0 \mid \boldsymbol{\theta}] = P_v(\boldsymbol{\theta}^*(\boldsymbol{\theta})). \end{aligned}$$

The following lemma shows that when $F(z)$ is absolutely continuous, any decision rule in the class (2.3) selects exactly one population.

LEMMA 3.2. *For $j \neq i$, $Z_{ij}(\mathbf{X})$ is absolutely continuous.*

PROOF. Let A be any set on the real line with Lebesgue measure zero and let $Q = \{\mathbf{x} \mid Z_{ij}(\mathbf{x}) \in A\}$. Define $U_2 = X_{i2} - X_{i1}, \cdots, U_n = X_{in} - X_{i1}$. From (3.4) it follows that $Z_{ij}(\mathbf{X}) = X_{i1} + Z_{ij}((X_{11}, \cdots, X_{i-1n}, 0, U_2, \cdots, U_n, X_{i+11}, \cdots, X_{kn})')$.

Consider the line in R^{kn} defined by letting $x_{11} = x_{11}^0, \cdots, x_{i-1n} = x_{i-1n}^0$, $x_{i2} = x_{i1} + u_2^0, \cdots, x_{in} = x_{i1} + u_n^0$, $x_{i+11} = x_{i+11}^0, \cdots, x_{kn} = x_{kn}^0$ where $x_{11}^0, \cdots, x_{i-1n}^0$, $u_2^0, \cdots, u_n^0$, $x_{i+11}^0, \cdots, x_{kn}^0$ are fixed constants. The section of Q by any such line has Lebesgue measure zero. Hence Q has Lebesgue measure zero. By the absolute continuity of $F(x)$ it follows that $P(Q) = 0$.

THEOREM 3.1. *For any procedure in the class of rules* (2.3),

$$\inf_{\theta \in \Theta} P[\textit{select a good population} \mid \boldsymbol{\theta}]$$
$$= P[\textit{select the } k\text{th } \textit{population} \mid \theta_k - \Delta = \theta_{k-1} = \cdots = \theta_1]$$

where the latter probability is independent of the parameter θ_k.

PROOF. Fix $\boldsymbol{\theta}$ in Θ. A procedure in (2.3) is not altered by the renumbering of the populations. Hence without loss of generality we assume that $\theta_k \geqq \theta_{k-1} \geqq \cdots \geqq \theta_1$. Let r denote the smallest positive integer such that $\theta_r > \theta_k - \Delta$. Now P[select a good population $\mid \boldsymbol{\theta}] = \sum_{j=r}^k P_j(\boldsymbol{\theta})$. Define $\boldsymbol{\theta}^{**}$ such that $\theta_i^{**} = \theta_i$ for $i = r, \cdots, k$ and $\theta_i^{**} = \theta_k - \Delta$ for $i = 1, \cdots, r-1$. Application of Lemma 3.1 for $s = 1, \cdots, r-1$ shows that

$$\textstyle\sum_{j=r}^k P_j(\boldsymbol{\theta}) \geqq \sum_{j=r}^k P_j(\boldsymbol{\theta}^{**}).$$

Lemma 3.2 yields that

$$\textstyle\sum_{j=r}^k P_j(\boldsymbol{\theta}^{**}) = 1 - \sum_{j=1}^{r-1} P_j(\boldsymbol{\theta}^{**}).$$

Define $\boldsymbol{\theta}^0$ such that $\theta_k^0 = \theta_k$ and $\theta_i^0 = \theta_k - \Delta$ for $i = 1, \cdots, k-1$. Application of Lemma 3.1 for $s = r, \cdots, k-1$ shows that

$$\textstyle 1 - \sum_{j=1}^{r-1} P_j(\boldsymbol{\theta}^{**}) \geqq 1 - \sum_{j=1}^{k-1} P_j(\boldsymbol{\theta}^0) = P[\text{select the } k\text{th population} \mid \boldsymbol{\theta}^0].$$

Since a procedure in the class (2.3) depends only on the $Z_{ij}(\mathbf{X})$'s, it follows from equation (3.4) that P[select the kth population $\mid \theta_k - \Delta = \theta_{k-1} = \cdots = \theta_1$] is independent of the parameter θ_k.

4. Asymptotic efficiency. In this section we compare two procedures S_1 and S_2 in the class of rules (2.3). For each procedure S_j, the sample size $n_j = n_j(\Delta, \gamma, F)$ is determined so that

$$\inf_{\theta \in \Theta} P[\text{select a good population} \mid \boldsymbol{\theta}, \Delta, F] = \gamma, \tag{4.1}$$

where $1 > \gamma > 1/k$. Assume that for each procedure S_j, the sample size n_j and its corresponding Δ have the relationship that as $n_j \to \infty$,

$$\Delta^{(n_j)} = \frac{K_j}{(n_j)^{\frac{1}{2}}} + o\left(\frac{1}{(n_j)^{\frac{1}{2}}}\right), \tag{4.2}$$

where K_j is a positive constant. We seek an asymptotic comparison of the procedures and hence we examine the ratio of the respective sample sizes required by (4.1) under a sequence of Δ values approaching zero.

For procedure S_1, consider the sample size $n_1(\Delta^{(n_2)}, \gamma, F)$, abbreviated $n_1(n_2)$, determined by (4.1) where the goodness criterion $\Delta^{(n_2)}$ is given by (4.2) with $j = 2$. As $n_2 \to \infty$, $\Delta^{(n_2)} \to 0$ and $n_1(n_2) \to \infty$. Thus (4.2) implies that

$$\Delta^{(n_2)} = \frac{K_1}{[n_1(n_2)]^{\frac{1}{2}}} + o\left(\frac{1}{[n_1(n_2)]^{\frac{1}{2}}}\right). \tag{4.3}$$

The asymptotic efficiency of S_1 relative to S_2 is then defined to be

$$\text{ARE}(S_1, S_2; F) = \lim_{n_2 \to \infty} \frac{n_2}{n_1(n_2)} = K_2{}^2/K_1{}^2. \tag{4.4}$$

This Pitman-type efficiency for selection procedures was considered by Lehmann [5].

THEOREM 4.1. *For procedure S_1, let $\Psi(\mathbf{x}_i, \mathbf{x}_j)$ be the Hodges–Lehmann estimate of $\theta_i - \theta_j$ corresponding to the two-sample F_0-scores test and for procedure S_2, let $\Psi(\mathbf{x}_i, \mathbf{x}_j) = \bar{x}_i - \bar{x}_j$. If $F(x)$ is the distribution function of an absolutely continuous random variable with finite variance, σ^2, and if the regularity conditions of Lemma* 7.2 *of* [8] *are satisfied,*

$$\text{ARE}(S_1, S_2; F) = \sigma^2 B^2/A^2, \tag{4.5}$$

where $J = F_0{}^{-1}$, $A^2 = \int_0^1 J^2(x)\,dx - (\int_0^1 J(x)\,dx)^2$ and $B = \int J[F(x)]f^2(x)\,dx$.

PROOF. Let $(U_1, \cdots, U_{k-1})$ be a $k-1$ variate normal random variable with $E(U_i) = 0$, $\text{Var}(U_i) = 1$ for $i = 1, \cdots, k-1$ and $\text{Cov}(U_i, U_r) = \frac{1}{2}$ for $i \neq r$. Let d be determined by $G(d/2^{\frac{1}{2}}, \cdots, d/2^{\frac{1}{2}}) = \gamma$ where G is the distribution function of $(U_1, \cdots, U_{k-1})$. By Lemma 1 of reference [5], it is seen that S_2 satisfies (4.2) with $K_2 = \sigma d$. There remains to show that S_1 satisfies (4.2) with $K_1 = Ad/B$.

Theorem 3.1 and equation (4.1) imply that for procedure S_1,

$$\begin{aligned}\gamma &= \inf_{\boldsymbol{\theta} \in \Theta} P[\text{select a good population} \mid \boldsymbol{\theta}, \Delta, F] \\ &= P[Z_{1k}^{(1)} \leqq 0, \cdots, Z_{k-1k}^{(1)} \leqq 0 \mid \theta_k - \Delta = \theta_{k-1} = \cdots = \theta_1, \Delta, F].\end{aligned}$$

By Theorem 3.1 of reference [3] and the fact that the convergence is uniform in the arguments of the cumulative distribution function, it follows that

$$\lim_{n_1 \to \infty} P[U_i \leqq \Delta^{(n_1)}(\tfrac{1}{2} n_1 B^2/A^2)^{\frac{1}{2}} \text{ for } i = 1, \cdots, k-1] = \gamma.$$

This equation is satisfied if and only if (4.2) is satisfied for $j = 1$ with $K_1 = Ad/B$.

Note that (4.5) is just the Pitman efficiency of the two-sample F_0-scores test relative to the t-test. This efficiency result corresponds to those in references [5], [1], and [9]. The advantage of the approach taken here, is that it eliminates the difficulties concerning the least favorable parameter point without the use of artificial restrictions on the parameter space.

Acknowledgments. The author would like to thank R. V. Hogg and J. D. Cryer for their comments on this paper and also G. G. Woodworth for his helpful suggestions.

REFERENCES

[1] BARTLETT, N. S. and GOVINDARAJULU, Z. (1968). Some distribution free statistics and their application to the selection problem. *Ann. Inst. Statist. Math. Tokyo* **20** 79–97.

[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] BHUCHONGKUL, S. and PURI, M. L. (1965). On the estimation of contrasts in linear models. *Ann. Math. Statist.* **36** 198–202.

[4] HODGES, J. L., Jr. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[5] LEHMANN, E. L. (1963). A class of selection procedures based on ranks. *Math. Ann.* **150** 268–275.

[6] LEHMANN, E. L. (1963). Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957–966.

[7] MCDONALD, G. C. and GUPTA, S. S. (1969). On some classes of selection procedures based on ranks. Mimeograph Series No. 190, Department of Statistics, Purdue Univ.

[8] PURI, M. L. (1964). Asymptotic efficiency of a class of c-sample tests. *Ann. Math. Statist.* **35** 102–121.

[9] PURI, M. L. and PURI, P. S. (1969). Multiple decision procedures based on ranks for certain problems in analysis of variance. *Ann. Math. Statist.* **40** 619–632.

[10] RIZVI, M. H. and WOODWORTH, G. G. (1968). On selection procedures based on ranks: counterexamples concerning least favorable configurations. Technical Report No. 114, Department of Statistics, Stanford Univ.

[11] SOBEL, M. (1967). Nonparametric procedures for selecting the t populations with the largest α-quantiles. *Ann. Math. Statist.* **38** 1804–1816.