A SEQUENTIAL PROCEDURE FOR SELECTING THE POPULATION WITH THE LARGEST MEAN FROM k NORMAL POPULATIONS[1]

By Edward Paulson

Queens College

1. Summary. In this paper sequential procedures are given for selecting the normal population with the greatest mean when (a) the k populations have a common known variance or (b) the k populations have a common but unknown variance, so that in each case the probability of making the correct selection exceeds a specified value when the greatest mean exceeds all other means by at least a specified amount.

The procedures in the present paper all have the property that inferior populations can be eliminated from further consideration as the experiment proceeds.

2. Introduction. A problem which seems to be of considerable practical interest may be described as follows. An experimenter is concerned with comparing k categories, such as k new drugs, k new machines, k new alloys, k new teaching methods, etc. We suppose that all k categories are of an experimental nature (that is, no standard or control is being used). The experimenter is practically certain that the k categories differ among themselves, and his objective is to select the "best" category on the basis of the information supplied by taking measurements with each category. It is assumed that all measurements are normally and independently distributed and each category is characterized by its population mean. The best category is defined (for convenience) as the category with the largest population mean.

What the experimenter requires then is a statistical procedure which will tell him what measurements to take and what population to select so that the probability of making a correct decision is controlled at some preassigned level when the largest mean exceeds all the other means by at least a specified amount.

This problem was formulated and solved by Bechhofer [2] for the case in which the k variances are known, and the experimenter is restricted to a fixed-sample size procedure. The result in [2] was extended by Bechhofer, Dunnett, and Sobel [3] for the case of a common but unknown variance when the experimenter is restricted to a two stage procedure. The case of a common but unknown variance is treated sequentially in [4], [5] where numerical illustrations (but no mathematical details) are given. Bechhofer, Kiefer, and Sobel are making a detailed study of sequential multiple decision procedures and plan to publish a monograph on this topic.

In the present paper the writer has modified some of his recent work [6] to arrive at sequential decision procedures for the present problem which differ

Received 1 November 1962; revised 28 July 1963.

[1] This research was supported by the National Science Foundation under Grant NSF-G23665.

substantially both from the Bechhofer-Kiefer-Sobel approach and from another approach suggested by Stein [8].

3. Notation and formulation of the problem. The k categories or populations are denoted by $\pi_1, \pi_2, \cdots, \pi_k$. Let X_{is} denote the sth measurement with category π_i ($i = 1, 2, \cdots k$, and $s = 1, 2, \cdots$). We assume throughout the paper that X_{is} is normally distributed with mean m_{is} and variance σ^2, and that the measurements $\{X_{is}\}$ are independent random variables for all values of i and s. We restrict our attention to the following two situations:

A (one way classification) where $m_{is} = \mu_i$

B (randomized block design) where $m_{is} = \mu_i + c_s$.

We assume that $\mu_1, \mu_2, \cdots, \mu_k$ are unknown parameters and the best category is the one with the largest μ. Denote the ranked μ's by

$$\mu_{[1]} \leqq \mu_{[2]} \leqq \mu_{[3]}, \cdots, \leqq \mu_{[k]}.$$

The problem of this paper is to design a sequential procedure for selecting the best category so that when $\mu_{[k]} - \mu_{[k-1]} \geqq \Delta$ the probability of making the correct decision and selecting $\pi_{[k]}$ is $\geqq 1 - \alpha$. Here Δ and α are constants which are specified by the experimenter in advance of the experiment on the basis of practical considerations, and $\pi_{[j]}$ is the category with parameter $\mu_{[j]}$.

4. The case of a common known variance. We will start by specifying a class S_λ of sequential procedures depending on a parameter λ where $0 < \lambda < \Delta$, and then show that for each value of λ in this range the corresponding sequential procedure has the required property that $P[\pi_{[k]} \text{ is selected} \mid \mu_{[k]} - \mu_{[k-1]} \geqq \Delta] \geqq 1 - \alpha$. Let $a_\lambda = [\sigma^2/(\Delta - \lambda)]\log((k-1)/\alpha)$, where all logarithms are to the base e, and let $W_\lambda =$ the largest integer less than a_λ/λ. At the first stage of the experiment we take one measurement with each category, obtaining measurements $(X_{11}, X_{21}, \cdots, X_{k1})$. Then we eliminate from further consideration any category π_j for which

$$X_{j1} < \max\{X_{11}, X_{21}, \cdots, X_{k1}\} - a_\lambda + \lambda.$$

If all but one category are eliminated after the first stage of the experiment, we stop the experiment and select the remaining category as the best one. Otherwise we go on to the second stage of the experiment and take one measurement on each category not eliminated after the first stage. Proceeding by induction, at the rth stage of the experiment ($r = 2, 3, \cdots, W_\lambda$) we take one measurement on each category not eliminated after the $(r-1)$ stage, and then eliminate any remaining category π_j for which

$$\sum_{s=1}^{r} X_{js} < \max_\nu \left\{ \sum_{s=1}^{r} X_{\nu s} \right\} - a_\lambda + r\lambda,$$

where the max is taken over all categories left after the $(r-1)$ stage. If only one category is left after the rth stage, the experiment is terminated and the remaining category is selected, otherwise we go on to the $(r+1)$ stage. If more

than one category remains after the W_λ stage, the experiment is terminated at the $(W_\lambda + 1)$ stage by selecting the remaining category for which the sum of the $(W_\lambda + 1)$ measurements is a maximum.

Now we proceed to show that for each λ the corresponding sequential procedure has the required property. First let δ_2 denote the parameter configuration $\mu_{[k]} \geqq \mu_{[k-1]} + \Delta$, and let δ_2^* denote the parameter configuration $\mu_k \geqq \mu_j + \Delta$ for $j = 1, 2, \cdots, k - 1$. It is obvious from the symmetry of the sequential procedure that it is sufficient to prove that the probability of a correct decision is $\geqq 1 - \alpha$ when $[k] = k$ and the parameter point belongs to δ_2^*. When π_k is eliminated before the $(W_\lambda + 1)$ stage it follows from the specification of the sequential procedure that for at least one value of ν ($\nu = 1, 2, \cdots, k - 1$) there must exist an integer $n(\nu) \leqq W_\lambda$ so that

$$\sum_{s=1}^{n(\nu)} X_{ks} < \sum_{s=1}^{n(\nu)} X_{\nu s} - a_\lambda + n(\nu)\lambda.$$

Where π_k is eliminated at the $(W_\lambda + 1)$ stage, it is clear that for some value of ν ($\nu = 1, 2, \cdots, k - 1$)

$$\sum_{s=1}^{W_\lambda+1} X_{ks} < \sum_{s=1}^{W_\lambda+1} X_{\nu s} \leqq \sum_{s=1}^{W_\lambda+1} X_{\nu s} - a_\lambda + (W_\lambda + 1)\lambda$$

since $(W_\lambda + 1)\lambda - a_\lambda \geqq 0$ from the definition of W_λ. Now

$$\begin{aligned}
&P[\text{incorrect decision} \mid \delta_2^*] = P[\pi_k \text{ is eliminated} \mid \delta_2^*] \\
&\leqq P\left[\begin{array}{c}\text{for at least one value of } \nu (\nu = 1, 2 \cdots \cdots, k-1) \\ \text{there exists an integer } n'(\nu) \leqq W_\lambda + 1 \text{ so that} \\ \sum_{s=1}^{n'(\nu)} X_{ks} < \sum_{s=1}^{n'(\nu)} X_{\nu s} - a_\lambda + n'(\nu)\lambda\end{array} \,\middle|\, \delta_2^*\right] \\
&\leqq \sum_{\nu=1}^{k-1} P\left[\sum_{s=1}^{n} X_{ks} < \sum_{s=1}^{n} X_{\nu s} - a_\lambda + n\lambda \quad \text{for some} \quad n \leqq W_\lambda + 1 \,\middle|\, \delta_2^*\right] \\
&\leqq \sum_{\nu=1}^{k-1} P\left[\sum_{s=1}^{n} X_{ks} < \sum_{s=1}^{n} X_{\nu s} - a_\lambda + n\lambda \quad \text{for some} \quad n < \infty \,\middle|\, \delta_2^*\right].
\end{aligned}$$

Now we make use of the known result (see for Example [1] Section 2.1) that if $\{Y_i\}$ is a sequence of independent and identically distributed random variables with a negative expectation and $b > 0$,

$$P\left[\sum_{i=1}^{n} Y_i > b \quad \text{for some } n < \infty\right] \leqq e^{-t_0 b}$$

where t_0 is the non-zero root of $E(e^{tY}) = 1$. We find that

$$\begin{aligned}
&P\left[\sum_{s=1}^{n} X_{ks} < \sum_{s=1}^{n} X_{\nu s} - a_\lambda + n\lambda \quad \text{for some} \quad n < \infty \,\middle|\, \delta_2^*\right] \\
&\qquad = P\left[\sum_{s=1}^{n} (X_{\nu s} - X_{ks} + \lambda) > a_\lambda \quad \text{for some} \quad n < \infty \,\middle|\, \delta_2^*\right] \\
&\qquad \leqq \exp\{[(\mu_\nu - \mu_k + \lambda)/\sigma^2]a_\lambda\} \leqq \exp\{[(-\Delta + \lambda)/\sigma^2]a_\lambda\}
\end{aligned}$$

since $\mu_\nu - \mu_k \leqq -\Delta$ for all points in δ_2^*. Therefore $P[\pi_k \text{ is eliminated} \mid \delta_2^*] \leqq (k-1)\exp\{[(-\Delta + \lambda)/\sigma^2]a_\lambda\}$. Since by definition $a_\lambda = [\sigma^2/(\Delta - \lambda)]\log((k-1)/\alpha)$, we find that $P[\pi_k \text{ is eliminated} \mid \delta_2^*] \leqq \alpha$, so that

$$P[\pi_k \text{ is selected} \mid \delta_2^*] \geqq 1 - \alpha$$

and by symmetry $P[\pi_{[k]} \text{ is selected} \mid \delta_2] \geqq 1 - \alpha$.

At present the "optimum" choice of λ is unknown. Some preliminary calculations indicate that $\lambda = \Delta/4$ should be a reasonably good choice. At present we recommend using $\lambda = \Delta/4$, and conjecture that any loss in efficiency using $\lambda = \Delta/4$ rather than the "optimum" value of λ will be small from a practical point of view. The resulting sequential test when $\lambda = \Delta/4$ can be summarized as follows: at the rth stage eliminate any remaining category π_j for which

$$\sum_{s=1}^{r} X_{js} < \max\left\{\sum_{s=1}^{r} X_{\nu s}\right\} - \frac{4\sigma^2}{3\Delta}\log((k-1)/\alpha) + \frac{r\Delta}{4},$$

where the max is taken with respect to all the categories not eliminated after the $(r-1)$ stage.

In order to get some idea of the efficiency of the resulting sequential procedure when $\lambda = \Delta/4$, a number of sampling experiments were carried out. (The Computations were programmed by Kurt Fuchel and done on the Merlin computer of the Brookhaven National Laboratory.) The results are summarized in Table I, where $\bar{T}_0$ denotes the mean of the total number of measurements from all populations required to make a decision when $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k]}$ (the δ_0 configuration), $\bar{T}_1$ denotes the mean when $\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \Delta$ (the δ_1 configuration), and N denotes the number of measurements with each category required by the fixed-sample size procedure. The values of N were obtained from the tables in [2].

In Table II, the empirically determined average sample size for the procedure of this section with $\lambda = \Delta/4$ (abbreviated the P procedure) was expressed as a percentage of the corresponding single-sample size. The table indicates that the

TABLE I

*A summary of the results of experimental sampling**

k	α	Number of Trials for $\bar{T}_0$	$\bar{T}_0$	Number of Trials for $\bar{T}_1$	$\bar{T}_1$	kN
4	.05	50	755 (27)	100	443 (16)	850
4	.01	50	1178 (47)	100	644 (24)	1442
10	.01	50	3226 (81)	50	1982 (86)	4506

* The values $\sigma = 1$ and $\Delta = .2$ were used throughout. The estimated standard deviation of each mean is written in parenthesis below the mean.

TABLE II

*Ratio (in percent) of the experimentally determined average total sample size divided by the corresponding total fixed-sample size**

k	α	Ratio for the P' Procedure Under δ_0	Ratio for the P Procedure Under δ_1
4	.05	89	52
		(3.2)	(1.9)
4	.01	82	45
		(3.3)	(1.7)
10	.01	72	44
		(1.8)	(1.9)

* The values $\sigma = 1$ and $\Delta = .2$ were used throughout. The estimated standard deviation of each ratio is written in parenthesis below the ratio.

sequential procedure with $\lambda = \Delta/4$ is substantially more efficient than the fixed-sample size procedure.

5. The case of a common but unknown variance. In this section we shall somewhat tentatively propose a sequential procedure which we feel is probably fairly adequate for most practical situations, although it is obviously far from being a perfect solution.

We start by taking a sample of n_0 measurements with each category. Let

$$\bar{X}_{i\cdot} = \sum_{s=1}^{n_0} X_{is}/n_0\,,\quad \bar{X}_{\cdot s} = \sum_{i=1}^{k} X_{is}/k,\quad \bar{X} = \sum_{i=1}^{k}\sum_{s=1}^{n_0} X_{is}/kn_0\,.$$

Under situation A, let $s^2 = \sum_{i=1}^{k}\sum_{s=1}^{n_0}(X_{is} - X_{i\cdot})^2/k(n_0 - 1)$ be the usual estimate of σ^2 with $f = k(n_0 - 1)$ degrees of freedom, while under situation B let $s^2 = \sum_{i=1}^{k}\sum_{s=1}^{n_0}(X_{is} - \bar{X}_{i\cdot} - \bar{X}_{\cdot s} + \bar{X})^2/(k-1)(n_0 - 1)$ be the usual estimate of σ^2 with $f = (k-1)(n_0 - 1)$ degrees of freedom, so that in each situation fs^2/σ^2 has the χ^2 distribution with f degrees of freedom. Next let

$$g = [(k-1)/\alpha)^{2/f} - 1]f/2,$$

let $a^* = 4s^2g/3\Delta$, and let W^* denote the largest integer less than $4a^*/\Delta$.

Having obtained n_0 measurements with each category, we compute s^2 and the values of g, a^*, and W^*. If $n_0 > W^*$ we stop the experiment and select the category with the largest cumulative sum. If $n_0 \leqq W^*$, then we eliminate any category π_j for which

$$\sum_{s=1}^{n_0} X_{js} < \max\left\{\sum_{s=1}^{n_0} X_{vs}\right\} - a^* + n_0\,\Delta/4,$$

where the max is taken over all k categories. If only one category is left, the experiment is terminated, otherwise we go on to the $(n_0 + 1)$ stage and take one measurement with each category not eliminated after the n_0th stage. Proceeding by induction, after the $(n_0 + t)$ stage $(t = 1, 2, \cdots, W^* - n_0)$ we eliminate any category π_j for which

$$\sum_{s=1}^{n_0+t} X_{js} < \max\left\{\sum_{s=1}^{n_0+t} X_{\nu s}\right\} - a^* + (n_0 + t)\Delta/4,$$

where the max is taken over all categories not eliminated after the $(n_0 + t - 1)$ stage. If more than one category is left after the W^* stage, we terminate the experiment at the $(W^* + 1)$ stage by selecting the remaining category with the greatest cumulative sum.

From the argument given in Section 4 it follows that

$$P[\pi_k \text{ is not selected} \mid \delta_2^*] \leqq \sum_{\nu=1}^{k-1} P\left[\sum_{s=1}^{n} X_{ks} < \sum_{s=1}^{n} X_{\nu s} - a^* + n\Delta/4 \quad \text{for some} \quad n, \quad n_0 \leqq n < \infty \;\middle|\; \delta_2^*\right].$$

Making use of the fact that the set of means $\{\bar{X}_{i\cdot}\}$ is independent of s^2, and that $\mu_k - \mu_\nu \geqq \Delta$ for all points in δ_2^*, and then following the detailed argument given on page 553 of [7], it is a routine matter to verify that

$$P[\pi_k \text{ is not selected} \mid \delta_2^*] \leqq \alpha.$$

In practical applications, the choice of n_0 is very important in determining the efficiency of the resulting sequential procedure. It is necessary to avoid taking n_0 too small since the resulting large value of g makes the sequential procedure inefficient, and it is also necessary to avoid having n_0 too large, as this would obviously also reduce the efficiency of the sequential procedure because no category can be eliminated or selected until the n_0th stage of the experiment.

Although an optimum procedure for selecting n_0 is unknown, it is hoped that the somewhat tentative procedure for selecting n_0 that shall now be described will prove useful. Two cases are distinguished: (1) no information about σ is available, and (2) an approximation σ' to the unknown common standard deviation σ is known from previous experience. For Case (1) we suggest selecting f (which determines n_0) so that g shall not exceed its limiting value by more than a specified amount (say ten percent); that is, select f to be the smallest integer so that $[((k-1)/\alpha)^{2/f} - 1]\frac{f}{2} < (1.1)\log[(k-1)/\alpha]$. If we denote the solution to this inequality by $\bar{f}$ (and the corresponding value of n_0 by $\bar{n}_0$) then for $\alpha = .05$ and $k = 2, 4, 10, 20$ the corresponding (approximate) value of $\bar{f}$ is 30, 45, 55, and 60 respectively, while for $\alpha = .01$ and $k = 2, 4, 10, 20$ the corresponding (approximate) value of $\bar{f}$ is 50, 60, 70, and 80 respectively. For case (2) let σ' be an approximation to σ known from previous experience, and let $N(\sigma')$ be the number of stages required by the fixed-sample size procedure corresponding to $\sigma = \sigma'$. Then we suggest taking $n_0 = \min\{N(\sigma')/3, \bar{n}_0\}$. However, when $N(\sigma')/3 < \bar{n}_0$ and the resulting value of g exceeds $\log[(k-1)/\alpha]$ by more than a specified amount (say by more than 25 percent) it is suggested that then the procedure of [5] should be preferred.

REFERENCES

[1] BARTLETT, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge Univ. Press.

[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] BECHHOFER, R. E., DUNNETT, C. W. and SOBEL, M. (1954). A two-sample multiple-decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41** 170–176.

[4] BECHHOFER, R. E. (1958). A sequential multiple decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs. *Biometrics* **14** 408–429.

[5] BECHHOFER, R. E. and BLUMENTHAL, S. (1962). A sequential multiple-decision procedure selecting the best one of several normal populations with a common unknown variance II: Monte Carlo sampling results and new computing formulae. *Biometrics* **18** 52–67.

[6] PAULSON, E. (1962). A sequential procedure for comparing several experimental categories with a standard or control. *Ann. Math. Statist.* **33** 438–443.

[7] PAULSON, E., (1963). A sequential decision procedure for choosing one of k hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.* **34** 549–554.

[8] STEIN, C. (1948). The selection of the largest of a number of means, (abstract). *Ann. Math. Statist.* **19** 429.