Gertrude M. Cox

AN OPTIMUM PROPERTY OF BECHHOFER'S SINGLE-SAMPLE MULTIPLE-DECISION PROCEDURE FOR RANKING MEANS AND SOME EXTENSIONS[1]

by

Wm. Jackson Hall
Institute of Statistics
University of North Carolina

Institute of Statistics
Mimeograph Series No. 118
September, 1954

1. This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

UNCLASSIFIED

Security Information

Bibliographical Control Sheet

1. O.A.: Institute of Statistics, North Carolina State College of the University of North Carolina

   M.A.: Office of Scientific Research of the Air Research and Development Command

2. O.A.: CIT Report No. 11

   M.A.: OSR-TN-54-279

3. AN OPTIMUM PROPERTY OF BECHHOFER'S SINGLE-SAMPLE MULTIPLE-DECISION PROCEDURE FOR RANKING MEANS AND SOME EXTENSIONS

4. William Jackson Hall

5. September, 1954

6. 13

7. None

8. AF 18(600)-458

9. RDO No. R-354-20-8

10. UNCLASSIFIED

11. None

12. Bechhofer [1] has considered the problem of ranking means of k normal populations with known variances, or more generally, of grouping the populations according to ranks. He suggests, with only intuitive justification, grouping the population means according to the ranked sample means, and gives tables for finding the minimum sample size which can be used which will guarantee a specified probability of a correct grouping when the population means satisfy lower bounds on the "distances" between groups. This paper gives justification for the use of his procedures when the population variances and the sample sizes are specified as equal among the populations, proving that if the bounds on the "distances" are to be satisfied, the sample size cannot be reduced by using any other type of procedure--that is, that Bechhofer's procedure is a most economical multiple-decision rule, as defined in [3]. Similar results, with some limitation, are obtained for problems of ranking other population parameters and for a distribution-free ranking problem as well; difficulties in computing the most economical sample size for these non-normal problems are discussed and approximations using Bechhofer's tables are indicated.

AN OPTIMUM PROPERTY OF BECHHOFER'S SINGLE-SAMPLE MULTIPLE-DECISION PROCEDURE FOR RANKING MEANS AND SOME EXTENSIONS

Wm. Jackson Hall, University of North Carolina

Summary. Bechhofer [1] has considered the problem of ranking means of k normal populations with known variances, or, more generally, of grouping the populations according to ranks. He suggests, with only intuitive justification, grouping the population means according to the ranked sample means, and gives tables for finding the minimum sample size which can be used which will guarantee a specified probability of a correct grouping when the population means satisfy lower bounds on the "distances" between groups. This paper gives justification for the use of his procedures when the population variances and the sample sizes are specified as equal among the populations, proving that if the bounds on the "distances" are to be satisfied, the sample size cannot be reduced by using any other type of procedure--that is, that Bechhofer's procedure is a most economical multiple-decision rule, as defined in [3]. Similar results, with some limitation, are obtained for problems of ranking other population parameters and for a distribution-free ranking problem as well; difficulties in computing the most economical sample size for these non-normal problems are discussed and approximations using Bechhofer's tables are indicated.

1. Choosing the "Best" Normal Population. We first suppose our goal is:

Goal I: to choose the population with the largest mean after taking a sample of size n, as yet unspecified, from each of k normal populations π_1, π_2, ..., π_k with unknown means $\mu_1, \mu_2, \ldots, \mu_k$ but equal and known variances σ^2. Thus, we have a k-decision problem; we denote by A_i the decision to choose the population π_i as the "best" population (i = 1, 2, ..., k), that is, the decision that $\mu_i = \mu_{[k]}$ where $\mu_{[1]}, \ldots, \mu_{[k]}$ are the ranked means (in increasing order).

We shall freely use the notation introduced by Bechhofer.[2] We require, given $\gamma (0 < \gamma < 1)$ and $\delta (\delta > 0)$, that the probability of a correct decision be at least γ if $\mu_{[k]} - \mu_{[k-1]} \geq \delta$ and subject to this restriction we wish to derive a decision procedure with a minimum sample size.

Let $\underline{x}_j = (x_{1j}, x_{2j}, \ldots, x_{kj})$, $j = 1, \ldots, n$, denote the sample values where x_{ij} is the jth sample value from π_i. We consider $\underline{x}_j$ as one observation from a k-variate population and denote the (k-variate) density function of $\underline{x}_j$ by $f(\underline{x}, \underline{\mu})$ where $\underline{\mu} = (\mu_1, \ldots, \mu_k)$. Denote

$$\omega_i = \{\underline{\mu}: \ \mu_i \geq \mu_j + \delta \ \text{ for all } j \neq i\}$$

$$= \{\underline{\mu}: \ \mu_i = \mu_{[k]} \text{ and } \mu_{[k]} - \mu_{[k-1]} \geq \delta\} \quad (i = 1, \ldots, k).$$

Clearly, using the terminology of Chapter II of [3], our problem is to find a M.E. k-d.r. (most economical k-decision rule) relative to the vector $\underline{\gamma}=(\gamma, \ldots, \gamma)$ for discriminating among $\underline{\omega} = (\omega_1, \ldots, \omega_k)$. We shall use the first minimax method of [3] to obtain such a d.r.

Let n be fixed. Consider the conditional a priori distributions $\lambda_1, \ldots, \lambda_k$ over $\omega_1, \ldots, \omega_k$, respectively, where λ_i assigns probability 1 to the parameter point $\underline{\mu}$ with the ith coordinate $\mu_0 + \delta$ and all other coordinates μ_0, μ_0 arbitrary but fixed, and consider the simple discrimination problem of finding a minimax d.r. D^0 w.r.t. the weight function $W(f_i, A_j) = (-1/\gamma$ if $i = j$ and 0 otherwise) for

---

2. All references to Bechhofer refer to [1] unless otherwise indicated.

discriminating among $\underline{f} = (f_1, \ldots, f_k)$ where $f_i = \int_{\omega_i} f(\underline{x}, \underline{\mu})\, d\lambda_i(\underline{\mu})$ $(i = 1, \ldots, k)$; that is, f_i is a k-variate normal density function with mean vector $\underline{\mu} = (\mu_1, \ldots, \mu_k)$ where $\mu_i = \mu_0 + \delta$ and $\mu_j = \mu_0$, $j \neq i$, and covariance matrix $P = (\sigma_{ij}) = (\sigma^2 \delta_{ij})$. Using the results of Chapter I in [3], a non-randomized likelihood ratio d.r. D^0 satisfying (1.11) in [3], i.e., satisfying

$$p_1(D^0) = \ldots = p_k(D^0) \tag{1}$$

where $p_i(D) = \Pr(D \text{ chooses } A_i \mid f_i)$, is a minimax d.r. Now a likelihood ratio d.r. is determined by the ratios $a_i f_i^n / a_j f_j^n$ for some positive constants $a_1, \ldots, a_k$ where f_i^n is the corresponding likelihood for samples of size n. It is easily verified that $f_i^n / f_j^n = \exp [\delta n(\bar{x}_i - \bar{x}_j)/\sigma^2]$ where $\bar{x}_i$ denotes the sample mean of π_i. Hence, a minimax d.r. D^0 chooses A_i if $a_i f_i^n > a_j f_j^n$ for all $j \neq i$, or, equivalently, if $y_{ij} > c_{ij}$ for all $j \neq i$ where $y_{ij} = \bar{x}_i - \bar{x}_j$ and $c_{ij} = (\sigma^2/\delta n) \log(a_j/a_i)$, and the c_{ij}'s are determined so that (1) holds. (We ignore the possibility of equality of $a_i f_i^n$ and $a_j f_j^n$ throughout since this is an event of probability 0.) Now, if f_i is true, $\underline{y}_i = (y_{i1}, \ldots, y_{i,i-1}, y_{i,i+1}, \ldots, y_{ik})$ has a (k-1)-variate normal distribution with means $(\delta, \ldots, \delta)$ and covariance matrix $P_1 = (\tau_{ij})$ where $\tau_{ij} = (2\sigma^2/n$ if $i = j$ and σ^2/n if $i \neq j)$; denote such a distribution by G (which is independent of μ_0). Then $p_i(D^0) =$

$$\int_{c_{i1}}^{\infty} \cdots \int_{c_{i,i-1}}^{\infty} \cdots \int_{c_{i,i+1}}^{\infty} \cdots \int_{c_{ik}}^{\infty} dG = H(c_{i1}, \ldots, c_{i,i-1}, c_{i,i+1}, \ldots, c_{ik}), \text{ say.}$$

Rearrange the subscripts so that a_1 and a_k are the smallest and largest, respectively, of the a_i's; then $c_{ij} \geq 0 \geq c_{kj}$ for all j with equalities if and only if all a_i's are equal. Now H is a decreasing function in each of its arguments so that $p_k(D^0) \geq H(0, \ldots, 0) \geq p_1(D^0)$ with both equalities if and only if all c_{ij}'s are zero. Hence, the requirement (1) implies that all c_{ij}'s are zero and hence that all a_i's are equal. Therefore, D^0 chooses A_i if $y_{ij} > 0$ for all $j \neq i$, that is, if $\bar{x}_i$ is the largest sample mean (i = 1, ..., k). Note that D^0 is independent of μ_0 which remains arbitrary.

Bechhofer proves that $\lambda_1, \ldots, \lambda_k$ are least favorable in the sense of Theorem 2.5 (ii) of [3] so it follows from that theorem that D^0 is a minimax d.r. w.r.t. the simple weight function

$$W(\underline{\mu}, A_j) = \begin{cases} -1/\gamma & \text{if } \underline{\mu} \in \omega_j \\ 0 & \text{otherwise} \end{cases} \qquad (j = 1, \ldots, k)$$

for discriminating among $\underline{\omega}$.

By Theorem 2.7 of [3], to find a M.E. d.r. we need consider only such minimax d.r.'s D_n^0 for various values of the sample size n. Bechhofer has given tables for finding the minimum n, say N, for which D_n^0 meets our requirements. Hence D_N^0 is a M.E. d.r. for discriminating among $\underline{\omega}$; thus Bechhofer's procedure is most economical for Goal I.

2. Grouping Normal Means. We now consider the more general goal treated by Bechhofer of which Goal I is a special case:

Goal II: to find the k_s "best" populations, the k_{s-1} "second best" populations, the k_{s-2} "third best" populations, etc., and finally the k_1 "worst" populations, given $k_1, k_2, \ldots, k_s$ ($s \leq k$, $\sum_{i=1}^{s} k_i = k$), "best" being interpreted in the sense of largest population means. Denote $\hat{k}_i = \sum_{j=1}^{i} k_j$ and $\delta_{\hat{k}_i+1,\hat{k}_i} = \mu_{[\hat{k}_i+1]} - \mu_{[\hat{k}_i]} = \delta_i$. We assume as in Section 1 that all variances are equal and that the same number of samples are to be taken from each population.

Let $A_{i_1 i_2 \ldots i_k}$ be the decision:

$\pi_{i_1}, \ldots, \pi_{i_{\hat{k}_1}}$ are the "worst" populations,

$\pi_{i_{\hat{k}_1+1}}, \ldots, \pi_{i_{\hat{k}_2}}$ are the "next worse",

....

$\pi_{i_{\hat{k}_{s-1}+1}}, \ldots, \pi_{i_{k_s}}$ are the "best" populations

where $(i_1, \ldots, i_k)$ is a permutation of $(1, \ldots, k)$ such that $i_1 < i_2 < \ldots < i_{\hat{k}_1}$, $i_{\hat{k}_1+1} < i_{\hat{k}_1+2} < \ldots < i_{\hat{k}_2}, \ldots, i_{\hat{k}_{s-1}+1} < i_{k_{s-1}+2} < \ldots < i_{\hat{k}_s} \equiv i_k$. Thus, there are $m = k!/(k_1!k_2! \ldots k_s!)$ alternative decisions. Correspondingly, given positive constants $\delta_1^*, \delta_2^*, \ldots, \delta_{s-1}^*$, denote

$$\begin{aligned}(2)\quad \omega_{i_1 i_2 \ldots i_k} = (\underline{\mu}:\ & \mu_j \le \mu_{[\hat{k}_1]} \text{ for } i=i_1, i_2, \ldots, i_{\hat{k}_1}, \\ & \mu_{[\hat{k}_1]} < \mu_j \le \mu_{[\hat{k}_2]} \text{ for } j = i_{\hat{k}_1+1}, i_{\hat{k}_1+2}, \ldots, i_{\hat{k}_2}, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & \mu_{[\hat{k}_{s-1}]} < \mu_j \le \mu_{[\hat{k}_s]} \text{ for } j = i_{\hat{k}_{s-1}+1}, i_{\hat{k}_{s-1}+2}, \ldots, i_k, \\ & \text{and } \delta_i \ge \delta_i^* \text{ for } i = 1, \ldots, s-1),\end{aligned}$$

Given γ $(0 < \gamma < 1)$, we wish to find a M.E. m-d.r. relative to the vector $\underline{\gamma} = (\gamma, \ldots, \gamma)$ for discriminating among $\underline{\omega}$, defined by (2), that is, a d.r. D based on a minimum sample size subject to $Pr(D \text{ chooses } A_{i_1 i_2 \ldots i_k} \mid \underline{\mu}) \geq \gamma$ if $\underline{\mu} \in \omega_{i_1 i_2 \ldots i_k}$ for all m values of the subscripts.

To obtain such a d.r. we use a method analogous to that in Section 1; the argument is brief because of this analogy and the notational complexities.

Let n be fixed. Let $\lambda_{i_1 \ldots i_k}$ be a conditional a priori distribution over $\omega_{i_1 \ldots i_k}$ assigning probability 1 to the parameter point $\underline{\mu}$ with coordinates

$$\mu_{i_1} = \mu_{i_2} = \ldots = \mu_{i_{\hat{k}_1}} = \mu_0$$

$$\mu_{i_{\hat{k}_1+1}} = \mu_{i_{\hat{k}_1+2}} = \ldots = \mu_{i_{\hat{k}_2}} = \mu_0 + \delta_1^*$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mu_{i_{\hat{k}_{s-1}+1}} = \mu_{i_{\hat{k}_{s-1}+2}} = \ldots = \mu_{i_{\hat{k}_s}} = \mu_0 + \delta_1^* + \delta_2^* + \ldots + \delta_{s-1}^*,$$

μ_0 arbitrary, and denote

$$f_{i_1 \ldots i_k} = \int_{\omega_{i_1 \ldots i_k}} f(\underline{x}, \underline{\mu}) \, d\lambda_{i_1 \ldots i_k},$$

and correspondingly for the other values of the subscripts. Consider the simple discrimination problem of discriminating among the $f_{i_1 \ldots i_k}$'s. A likelihood ratio d.r. where the $a_{i_1 \ldots i_k}$'s are chosen so that the $p_{i_1 \ldots i_k}(D)$'s are all equal is a minimax d.r. for this problem. Now it is easily verified that for any set of $a_{i_1 \ldots i_k}$'s,

$$a_{i_1\ldots i_k} f_{i_1\ldots i_k} > a_{j_1\ldots j_k} f_{j_1\ldots j_k} \Longleftrightarrow$$

$$y_{i_1\ldots i_k, j_1\ldots j_k} \equiv \sum_{t=1}^{s-1} \delta_t^* z_t(i_1\ldots i_k, j_1\ldots j_k) > \frac{\sigma^2}{n} \log \frac{a_{j_1\ldots j_k}}{a_{i_1\ldots i_k}}$$

$$\equiv c(i_1\ldots i_k, j_1\ldots j_k)$$

where the z_t's are various contrasts of the sample means with all coefficients restricted to the values -1, 0,+1, the coefficients depending on $(i_1\ldots i_k, j_1\ldots j_k)$. (It may be helpful to consider a special case such as $k=4$, $s=3$, $k_3=1$, $k_2=1$, $k_1=2$.) For any set of subscripts $(i_1\ldots i_k)$, the m-1 $y_{i_1\ldots i_k, j_1\ldots j_k}$'s corresponding to various sets of subscripts $(j_1\ldots j_k)$ unequal to $(i_1\ldots i_k)$ are random variables having a joint (m-1)-variate normal distribution (independent of μ_0). Because of the symmetry of the problem (we assume no a priori knowledge about the relative magnitude of the population means), the same normal distribution, say G', will occur for any set of $(i_1\ldots i_k)$'s, as long as the order of the variables is properly arranged. Hence, we may write $p_{i_1\ldots i_k}(D^0) = \int\int \ldots \int dG'$ where the upper limits of the (m-1)-fold integral are all $+\infty$ and the lower limits are the m-1 $c(i_1\ldots i_k, j_1\ldots j_k)$'s for various values of the $(j_1\ldots j_k)$'s unequal to $(i_1\ldots i_k)$. Now the c's are to be determined so that all $p_{i_1\ldots i_k}(D^0)$'s are equal. Let $a_{i_1'\ldots i_k'}$ and $a_{i_1''\ldots i_k''}$ be the smallest and largest of the a's, respectively; then an argument analogous to that in Section 1 proves that all c's must be zero and hence all a's must be equal.

Consider two sets of subscripts $i_1\ldots i_k$ and $j_1\ldots j_k$, denoted simple $\underline{i}$ and $\underline{j}$, which differ only in that, for some ℓ $(1 \leq \ell \leq s-1)$, one subscript, say ℓ', in the ℓth group and one subscript, say ℓ'', in the $(\ell+1)$th group of $\underline{i}$ are interchanged in $\underline{j}$ (and then the subscripts rearranged within groups so that they are in increasing order). Then, clearly,

$$\frac{a_{\underline{i}} f^n_{\underline{i}}}{a_{\underline{j}} f^n_{\underline{j}}} = \frac{\exp\{-\frac{1}{2\sigma^2}\Sigma[(x_{\ell'}-\mu_0-\delta^*_1-\cdots-\delta^*_{\ell-1})^2+(x_{\ell''}-\mu_0-\delta^*_1-\cdots-\delta^*_{\ell})^2]\}}{\exp\{-\frac{1}{2\sigma^2}\Sigma[(x_{\ell''}-\mu_0-\delta^*_1-\cdots-\delta^*_{\ell-1})^2+(x_{\ell'}-\mu_0-\delta^*_1-\cdots-\delta^*_{\ell})^2]\}}$$

$$= \exp\left[\frac{n}{\sigma^2}\delta^*_{\ell}(\bar{x}_{\ell''} - \bar{x}_{\ell'})\right]$$

where the summations are over the sample, the sample subscripts having been omitted. Hence, the corresponding contrasts z_t $(t = 1, \ldots, s-1)$ have all coefficients zero except the ℓth contrast which is $\bar{x}_{\ell''} - \bar{x}_{\ell'}$. Thus $(a_{\underline{i}} f^n_{\underline{i}} > a_{\underline{j}} f^n_{\underline{j}})$ is equivalent to $(\bar{x}_{\ell''} > \bar{x}_{\ell'})$. Such a statement holds for every two sets of subscripts corresponding to two consecutive groups. Hence, the decision to choose $A_{i_1 \ldots i_k}$ by D^0 implies $\bar{x}_{j_1} > \bar{x}_{j_2}$ for every pair of subscripts (j_1, j_2), j_1 belonging to the $(t+1)$th group $(i_{\hat{k}_t+1}, \ldots, i_{\hat{k}_{t+1}})$ and j_2 belonging to the tth group $(i_{\hat{k}_{t-1}+1}, \ldots, i_{\hat{k}_t})$, $t = 1, \ldots, s-1$. Clearly, then, a minimax d.r. D^0 (for a fixed sample size) groups the populations according to the sample means; i.e., denoting the ranked sample means (in increasing order) $\bar{x}_{[1]}, \bar{x}_{[2]}, \ldots, \bar{x}_{[k]}$, and the corresponding populations by $\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(k)}$, D^0 chooses the alternative A corresponding to:

$\pi_{(1)}, \ldots, \pi_{(\hat{k}_1)}$ are the worst populations

$\pi_{(\hat{k}_1+1)}, \ldots, \pi_{(\hat{k}_2)}$ are the next worse

....

$\pi_{(\hat{k}_{s-1}+1)}, \ldots, \pi_{(k)}$ are the best populations.

Note that D^0 is independent of μ_0 which remains arbitrary.

Bechhofer proves that the least favorable configuration, which is assigned probability 1 by the $\lambda_{i_1 \ldots i_k}$'s, is least favorable in the sense of Theorem 2.5 (ii) of [3]. Hence, D^0 is a minimax d.r. for discriminating among $\underline{\omega}$, defined by (2), and D_N^0, where N is chosen according to Bechhofer's rules, is a M. E. d.r. for this problem. Thus, we have proved Bechhofer's procedure to be most economical for Goal II.

3. Ranking the Parameters of Distributions of the Laplacian Class. We shall show that similar results, with some limitation, hold for problems of ranking (or grouping according to ranks) the parameters $\theta_1, \ldots, \theta_k$ of k populations whose distributions, differing only in a parameter θ, belong to the Laplacian class of distributions, defined as all distributions having a density function or probability function of the form

$$g(x,\theta) = A(x) \exp [\rho(\theta)\, t(x) + \tau(\theta)] \tag{3}$$

where ρ is an increasing function of θ. This class includes the normal distributions with known variance, the normal distribution with known mean, and the binomial, negative binomial, and Poisson distributions. We require, as before, that the same number of samples be taken from each population.

We consider only extensions to Section 1; extensions to Section 2 are completely analogous. Denote by $f(\underline{x},\underline{\theta})$ the k-variate density (or probability) function of k independent variables, the ith variable having the density (or probability) function $g(x_i,\theta_i)$ $(i = 1, \ldots, k)$. Except for one point, it is easily verified that the derivation of Section 1 is valid for this case concerning any distribution of the Laplacian class, replacing population means by θ's, the sample means $(\bar{x}_1, \ldots, \bar{x}_k)$ by $(\bar{t}_1,\ldots,\bar{t}_k)$ where $\bar{t}_i$ is the average over the sample values of t(x) for the ith population, and the (k-1)-variate normal distribution, G, of the $\underline{y}_i$ vectors by a (k-1)-variate joint distribution of $\bar{t}_i-\bar{t}_j$ $(j \neq i,$ $j = 1, \ldots, k)$ where the $\bar{t}_j$'s are no longer necessarily normal[3]. But none of the arguments of Section 1 depend on the nature of the G distribution except the fact that G is independent of μ_0 (here denoted θ_0); this fact, the exceptional point referred to above, will not necessarily hold for other distributions of the Laplacian class--in fact, it does not hold for any of the other distributions listed above. This is intuitively not surprising since none of the parameters considered other than the normal mean have an unlimited range from plus to minus infinity and hence we might expect a minimax d.r. to depend on the "location" fixed by θ_0.

Let D_n^0 denote a decision rule based on a sample of size n which chooses as the best population the one with the largest sample $\bar{t}_i$, and suppose N is the sample size taken from (hypothetical) tables similar to Bechhofer's but constructed from the appropriate G distribution for a specified θ_0. Then, if θ_0 is specified in the definition of the parameter spaces ω_i, thusly

---

3. It should also be noted that events $a_i f_i^n = a_j f_j^n$ may now have positive probability and must be accounted for by allowing randomized d.r.'s--that is, "ties" must be "broken" by randomization.

(4) $$\omega_i^0 = \{\underline{\theta}: \ \theta_i = \theta_{[k]} - \theta_{[k-1]} \geq \delta, \text{ and } \theta_{[j]} = \theta_0 \text{ for some specified value of } j\} \quad (i = 1, \ldots, k),$$

we have as in Section 1 that D_N^0 is a M.E. d.r. relative to $\underline{\gamma} = (\gamma, \ldots, \gamma)$ for discriminating among $\underline{\omega}^0$. To find a M.E. d.r. for discriminating among $\underline{\omega}$ (with θ_0 unspecified), it would be necessary to find a "least favorable" value of θ_0 for the particular problem at hand. This may make the "least favorable configuration" a rather unrealistic one; this might be overcome, however, by taking a "middle course" of limiting θ_0 to some interval.

Tables for finding the M.E. sample size for such non-normal problems analogous to Bechhofer's tables for the normal case have not been constructed. The compilation of such tables would involve some unsolved distribution problems; for example, if the θ's are parameters of binomial distributions, we require the joint distribution of m-1 dependent y_i's, each of which is the difference between two binomial variables with the same parameter n but different θ's. However, by transforming $t = t(x)$ into an approximately normally distributed variable, Bechhofer's tables might be used to advantage to approximate the M.E. sample size, as suggested by Bechhofer in Section 7 of [1]; for example, see Bechhofer's Example 4. Some indication of the accuracy of such approximations may be given by the comparison of a similar approximation with the exact computation made by Bechhofer and Sobel in [2].

4. <u>A Distribution-Free Ranking Problem.</u> Suppose $F_1, \ldots, F_k$ are unknown continuous distribution functions corresponding to the populations $\pi_1, \ldots, \pi_k$, and denote $\theta(F_i) = \theta_i = \int_0^\infty dF_i \ (i = 1, \ldots, k)$. (Alternatively, some constant

other than zero may be specified as the lower limit of integration in the definition of Θ_i.) Consider the goal analogous to Goal I: to choose the population with the largest value of Θ, after taking an equal number of samples from each population. If the F_i's differ only in location parameter, then the ranked Θ_i's correspond to ranked means or medians. A similar development may be given for the analogous Goal II.

We shall show that, for a fixed sample size, a d.r. D^0 which chooses Θ_i as the largest Θ if the sample from π_i has the largest number of positive values ($i = 1, \ldots, k$) is a minimax d.r. for a simple weight function for discriminating among $\underline{\omega}^0$, defined by

$$\omega_i^0 = \{F_1, \ldots, F_k: \ \Theta_i = \Theta_{[k]}, \ \Theta_{[k-1]} \geq \delta, \text{ and } \Theta_{[j]} = \Theta_0 \text{ for some specified value of } j\} \ (i=1,\ldots,k);$$

and hence, by choosing the sample size as indicated in Section 3 for the proper distribution G, D^0 is a M.E. d.r.

Let

$$h(x,\Theta) = \begin{cases} \Theta^{t(x)}(1 - \Theta)^{1-t(x)}/C & \text{if } |x| \leq C \\ 0 & \text{otherwise} \end{cases}$$

where $t(x) = 1$ if $x \geq 0$ and 0 otherwise and C is an arbitrary positive constant, and denote by $f(\underline{x},\underline{\Theta})$ the corresponding joint distribution of $(x_1, \ldots, x_k)$ where x_i has the density $h(x_i,\Theta_i)$ ($i = 1, \ldots, k$). Let λ_i assign probability 1 to $\{F_1, \ldots, F_k: \ dF_\ell/dx = h(x,\Theta_\ell) \ (\ell = 1, \ldots, k) \text{ and } \Theta_i = \Theta_0+\delta, \ \Theta_j = \Theta_0 \text{ for all } j \neq i\}$ ($i = 1, \ldots, k$), and consider the simple discrimination problem of discriminating among $f_1, \ldots, f_k$ where f_i is the "average" (w.r.t. λ_i) density function of $(x_1, \ldots, x_k)$. Since $h(x,\theta_i)$ is of the form (3), Section 3 implies that D^0 defined above is a minimax d.r. for this simple discrimination problem. That the λ_i's are least favorable follows as for the previous cases, noting first that the probability that D^0 will choose θ_i when $F_1, \ldots, F_k$ are in ω_i^0 depends on $F_1, \ldots, F_k$ only through $\theta_1, \ldots, \theta_k$.

The distribution G required to compute the M.E. sample size is the same distribution required for the case of choosing the largest of k binomial parameters. (See the last paragraph of Section 3.)

5. Directions for Further Research. The author is at present considering the following problems: (1) to prove Bechhofer and Sobel's procedure for ranking variances of normal populations with unknown means [2] is M.E.; (2) to find M.E. solutions when different values of $\gamma = \gamma_i$ are specified corresponding to each θ_i; (3) to apply the theory of Section 4.3 of [3] to find most economical solutions when the restriction of equal sample sizes (and of equal variances for the normal case) is dropped; (4) to find a least favorable distribution of θ_0 for various distributions of the Laplacian class; and (5) to find methods for obtaining the M.E. sample size more accurately for the non-normal cases--in particular, for the binomial and distribution-free cases.

REFERENCES

[1] Bechhofer, Robert E., "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances", Annals of Mathematical Statistics, 25 (1954), 16-39.

[2] Bechhofer, Robert E., and Sobel, Milton, "A Single-Sample Multiple Decision Procedure for Ranking Variances of Normal Populations", Annals of Mathematical Statistics, 25 (1954), 273-289.

[3] Hall, Wm. Jackson, "Most Economical Multiple-Decision Rules", Institute of Statistics Mimeograph Series No. 115, University of North Carolina (1954).