The Annals of Statistics
1979, Vol. 7, No. 6, 1321–1328

MOST ECONOMICAL ROBUST SELECTION PROCEDURES FOR LOCATION PARAMETERS[1]

By S. R. Dalal and W. J. Hall
University of Rochester

Consider samples of size n from each of k symmetric populations, differing only in their location parameters. The decision problem is to select the best population—the one with the largest location parameter—with control on the probability of correct selection (PCS) whenever the largest parameter is at least Δ units larger than all others, and whenever the common error distribution belongs to a specified neighborhood of the standard normal. It is shown that, if the sample size n is chosen according to a formula given herein, and Huber's M-estimate is applied to each of the k samples with the population having the largest estimate being selected as best, that the PCS goal is achieved asymptotically (as $\Delta \downarrow 0$)—the procedure is *robust*. Moreover, no other selection procedure can achieve this goal asymptotically with a smaller sample size—the procedure is *most economical*. Comparisons with other procedures are given. These results are based on a uniform asymptotic normality theorem for Huber's M-estimate, contained herein.

1. Introduction and summary. Let X_{ij} ($j = 1, \cdots, n$; $i = 1, \cdots, k$) be independent observations from k populations with respective distribution functions (df's) $F(x - \theta_i)$, and let $\theta_{[1]} \leqslant \cdots \leqslant \theta_{[k]}$ denote the ordered θ_i's. We consider the problem of selecting the 'best' population, namely the one with the largest location parameter θ; the methods are readily extended to other ranking and selection goals, as introduced by Bechhofer (1954) and treated recently by Ghosh (1973); see also Gibbons, et al., (1977).

Let Δ (> 0) and P^* ($> 1/k$) be specified and write $S = S_\Delta$ for the subset of the parameter space R^k of $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ where $\theta_{[k]} \geqslant \theta_{[k-1]} + \Delta$. The problem is to choose a value, say N, for the (common) sample size n and a *selection procedure* for choosing the best population which assures a PCS (*probability of correct selection*) of at least P^* whenever $\boldsymbol{\theta} \in S$. When F is Φ—and the errors are thus standard normal—Bechhofer (1954) showed that, by choosing $N = d^2/\Delta^2$ (or $\geqslant$) with $d = P^{-1}(P^*)$ and

$$(1) \qquad P(d) \equiv \int \Phi^{k-1}(x + d)\, d\Phi(x),$$

the procedure which selects the population with the largest sample mean meets this *goal*: $\text{PCS}_N(\Phi, \boldsymbol{\theta}) \geqslant P^*$ when $\boldsymbol{\theta} \in S_\Delta$, with equality at a *least favorable configuration* (lfc) where all but one of the populations have the same θ-value and the remaining

Received October 1977; revised August 1978.
[1]Supported by the U. S. Army Research Office and revision supported by N.S.F. grant No. MCS78-02160.
AMS 1970 *subject classifications.* Primary 62F07; secondary 62G35.
Key words and phrases. Selection procedures, robust procedures, Huber's M-estimate, location parameters, uniform asymptotic normality.

population's θ-value is Δ units larger. Hall (1959) showed this procedure to be *most economical*, in the sense that no competing procedure could meet the goal with a smaller sample size.

Other authors have introduced various competing procedures, especially procedures based on various nonparametric statistics (e.g., Lehmann (1963), Randles (1970) and Ghosh (1973); other references are in Ghosh). By selecting as the best population the one with the largest location estimate, now using a nonparametric estimate of location, the procedures are not so sensitive to the assumption of normal errors. However, they are not fully nonparametric, even asymptotically, in that the rule for choosing N is: $N \sim d^2\tau^2/\Delta^2$ (as $\Delta\downarrow 0$) where τ^2/n is the (asymptotic) variance of the location estimator T_n used by the selection rule. And $\tau = \tau(F)$, so now assumptions about F are needed in order to determine N. For example, one may use the Hodges-Lehmann location estimate and then $\tau^2 = (2(3)^{\frac{1}{2}}\int f^2)^{-2}$, assuming F has density f (Ghosh, 1973); specifically, $\tau^2(\Phi) = \pi/3$. The ratio of N's for procedures based on a location estimator T_n and on the sample mean, respectively, is asymptotically $\tau^2(\Phi)$—assuming both N's were chosen to meet the PCS requirement at $F = \Phi$. Such ARE's are discussed by Ghosh.

Asymptotically, the choice of $N = N(\Delta) = d^2\tau^2(\Phi)/\Delta^2 + o(\Delta^{-2})$ works for other F's with the same or smaller $\tau(F)$ in that $\lim_{\Delta\downarrow 0}\inf_{\theta\varepsilon S_\Delta}\mathrm{PCS}_{N(\Delta)}(F, \theta) \geqslant P^*$ for each such F. However, one can *not* claim that, for some small $\Delta = \Delta(\varepsilon)$, $\inf_{\theta\varepsilon S}\mathrm{PCS}_{N(\Delta)}(F, \theta) > P^* - \varepsilon$ for each such F; for this, we would need the $o(\Delta^{-2})$ term above to be uniform in F, thus allowing the insertion of an "inf over F" in front of the PCS. Thus, the asymptotic formulation of the problem is: *find a sequence of selection procedures* (one for each sample size) *and a sample size formula* $N = N(\Delta)$ *such that*

$$\liminf_{\Delta\downarrow 0}\inf_{F\in\mathcal{F},\, \theta\in S_\Delta}\mathrm{PCS}_{N(\Delta)}(F, \theta) \geqslant P^* \ (> 1/k). \tag{2}$$

A procedure satisfying (2), for some suitable $\mathcal{F}$ containing Φ, is said to be *asymptotically robust* at $\Phi \in \mathcal{F}$—since we have P^*-protection at F's near Φ. If every other procedure satisfying (2), but with $N = N'(\Delta)$, has the property $\liminf_{\Delta\downarrow 0}[N'(\Delta)/N(\Delta)] \geqslant 1$, then the procedure is *asymptotically most economical* for $\mathcal{F}$ as well—N is minimal.

Thus our approach differs in that we set out to choose a suitable family $\mathcal{F}$ of possible error distributions and determine N (minimally) so as to meet the PCS goal for every $F \in \mathcal{F}$—specifically for a least favorable F^0 in $\mathcal{F}$. This continues the same minimax spirit of these selection procedures: we not only guard against a least favorable configuration of location parameter values, but against a least favorable error distribution as well—the roles of θ and F are treated in the same way. For this development, we follow the 'contamination' or 'gross error' model approach of Huber (1964).

Let $\mathcal{F} = \{F | F = (1 - \gamma)\Phi + \gamma H,\ H$ an arbitrary df symmetric at $0\}$, for specified $\gamma \in [0, 1)$. This is a (symmetric) 'contamination neighborhood' of the ideal model Φ, allowing for a proportion γ or less of 'gross errors' in the model. Huber's

M-estimate T^0 is designed for this setting: T^0 is implicitly defined so that it is the mean of the Winsorized sample, having 'pulled in' to $T^0 \pm c$ all observations more than c units away from T^0. The censoring point $c = c(\gamma)$ is implicitly defined by $(1-\gamma)^{-1} = 1 - 2\Phi(-c) + 2\phi(c)/c$. This estimator is asymptotically normal whenever F is symmetric at 0: $n^{\frac{1}{2}}(T_n^0 - \theta)/\sigma(F) \to N(0, 1)$ in law, where

$$\sigma^2(F) = E_F(X^2 \wedge c^2)/[2F(c) - 1]^2. \tag{3}$$

Indeed, T_n^0 is the mle under the least favorable F^0, which has density $(1-\gamma)\phi(x)\exp\{\frac{1}{2}[(|x| - c)^+]^2\}$ and is in $\mathcal{F}$, and $\sigma_0^2 \equiv \sigma(F^0)^2 = 1/[1 - 2(1 - \gamma)\phi(c)/c]$. Huber (1964, Section 6) proved an *asymptotic minimax theorem*: (a) *F^0 is least favorable in that $\sigma(F) \leqslant \sigma_0$ for all $F \in \mathcal{F}$, and* (b) *if T_n is any sequence of (location equivariant) estimators for which $n^{\frac{1}{2}}(T_n - \theta)/\sigma \to N(0, 1)$ for some σ when the error distribution is F^0, then $\sigma_0 \leqslant \sigma$.*

We shall show that a selection procedure based on Huber's M-estimate T^0, with sample size chosen to meet the PCS goal (asymptotically) at F^0, meets the goal for every $F \in \mathcal{F}^0$, an equicontinuous (at c) subset of $\mathcal{F}$, and the appropriate asymptotic sample size N is $d^2\sigma_0^2/\Delta^2$; moreover, this N is minimal. We conclude therefore that the procedure is (asymptotically) a *most-economical robust selection procedure.*

The ratio of sample sizes for this procedure relative to Bechhofer's $\bar{X}$-procedure is thus σ_0^2. This ratio differs conceptually from the ARE (see Ghosh); it measures how much the sample size needs to be increased to expand the P^*-protection from Φ to $\mathcal{F}^0$, whereas the ARE compares two procedures both evaluated at Φ. A table of σ_0^2 appears below:

γ	c	σ_0^2	c	γ	σ_0^2
0.5	0.436	5.930	0.50	.4417	4.678
0.2	0.862	2.045	0.75	.2591	2.469
0.1	1.140	1.490	1.00	.1428	1.709
0.05	1.398	1.256	1.25	.0749	1.371
0.02	1.717	1.116	1.50	.0376	1.199
0.01	1.945	1.065	1.75	.0181	1.107
0.005	2.163	1.037	2.00	.0084	1.057
0.002	2.436	1.017	2.25	.0038	1.029
0.001	2.633	1.010	2.50	.0016	1.014
0.0005	2.822	1.005	2.75	.0007	1.007
0.0002	3.062	1.002	3.00	.0003	1.003

Thus, it requires a $6\frac{1}{2}\%$ increase in sample size to allow for 1% contamination; and if we censor at $c = \pm 1.5$ units from T^0, we allow for almost 4% contamination with a 20% increase in sample size. The PCS of our procedure when F is actually Φ is at least $P(d\sigma_0)$, which is greater than P^*, while the PCS for the $\bar{X}$-procedure when $F \neq \Phi$ but $\in \mathcal{F}^0$ may be smaller than P^*; indeed, if $\gamma \geqslant 1 - (2\pi/e)^{-\frac{1}{2}} \doteq 0.34$, $\mathcal{F}$

contains the standard Cauchy distribution F_C, and at F_C (and a lfc) the PCS of the $\bar{X}$ procedure tends to $1/k$.

To prove our optimality claims, we first need to strengthen Huber's asymptotic normality result quoted above to *uniform* (in $\mathcal{F}^0$) *asymptotic normality*. This is the subject of Section 2. We first comment on an implicit restriction.

A selection problem is a *multiple-decision problem*, but for asymptotic investigations we (and others) confine attention to *selection procedures based on a* (location equivariant) *estimator sequence* $T = \{T_n\}$: T_n is a numerical-valued function of n i.i.d. observations and is applied to each of the k samples separately, yielding $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$, say, and population i is selected as best iff $\hat{\theta}_i$ is the largest (with ties broken at random). Whether or not this class of selection procedures is (asymptotically) *essentially complete* is not known, though it seems likely. (For example, one might consider a compound estimator T_n depending on data from all populations, such as Stein's, but Stein's is asymptotically equivalent to $T_n = \bar{X}_n$, and hence does not provide a counterexample.)

Finally, it should be recognized that the known scale assumption regarding the core distribution Φ of $\mathcal{F}$ is a limitation which is frequently unrealistic. However, it is only an assumption about Φ, and therefore is not so specific about F. And for the purposes of selecting a suitable sample size it only need be a bound, or conservative guess, at scale. To totally avoid an assumption on scale, it is necessary to go to a two-stage or sequential procedure (see Gibbons, et al., 1977). It should be possible to adapt the parametric selection procedures of this sort to accommodate a robust estimate of scale (and location), but the most-economical property will not likely be achievable, since robust optimality is not yet available in estimation except when scale is assumed known.

2. Uniform asymptotic normality of Huber's *M*-estimate. Let $(X_1, \cdots, X_n)$ represent a sample of size n from a population with df $F(x - \theta)$, $F \in \mathcal{F}$, and the location parameter θ is to be estimated. We confine attention to Huber's M-estimator with $\psi(x) = \max(-c, \min(x, c))$ where $c = c(\gamma)$ (see Section 1). The corresponding M-estimator T^0 is the (midpoint of the interval of) solution(s) of $\Sigma_{i=1}^n \psi(x_i - T^0) = 0$. Huber has shown that $n^{\frac{1}{2}}(T^0 - \theta)$ is asymptotically normal for $F \in \mathcal{F}$ (and even more generally); we show here that this asymptotic normality is uniform in F for F in a subclass $\mathcal{F}^{0'}$ of $\mathcal{F}$. Similar methods could be used for other M-estimators with bounded, continuous and monotone ψ-functions.

Let $\mathcal{F}^0$ be an equicontinuous (at c) subclass of $\mathcal{F}$—i.e.,

$$\lim_{h \downarrow 0} \sup_{F \in \mathcal{F}^0} [F(c + h) - F(c - h)] = 0. \tag{4}$$

(Later, $\mathcal{F}^0$ must be big enough to include Huber's least favorable F^0—normal in the center, exponential in the tails, and continuous at c.) Denote (as in Huber) $\lambda_F(t) = \int \psi(x - t)\, dF(x)$. For symmetric $F \in \mathcal{F}$, $\lambda_F(\cdot)$ is strictly decreasing, and 0 is readily seen to be the unique solution of $\lambda_F(t) = 0$. Additional properties of λ_F are:

LEMMA 1. *Let* $\mathcal{F}_*$ *be a class of* df's *symmetric at* 0 *and* $\mathcal{F}_*^0$ *an equicontinuous at* c *subclass* (e.g., $\mathcal{F}$ *and* $\mathcal{F}^0$).

(i) *For* $F \in \mathcal{F}_*$ *and* $0 < |h| < c$, $\lambda_F(h)/h + \int_{-c}^{c} dF(x) = -\int_{c-h}^{c+h}((F(x) - F(c))/h)dx$.

(ii) $\lim_{h\downarrow 0}\sup_{F\in\mathcal{F}_*^0}|\lambda_F(h)/h - \lambda_F'(0)| = 0$.

(iii) *For* $F \in \mathcal{F}_*^0$, $\lambda_F'(0) = -\int_{-c}^{c} dF(x) = 2F(-c) - 1$.

PROOF. For (i), consider $0 < h < c$; then

$$\lambda_F(h) + h\int_{-c}^{c} dF(x) = \int[\psi(x-h) - \psi(x) + hI(|x| \leqslant c)]\, dF(x)$$

since $E_F\psi = 0$

$$= \int_{-c}^{-c+h}(-x + h - c)\, dF + \int_{c}^{c+h}(x - h - c)\, dF$$

$$= \int_{c-h}^{c+h}(x - c)\, dF + h[2F(c) - F(c-h) - F(c+h)]$$

by symmetry

$$= -\int_{c-h}^{c+h}[F(x) - F(c)]\, dx \qquad \text{upon integrating by parts,}$$

and similarly for $-c < h < 0$.

Now note from (i) that $|\lambda_F(h)/h + \int_{(-c,c)} dF| \leqslant h^{-1}\int_{(c,c+h)}[F(x) - F(c)]\, dx + h^{-1}\int_{(c-h,c)}[F(c) - F(x)]\, dx \leqslant F(c+h) - F(c-h)$; then (ii) and (iii) follow from (i).

In what follows, (ii) is needed and equicontinuity is sufficient for it. That it is not necessary can be seen by considering a symmetric F with a density for positive x which is symmetric at c; then the term in "$|\cdots|$" in (ii) is identically zero, but the left-hand side of (4) need not be zero. (For $F \in \mathcal{F}$, we apply the above reasoning to H.) Nevertheless, equicontinuity at c is a simpler assumption to impose. We can and will assume (as noted above) that Huber's least favorable F is in $\mathcal{F}^0$.

Another way of assuring (ii) is to replace Huber's ψ by a ψ with a uniformly continuous derivative ψ'; then $\lambda_F'(0)$ is $-\int\psi'(x)\, dF(x)$ (assumed finite). For bounded and monotone ψ's, this additional condition would enable proof of Theorem 1 (below) for the corresponding M-estimators, with $\mathcal{F}^0$ enlarged to $\mathcal{F}$. (Huber considered such assumptions in his Section 3.) But we need the minimax property of Huber's T^0 and ψ in what follows, and therefore limit our attention hereafter to this case, with the equicontinuity assumption.

We also need a 'double subscript version' of the Berry-Esseen theorem (e.g., Feller (1971), page 542):

LEMMA 2. (Berry-Esseen). *For each* $n = 1, 2, \cdots$, *let* X_{kn} $(k = 1, \cdots, n)$ *be* i.i.d. rv's *with* $EX_{kn} = 0$, $EX_{kn}^2 = \sigma_n^2$, $E|X_{kn}|^3 = \lambda_n$, *and let* F_n *be the* df *of the normalized sum* $n^{-\frac{1}{2}}(X_{1n} + \cdots + X_{nn})/\sigma_n$. *Then* $\sup_x|F_n(x) - \Phi(x)| \leqslant 3\lambda_n\sigma_n^{-3}n^{-\frac{1}{2}}$.

We are now ready for the uniform convergence theorem ($\sigma(F)$ was defined in (3)):

THEOREM 1. $\sup_{F \in \mathcal{F}^0} \sup_t |P_F\{n^{\frac{1}{2}}(T_n^0 - \theta)/\sigma(F) \leqslant t\} - \Phi(t)| \to 0$ *as* $n \to \infty$.

PROOF. Take $\theta = 0$. By use of Polya's device, it is sufficient to confine attention to t in a compact set $\mathfrak{T}$.

Write $s_n(t, F) = n^{-\frac{1}{2}} t\sigma(F)$ and $\sigma_n^2(t, F) = \operatorname{Var}_F \psi(X - s_n)$. By the uniform continuity of ψ, $\sigma_n^2 \to E_F \psi^2$ (uniformly, for $F \in \mathcal{F}^0$ and $t \in \mathfrak{T}$) $\geqslant (1 - \gamma) E_\Phi \psi^2 > 0$, so that σ_n is uniformly bounded away from zero for all large n. Writing $u_n(t, F) = -n^{\frac{1}{2}} \sigma_n^{-1} \lambda_F(s_n)$, and using Lemma 1, we can show that $u_n \to t$, likewise uniformly.

Now $P_F\{n^{\frac{1}{2}} T_n^0/\sigma(F) \leqslant t\} = P_F\{T_n^0 \leqslant s_n\} \leqslant P_F\{\Sigma_{i=1}^n \psi(X_i - s_n) \leqslant 0\} = P_F\{n^{-\frac{1}{2}} \sigma_n^{-1} \Sigma_{i=1}^n [\psi(X_i - s_n) - \lambda_F(s_n)] \leqslant u_n\} \leqslant \Phi(u_n) + 24\sigma_n^{-3} n^{-\frac{1}{2}}$ (by Lemma 2 since the summands are bounded and have third absolute moments at most $8c^3$) $\leqslant \Phi(t) +$ for n sufficiently large (uniformly, for $F \in \mathcal{F}^0$ and $t \in \mathfrak{T}$), by the uniformity properties noted above.

Similarly, $P_F\{n^{\frac{1}{2}} T_n^0/\sigma(F) < t\} \geqslant P_F\{\Sigma \psi(X_i - s_n) < 0\} \geqslant \Phi(t) - \varepsilon$ uniformly.

We now extend this to the selection procedure setting. Suppose we have k populations, sampled independently, each with possibly differing θ-values but otherwise identical df's $F \varepsilon \mathcal{F}$; T_{jn}^0 is Huber's estimator of θ_j based on n observations from population j; $P(\cdot)$ was defined in (1). We then have

COROLLARY. $\sup_{F \in \mathcal{F}^0} \sup_t |P_F\{n^{\frac{1}{2}}(T_{jn}^0 - \theta_j - T_{kn}^0 + \theta_k)/\sigma(F) \leqslant t$ *for all* $j < k\}$ $- P(t)| \to 0$ *as* $n \to \infty$.

PROOF. Take $\theta_1 = \cdots = \theta_k = 0$ and write $Z_{jn} = n^{\frac{1}{2}} T_{jn}^0/\sigma(F)$, with df F_n (under F). We need to consider $\sup_F \sup_t |D(t, F)|$ where $D(t, F) = \int F_n^{k-1}(x + t)\, dF_n(x) - \int \Phi^{k-1}(x + t)\, d\Phi(x)$. But since $F_n = \Phi + o_u(1)$ (uniformly $o(1)$, in F and t), likewise $F_n^{k-1} = \Phi^{k-1} + o_u(1)$, and therefore $D(t, F) = \int \Phi^{k-1}(x + t)\, d(F_n - \Phi)(x) + o_u(1) = \int [\Phi(x) - F_n(x)]\, d\Phi^{k-1}(x + t) + o_u(1) = o_u(1)$.

REMARK. The F's could vary from population to population so long as $\sigma(F)$ remained the same in all populations; in particular, F can vary arbitrarily (in a symmetric way) outside of an open interval containing $[-c, c]$.

3. Asymptotically optimal selection procedures. Let $T = \{T_n\}$ be any (translation equivariant) sequence of location estimators, and let T also denote the selection procedure based on T. We suppose the subscripts are arranged so that $\theta_1 \leqslant \cdots \leqslant \theta_k$. Let $\boldsymbol{\theta}_\Delta$ represent any $\boldsymbol{\theta}$ for which $\theta_k = \theta_{k-1} + \Delta$ and $\theta_1 = \theta_{k-1}$. Then $\boldsymbol{\theta}_\Delta$ is a lfc—i.e., $\inf_{S_\Delta} \mathrm{PCS}_n(F, \boldsymbol{\theta}) = \mathrm{PCS}_n(F, \boldsymbol{\theta}_\Delta)$. This is so since T forms a stochastically ordered family w.r.t. θ, and a theorem of Alam and Rizvi (1967) then implies $\boldsymbol{\theta}_\Delta$ is a lfc (also noted by Ghosh, 1973). Hence, we can and do confine attention to the parameter vector $\boldsymbol{\theta}_\Delta$ hereafter.

Let $N_{P^*}(T, \mathcal{F}, \Delta)$ denote the smallest sample size n for which $\inf_{F \in \mathcal{F}} \inf_{\boldsymbol{\theta} \in S_\Delta} \mathrm{PCS}_n(F, \boldsymbol{\theta}) \geqslant P^*$ when using the selection procedure T. An asymptotic formula for $N(T, F, \Delta) = N_{p^*}(T, \{F\}, \Delta)$ is now established (it is actually only needed at F^0, Huber's least favorable F, in what follows):

LEMMA 3. *If $T = \{T_n\}$ is translation equivariant and if $n^{\frac{1}{2}}(T_n - \theta)/\sigma \to N(0, 1)$ under F as $n \to \infty$, then $N(T, F, \Delta) \sim d^2\sigma^2/\Delta^2$ as $\Delta \downarrow 0$ (where $P(d) = P^*$, fixed).*

PROOF. Since ties can be ignored asymptotically, we have

$$\begin{aligned} \mathrm{PCS}_n(F, \boldsymbol{\theta}_\Delta) &= P_F\left\{ n^{\frac{1}{2}} \frac{T^0_{jn} - T^0_{kn}}{\sigma(F)} \leqslant 0 \quad \text{for all } j \right\} \\ &= P_F\left\{ n^{\frac{1}{2}} \frac{T^0_{jn} - \theta - T^0_{kn} + \theta + \Delta}{\sigma(F)} \leqslant n^{\frac{1}{2}} \frac{\Delta}{\sigma(F)} \quad \text{for all } j \right\} \\ &= P\left\{ n^{\frac{1}{2}} \Delta/\sigma(F) \right\} + o(1) \end{aligned}$$

by a simpler (nonuniform) version of the corollary, and this has limit $P(d) = P^*$ iff $n^{\frac{1}{2}}\Delta/\sigma(F) \to d$.

We now recall Huber's minimax results, quoted in Section 1. We use (a) in Theorem 2 (and Lemma 4) below and use (b) in Theorem 3. We first need:

LEMMA 4. *Given $\varepsilon > 0$, there exists Δ_ε such that for $\Delta < \Delta_\varepsilon$*

$$N_{P^*-\varepsilon}(T^0, F^0, \Delta) \leqslant N_{P^*-\varepsilon}(T^0, \mathcal{F}^0, \Delta) \leqslant N_{P^*}(T^0, F^0, \Delta).$$

PROOF. Writing $N(\Delta) = N_{P^*}(T^0, F^0, \Delta)$, we can find Δ_ε so that $N(\Delta_\varepsilon)$ is large and $\mathrm{PCS}_{N(\Delta)}(F) \geqslant P[N(\Delta)^{\frac{1}{2}}\Delta/\sigma(F)] - \frac{1}{2}\varepsilon$ for all $\Delta \leqslant \Delta_\varepsilon$ and all $F \in \mathcal{F}^0$ (by the corollary). But $\sigma(F) \leqslant \sigma_0$ for all $F \in \mathcal{F}^0$ by (a) of Huber's minimax theorem (see Section 1), and $P(\cdot)\uparrow$; also $N(\Delta)^{\frac{1}{2}}\Delta/\sigma_0 \to d$ (Lemma 3). Therefore $\mathrm{PCS}_{N(\Delta)}(F, \boldsymbol{\theta}_\Delta) \geqslant P(N(\Delta)^{\frac{1}{2}}\Delta/\sigma_0) - \frac{1}{2}\varepsilon \geqslant P(d) - \varepsilon = P^* - \varepsilon$ for Δ sufficiently small. Hence, this $N(\Delta) \geqslant N_{P^*-\varepsilon}(T^0, F, \Delta)$, proving the second inequality (the first being trivial).

We are now ready for the final two theorems; the first establishes the (asymptotic) sample size formula for guaranteeing the PCS goal throughout $\mathcal{F}^0$ when using the selection procedure T, and the second confirms the (asymptotic) optimality of selection procedures based on Huber's T^0 relative to others.

THEOREM 2. $1 \leqslant N(T^0, \mathcal{F}^0, \Delta)/N(T^0, F^0, \Delta) \to 1$ *as $\Delta\downarrow 0$, and $N(T^0, \mathcal{F}^0, \Delta) \sim d^2\sigma_0^2/\Delta^2$ as $\Delta\downarrow 0$; moreover, a selection procedure based on Huber's estimator T^0 with sample size $N \sim d^2\sigma_0^2/\Delta^2$ is asymptotically robust at $\Phi \in \mathcal{F}^0$.*

Here, as elsewhere, d is defined by $P(d) = P^*$ (see (1)); Bechhofer's (1954) Table I provides values of d (his table entries) for selected values of P^* for $k = 1(1)10$ $(t = 1)$.

PROOF. Writing $d = P^{-1}(P^*)$ and $d_\varepsilon = P^{-1}(P^* - \varepsilon)$, we have from Lemma 3:

$$N_{P^*-\varepsilon}(T^0, F^0, \Delta) \sim d_\varepsilon^2\sigma_0^2/\Delta^2, \qquad N_{P^*}(T^0, F^0, \Delta) \sim d^2\sigma_0^2/\Delta^2.$$

Since $P(\cdot)$ is continuous, we can choose ε small enough for $d^2 - d_\varepsilon^2 < \zeta^2/\sigma_0^2$, and then $\Delta^2(N_{P^*} - N_{P^*-\varepsilon}) < \zeta$ for small Δ. By Lemma 4, $\Delta^2(N_{P^*-\varepsilon}(\mathcal{F}^0) - N_{P^*-\varepsilon}(F^0))$

$< \zeta$ for Δ small. Replace P^* by $P^* + \varepsilon$, and conclude that $0 \leqslant N_{P^*}(\mathfrak{F}^0) - N_{P^*}(F^0) = o(\Delta^{-2})$ as $\Delta \downarrow 0$. But $\Delta^2 N_{P^*}(F^0) \to d^2\sigma_0^2$, and the theorem follows.

THEOREM 3. *If* $T = \{T_n\}$ *is translation equivariant, and if* $n^{\frac{1}{2}}(T_n - \theta)/\sigma \to N(0, 1)$ *for some* σ *when* $F = F^0$, *then*

$$\liminf_{\Delta \downarrow 0} \frac{N(T, \mathfrak{F}^0, \Delta)}{N(T^0, \mathfrak{F}^0, \Delta)} \geqslant \frac{\sigma^2}{\sigma_0^2} \geqslant 1;$$

i.e., *the selection procedure based on* T^0 *with* $N \sim d^2\sigma_0^2/\Delta^2$ *is asymptotically most economical for* $\mathfrak{F}^0$.

PROOF. By Lemma 3, $N(T, F^0, \Delta) \sim d^2\sigma^2/\Delta^2$. Therefore $N(T, \mathfrak{F}^0, \Delta) / N(T^0, \mathfrak{F}^0, \Delta) \geqslant N(T, F^0, \Delta)/N(T^0, \mathfrak{F}^0, \Delta) \to \sigma^2/\sigma_0^2$, using Theorem 2, and this is $\geqslant 1$ by (b) of Huber's minimax theorem (Section 1).

REFERENCES

ALAM, K. and RIZVI, M. H. (1966). Selection from multivariate normal populations. *Ann. Inst. Statist. Math.* **18** 307–318.

BECHHOFER, ROBERT E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25** 16–39.

DALAL, S. R. and HALL, W. J. (1973). Asymptotically most robust selection procedures. (Preliminary report). Abstract: *I.M.S. Bull.* **2** 118.

FELLER, WILLIAM (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York.

GHOSH, MALAY (1973). Nonparametric selection procedures for symmetric location parameter populations. *Ann. Statist.* **1** 773–779.

GIBBONS, JEAN DICKINSON, OLKIN, INGRAM and SOBEL, MILTON (1977). *Selecting and Ordering Populations: A New Statistical Methodology*. Wiley, New York.

HALL WM. JACKSON (1959). The most-economical character of some Bechhofer and Sobel decision rules. *Ann. Math. Statist.* **30** 964–969.

HUBER, PETER J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

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

RANDLES, RONALD H. (1970). Some robust selection procedures. *Ann. Math. Statist.* **41** 1640–1645.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER NY 14627