SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 789

May 1988
(updated April 1989)

**A COMPARISON OF THE PERFORMANCES
OF PROCEDURES FOR SELECTING THE
NORMAL POPULATION HAVING THE LARGEST MEAN
WHEN THE VARIANCES ARE KNOWN AND EQUAL**

by

Robert E. Bechhofer
and
David M. Goldsman*

Approved for public release; distribution unlimited.

Research partially supported by the U.S. Army Research office through the
Mathematical Sciences Institute of Cornell University.

*School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, GA 30332.

A COMPARISON OF THE PERFORMANCES
OF PROCEDURES FOR SELECTING THE
NORMAL POPULATION WITH THE LARGEST MEAN
WHEN THE VARIANCES ARE KNOWN AND EQUAL

Robert E. Bechhofer
School of Operations Research
and Industrial Engineering
Cornell University
Ithaca, NY 14853

David M. Goldsman
School of Industrial and
Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Key Words and Phrases: normal means selection problem; indifference-zone approach; selection procedures; single-stage, two-stage and sequential procedures; open, closed and truncated procedures; achieved $P\{CS\}$, expected stages, expected total number of observations; Monte Carlo sampling.

Dedicated to Ingram Olkin, affectionately known as "Red," on the occasion of his 65th birthday. Although his hair has greyed, his ability to make significant contributions to the statistical profession, and his capacity for warmth and good humor have not diminished with time.

ABSTRACT

We study the performance characteristics of procedures for selecting the normal population which has the largest mean when the variances are known and equal. The procedures studied are the single-stage procedure of Bechhofer, the closed two-stage procedure of Tamhane and Bechhofer, the open sequential procedure of Bechhofer, Kiefer and Sobel and a truncated version of that procedure by Bechhofer and Goldsman, the closed multi-stage procedure with elimination of Paulson and improved closed versions of that procedure by Fabian and by Hartmann. The performance characteristics studied are the achieved probability of a correct selection, the expected number of stages required to terminate experimentation, and the expected total number of observations required to terminate experimentation. Except for the single-stage procedure, all performance characteristics are estimated by Monte Carlo sampling. Based on these results, recommendations are made concerning which procedure to use in different circumstances.

TABLE OF CONTENTS

Abstract	i
1. Introduction and summary	1
2. Statement of the problem	2
3. Procedures considered	3
3.1 Single-stage procedure of Bechhofer	3
3.2 Two-stage procedure of Tamhane and Bechhofer	4
3.3 Sequential procedure of Bechhofer, Kiefer and Sobel, and a truncated version	5
3.4 Sequential procedure of Paulson, and the Fabian and Hartmann improvements	7
4. The performance of the procedures	9
5. Discussion of results	14
6. Concluding remarks	17
Acknowledgments	19
Bibliography	20

1. INTRODUCTION AND SUMMARY

Over the years considerable research effort has been devoted to devising procedures for selecting the normal population which has the largest population mean. Bechhofer (1954), adopting the so-called indifference-zone approach, developed a single-stage procedure for the case of common known variance. Following that paper, several authors, also adopting the indifference-zone approach, have proposed different procedures for dealing with this same problem. Among these is a closed two-stage procedure, an open sequential procedure, and several closed multi-stage procedures. All of these procedures guarantee the same indifference-zone probability requirement but each, at the time that it was introduced, had a different virtue and in some sense represented an improvement over the procedures proposed earlier. The purpose of the present article is to compare certain performance characteristics of these procedures—specifically their achieved probability of a correct selection, and for sequential and multi-stage procedures, their expected number of stages required to terminate experimentation and their expected total number of observations required to terminate experimentation.

In addition to the single-stage procedure of Bechhofer, the specific procedures that we study are the closed two-stage procedure of Tamhane and Bechhofer (1977,1979); the open sequential procedure of Bechhofer, Kiefer and Sobel (1968) and the truncated version of that procedure by Bechhofer and Goldsman (1987); the closed multi-stage procedure of Paulson (1964) and the improved version of that procedure by Fabian (1974) with further improvements by Hartmann (1988). The numerical estimates of the performance characteristics for the Tamhane–Bechhofer, Paulson, Fabian, and Hartmann procedures are new, and have not been reported in detail elsewhere; all of our results for these procedures were obtained by carefully executed Monte Carlo

sampling experiments. Based on these results we assess the virtues and drawbacks of the procedures, and discuss the alternatives available to the experimenter. Our findings should be of assistance to experimenters who wish to decide which of these procedures is appropriate to use in a particular real-life setting.

We formulate the selection problem in Section 2. The procedures under consideration will be described and their attributes reviewed in Section 3. Numerical estimates of the performance characteristics studied are given in Section 4. We discuss these results in Section 5. In Section 6 we state our conclusions.

2. STATEMENT OF THE PROBLEM

We assume that we have $k \geq 2$ normal populations $\Pi_1, \Pi_2, \dots, \Pi_k$ with unknown population means μ_i ($1 \leq i \leq k$) and a common known variance σ^2 . The ordered values of the μ_i are denoted by $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$. We further assume that the values of the $\mu_{[j]}$ ($1 \leq j \leq k$) are unknown, and that the pairing of the $\mu_{[j]}$ with the Π_i ($1 \leq i, j \leq k$) is completely unknown. The objective of the experiment, and the associated probability requirement, are stated in (2.1) and (2.2) below.

GOAL: To select the population associated with $\mu_{[k]}$. (2.1)

If the population selected by a procedure is indeed associated with $\mu_{[k]}$, then we say that a correct selection (CS) has been made. We limit consideration to procedures that guarantee the following indifference-zone requirement on the $P\{CS\}$:

PROBABILITY REQUIREMENT:

$$P\{CS\} \geq P^* \text{ whenever } \mu_{[k]} - \mu_{[k-1]} \geq \delta^*. \quad (2.2)$$

The quantities $\{\delta^*, P^*\}$ with $0 < \delta^* < \infty$, $1/k < P^* < 1$ are specified prior to the

start of experimentation. All of the aforementioned procedures have been shown to guarantee (2.2).

The present investigation is limited to the following representative cases: $k = 4$, $\delta^* = 0.2$, $P^* = 0.75, 0.90, 0.95, 0.99$, and three configurations of the population means, namely, (i) equally-spaced δ^* -apart (ES(δ^*)), i.e., $\mu_{[i+1]} - \mu_{[i]} = \delta^*$ ($1 \leq i \leq k-1$), (ii) least favorable (LF), i.e., $\mu_{[1]} = \mu_{[k-1]} = \mu_{[k]} - \delta^*$ and (iii) equal means (EM), i.e., $\mu_{[1]} = \mu_{[k]}$. The extent to which our findings can be extrapolated to additional $(k; \delta^*, P^*)$ and other configurations will be indicated.

3. PROCEDURES CONSIDERED

In order to make the present article self contained, we describe and critique in this section the procedures under consideration. The reader is referred to the cited articles for more detailed information concerning these procedures and their performance characteristics.

3.1 Single-stage procedure of Bechhofer

For this procedure (referred to herein as B) a common number n of independent observations X_{ij} ($1 \leq i \leq k$, $1 \leq j \leq n$) is taken in a single stage from each Π_i ($1 \leq i \leq k$), and the k sample means $\bar{x}_i = \sum_{j=1}^n x_{ij}/n$ ($1 \leq i \leq k$) are calculated. Let $\bar{x}_{[k]} = \max\{\bar{x}_1, \dots, \bar{x}_k\}$. The experimenter then selects the population that yielded $\bar{x}_{[k]}$ as the one associated with $\mu_{[k]}$. Here $n = [(\sigma c_{k,P^*}/\delta^*)^2]^+$ where c_{k,P^*} is a constant which is given (as $\sqrt{N} \lambda$) for $k = 2(1)10$ and selected P^* in Table I of Bechhofer (1954), and $[b]^+$ is the smallest integer greater than or equal to b .

The single-stage procedure can be quite conservative, i.e., if, unknown to the experimenter, the population means are in a very favorable configuration (e.g., widely spaced), then the actual achieved probability of a correct selection in the LF

configuration may exceed the specified P^* by a considerable amount. The experimenter thus has taken a larger number of observations than would have been required had the true configuration been known, and had that configuration then been taken into account when choosing n . The fact that observations are taken in a single stage does not permit the experimenter to gain information concerning the configuration of the population means as sampling progresses, as would be the case if observations were taken in (say) several stages. In particular, no information can be obtained as to whether or not any populations are likely to be contenders or non—contenders for "best." It is for this reason that multi—stage procedures were introduced.

3.2 Two—stage procedure of Tamhane and Bechhofer

Tamhane and Bechhofer (1977, 1979) proposed a two—stage selection procedure. At the end of the first stage, this procedure (referred to herein as T—B) screens out (permanently eliminates) populations indicated as being non—contending for best, and takes observations in the second stage only from those populations indicated as being contending. The cumulative sample mean based on first— and second—stage observations is calculated for populations which enter the second stage, and the experimenter selects the population that yielded the largest of these cumulative means. Thus this procedure can capitalize on favorable configurations of the population means permitting a reduction in the expected total number of observations required to terminate sampling, and in extreme situations leads to termination after the first stage. Implementation of the T—B procedure involves the use of three predetermined constants $(\hat{n}_1, \hat{n}_2, \hat{h})$ where $\hat{n}_1, \hat{n}_2$ are positive integers and $\hat{h} > 0$. The source of these constants is given below.

We now formally describe the T—B procedure. A common number $\hat{n}_1$ of independent observations $X_{ij}^{(1)}$ ($1 \leq j \leq \hat{n}_1$) is taken in the first stage from

Π_i ($1 \leq i \leq k$), and the k first-stage sample means $\bar{x}_i^{(1)} = \sum_{j=1}^{\hat{n}_1} x_{ij}^{(1)} / \hat{n}_1$ ($1 \leq i \leq k$) are calculated. Let $\bar{x}^{(1)} = \max\{\bar{x}_1^{(1)}, \dots, \bar{x}_k^{(1)}\}$. Determine the subset I of $\{1, 2, \dots, k\}$ where $I = \{i | \bar{x}_i^{(1)} \geq \bar{x}^{(1)} - \hat{h}\}$, and let Π_I denote the associated subset of $\{\Pi_1, \Pi_2, \dots, \Pi_k\}$. If Π_I consists of one population, stop sampling and select the population yielding $\bar{x}^{(1)}$ as the one associated with $\mu_{[k]}$. If Π_I consists of more than one population, proceed to the second stage and take $\hat{n}_2$ additional observations $X_{ij}^{(2)}$ ($1 \leq j \leq \hat{n}_2$) from each population in Π_I . Compute the cumulative sample means $\bar{x}_i = (\sum_{j=1}^{\hat{n}_1} x_{ij}^{(1)} + \sum_{j=1}^{\hat{n}_2} x_{ij}^{(2)}) / (\hat{n}_1 + \hat{n}_2)$ for $i \in I$. The experimenter then selects the population that yielded $\max\{\bar{x}_i | i \in I\}$ as the one associated with $\mu_{[k]}$.

For given k and specified $\{\delta^*, P^*\}$ the procedure depends on the three constants $(\hat{n}_1, \hat{n}_2, \hat{h})$. Here $\hat{n}_1 = [(\sigma \hat{c}_1 / \delta^*)^2]^+$, $\hat{n}_2 = [(\sigma \hat{c}_2 / \delta^*)^2]^+$ and $\hat{h} = \hat{d} \delta^* / \hat{c}_1$ where $(\hat{c}_1, \hat{c}_2, \hat{d})$ are found for $k = 2$ in Table I (E) of Tamhane and Bechhofer (1979) and for $k = 3(1)10, 12, 15, 25$ and selected P^* in Table II of that article. The choice $(\hat{n}_1, \hat{n}_2, \hat{h})$ guarantees (2.2) and minimizes $\max_{\mu} E\{\text{Total number of observations}\}$ where

$\mu = (\mu_1, \mu_2, \dots, \mu_k)$. This minimax two-stage procedure has the property that $E\{\text{Total number of observations}\} < kn$, uniformly in μ for all $\{\delta^*, P^*\}$, where n is the number of observations per population required by the corresponding single-stage procedure.

Multi-stage procedures offer the potential of further improvements over the two-stage procedure. We describe such procedures in Sections 3.3 and 3.4.

3.3 Sequential procedure of Bechhofer, Kiefer and Sobel, and a truncated version

Bechhofer, Kiefer and Sobel (1968), Section 12.6.1.1, proposed a fully sequential open selection procedure. For this procedure (referred to herein as B-K-S) independent

vector—observations consisting of one observation from each of the k populations are taken one vector—at—a—time until a stopping rule calls for termination of sampling. The procedure can capitalize on favorable configurations of the population means, thus leading to early termination of sampling. In fact, if $\mu_{[k]} - \mu_{[k-1]} \gg \sigma$, then with high probability sampling will stop after a single stage (requiring therefore only a total of k scalar observations).

Let x_{ij} ($1 \leq i \leq k$, $j = 1, 2, \dots$) denote the j th observation from Π_i . At stage m of experimentation ($m = 1, 2, \dots$), observe the random vector $(X_{1m}, X_{2m}, \dots, X_{km})$. Let $y_{im} = \sum_{j=1}^m x_{ij}$ ($1 \leq i \leq k$), and denote the ordered values of the y_{im} by $y_{[1]m} < \dots < y_{[k]m}$. Calculate $z_m = \sum_{i=1}^{k-1} \exp\{-\delta^*(y_{[k]m} - y_{[i]m})/\sigma^2\}$. Stop sampling when, for the first time, $z_m \leq (1-P^*)/P^*$. Let N , a random variable, denote the value of m at termination. After stopping, the experimenter selects the population that yielded $y_{[k]N}$ as the one associated with $\mu_{[k]}$.

This procedure has the drawback that if $\mu_{[k]} - \mu_{[1]}$ is small, then the distribution of N is highly skewed to the right, and large values of N can occur with sizable probability. (See Table I in Bechhofer—Goldsman (1987).) Also, the variance of N can be very large. To avoid these undesirable effects, B and G studied a truncated version of the B—K—S procedure (referred to herein as $(B-K-S)_T$); truncation is possible here because, when the population means are in the LF—configuration, the achieved $P\{CS\}$ is greater than P^* . For this version sampling is stopped when, for the first time, either $z_m \leq (1-P^*)/P^*$ or $m = n_0$, whichever occurs first. Let N denote the value of m at termination, and let $n_0(k; \delta^*, P^*) = n_0$ (say) be predetermined as the smallest integer that will guarantee (2.2) using $(B-K-S)_T$. After stopping, the experimenter selects the population that yielded $y_{[k]N}$ as the one associated with $\mu_{[k]}$. The truncated procedure retains all of the virtues of the original procedure, and in addition reduces $E\{N\}$ and $\text{Var}\{N\}$. Values of n_0 are given for $k = 2, 3, 4, 5$ in

Tables II, III, IV, V, respectively, of B-G (1987) and in Tables I, II, III, IV of B-G (1988); each table of B-G (1987) gives values for $P^* = 0.75, 0.90, 0.95, 0.99$ with $\delta^* = 0.2(0.2)0.8$ while each table of B-G (1989) gives values for $P^* = 0.75, 0.90, 0.95, 0.99$ with $\delta^* = 0.3(0.2)0.7$.

Neither B-K-S nor $(B-K-S)_T$ permits elimination of populations indicated as being non-contending for best. Such elimination-type procedures are considered in Section 3.4. (T-B which was considered in Section 3.2 is also such a procedure.)

3.4 Sequential procedure of Paulson, and the Fabian and Hartmann improvements

The B-K-S procedure has two drawbacks: i) It is open, i.e., before experimentation starts it is not possible to give a finite upper bound on the number of stages required to terminate experimentation, and ii) It does not eliminate from sampling populations which, based on the earlier stages, would appear to be out of contention for selection as "best." The following closed sequential procedure of Paulson (1964) (referred to herein as P), which employs elimination of populations, overcomes both of these drawbacks. For this procedure independent vector-observations are taken, one vector-at-a-time, until a stopping rule calls for termination of sampling; each vector consists of one observation from each population not eliminated at an earlier stage. Paulson's procedure is actually a family of procedures which depends on a constant λ ($0 \leq \lambda \leq \delta^*/2$) which is chosen by the experimenter. Paulson recommended the choice $\lambda = \delta^*/4$. However, the choice $\lambda = \delta^*/2$ has the desirable property that the maximum possible number of stages to termination of the experiment is minimized for given k and specified $\{\delta^*, P^*\}$.

Let x_{ij} ($1 \leq i \leq k, j = 1, 2, \dots$) denote the j th observation from Π_i . At stage m ($m = 1, 2, \dots$) let $y_{im} = \sum_{j=1}^m x_{ij}$ ($1 \leq i \leq k$). Denote the ordered values of the y_{im} for the R populations still retained at stage m by $y_{[1]m} < \dots < y_{[R]m}$. Here $R \geq 2$ is a random variable. Let $a_\lambda = (\sigma^2/(\delta^* - \lambda)) \ln\{(k-1)/(1-P^*)\}$ and let W_λ denote

the largest integer $< a_\lambda/\lambda$. Start sampling by taking one observation from each of the k populations. Eliminate from further sampling and consideration any population Π_i for which $a_\lambda - \lambda < y_{[k]1} - y_{i1}$. If all but one population is eliminated after the first stage, then stop sampling and select the remaining population as the one associated with $\mu_{[k]}$. Otherwise, proceed to the second stage and take one observation from each population not yet eliminated. In general, at stage m ($2 \leq m \leq W_\lambda$) take one observation from each population not eliminated after the $(m-1)$ st stage, and then eliminate from further consideration any remaining population Π_i for which $a_\lambda - m\lambda < y_{[R]m} - y_{im}$. If all but one population is eliminated after the m th stage, stop sampling and select the remaining population as the one associated with $\mu_{[k]}$; otherwise proceed to the $(m+1)$ st stage. If more than one population remains after stage W_λ , take one additional observation from each of these populations. Then select the population with the largest sum of the $(W_\lambda+1)$ observations as the one associated with $\mu_{[k]}$.

The Fabian (1974) improvement of the P-procedure (referred to herein as P-F) is obtained by replacing the constant $\{(k-1)/(1-P^*)\}$ by the smaller constants $\{(k-1)/2(1-P^*)\}$ for $\lambda = \delta^*/2$ and by $1/\theta$ for $\lambda = \delta^*/4$ where θ solves $(1-P^*)/(k-1) = \theta - \frac{1}{2} \theta^{4/3}$.

Hartmann (1988) improved the P-F procedure (and thus the P-procedure). For Hartmann's procedure (referred to herein as P-H) the constant $\{(k-1)/(1-P^*)\}$ of the P-procedure is replaced by $1/\theta$ where $\theta = 2[1 - (P^*)^{1/(k-1)}]$ for $\lambda = \delta^*/2$, and θ solves $1 - (P^*)^{1/(k-1)} = \theta - \frac{1}{2} \theta^{4/3}$ for $\lambda = \delta^*/4$. For $k > 2$ the constant a_λ is smaller for the P-H procedure than for the P-F procedure.

The P, P-F and P-H procedures can be truncated while still guaranteeing (2.2). The effect of truncation is to decrease the achieved $P\{CS\}$. At the same time $E\{N\}$, $\text{Var}\{N\}$ and $E\{T\}$, $\text{Var}\{T\}$ are reduced; here T is the total number of

observations required to terminate sampling. We do not study truncation of these procedures in the present article.

The P, P-F, and P-H procedures never require more than $W_{\lambda}+1$ stages to terminate sampling. The effectiveness of these procedures can be measured not only in terms of $E\{N\}$ but also in terms of $E\{T\}$. The appropriate measure(s) to use will depend on the particular practical situation at hand. For example, stages can often be equated to time; this would be the situation if only one vector of observations could be obtained per day (say) as with certain production processes. Thus the choice of measure might depend on whether it is more desirable to minimize the expected duration of the experiment than to minimize the expected total number of observations taken in the experiment. The answer will depend on various relative costs. Hartmann's (Fabian's) modification improves the performance of Fabian's (Paulson's) procedure uniformly in $(k; \delta^*, P^*)$ and μ for both measures.

4. THE PERFORMANCE OF THE PROCEDURES

Monte Carlo (MC) estimates of the achieved $P\{CS\}$, $E\{N\}$, and $E\{T\}$ are given for $k = 4$, $\delta^* = 0.2$ in Tables 4.1, 4.2, 4.3, 4.4 for $P^* = 0.75, 0.90, 0.95, 0.99$, respectively. Three configurations of the population means, namely, (i) equally-spaced δ^* -apart ($ES(\delta^*)$), (ii) least favorable (LF), and (iii) equal means (EM) were studied for B, T-B, B-K-S, $(B-K-S)_T$, P using $\lambda = \delta^*/4$ and $\delta^*/2$ (P, $\delta^*/4$; P, $\delta^*/2$), P-F using $\lambda = \delta^*/4$ and $\delta^*/2$ (P-F, $\delta^*/4$; P-F, $\delta^*/2$), and P-H using $\lambda = \delta^*/4$ and $\delta^*/2$ (P-H, $\delta^*/4$; P-H, $\delta^*/2$). All of the results for T-B, B-K-S and $(B-K-S)_T$, (P, $\delta^*/4$ and $\delta^*/2$), (P-F, $\delta^*/4$ and $\delta^*/2$), and (P-H, $\delta^*/4$ and $\delta^*/2$) were obtained by MC sampling. The MC results for B-K-S and $(B-K-S)_T$ in the

TABLE 4.1

Estimates of achieved probability of a correct selection ($\bar{p}$ or $\bar{w}$), expected number of stages to terminate sampling ($\bar{n}$), and expected total number of observations to terminate sampling ($\bar{t}$) for three configurations of the population means, and selected procedures.
($k = 4, \delta^* = 0.2, P^* = 0.75$)

Procedure	Equally-spaced δ^* apart: $\mu_{[i+1]} - \mu_{[i]} = \delta^* \ (1 \leq i \leq 3)$			Least favorable: $\mu_{[1]} = \mu_{[3]} = \mu_{[4]} - \delta^*$			Equal means: $\mu_{[1]} = \mu_{[4]}$	
	$\bar{p}$	$\bar{n}$	$\bar{t}$	$\bar{p}$ or $\bar{w}$	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$
B $n=71$	0.8796	71	284	0.7508	71	284	71	284
T-B $\hat{n}_1=34, \hat{n}_2=32$ $\hat{h}=0.6107$	0.8729 (0.0030)	65.2 (<0.1)	237.9 (0.3)	0.7361 (0.0028)	65.9 (<0.1)	258.9 (<0.1)	66.0 (<0.1)	261.8 (<0.1)
B-K-S	0.8493 (0.0033)	34.4 (0.2)	137.8 (0.7)	0.7759 (<0.0001)	53.3 (0.2)	213.1 (0.6)	72.0 (0.4)	288.0 (1.8)
(B-K-S) _T $n_0=94$	0.8439 (0.0033)	33.6 (0.2)	134.5 (0.7)	0.7507 (0.0004)	50.2 (0.1)	200.8 (0.4)	60.5 (0.2)	242.2 (1.0)
(P, $\delta^*/4$)	0.9388 (0.0022)	66.4 (0.3)	191.8 (0.7)	0.8621 (0.0023)	84.5 (0.3)	262.6 (0.7)	110.1 (0.4)	335.7 (1.1)
(P, $\delta^*/2$)	0.9626 (0.0017)	84.4 (0.3)	252.2 (0.8)	0.9065 (0.0019)	104.0 (0.2)	341.3 (0.7)	134.3 (0.4)	437.1 (1.1)
(P-F, $\delta^*/4$) $\theta = 0.10954$	0.9228 (0.0024)	57.9 (0.3)	167.7 (0.6)	0.8294 (0.0024)	73.6 (0.2)	227.1 (0.6)	92.6 (0.4)	280.8 (1.0)
(P-F, $\delta^*/2$)	0.9225 (0.0024)	58.9 (0.2)	176.3 (0.6)	0.8293 (0.0024)	72.8 (0.2)	233.3 (0.5)	87.2 (0.3)	277.7 (0.8)
(P-H, $\delta^*/4$) $\theta = 0.12154$	0.9164 (0.0025)	54.6 (0.3)	158.2 (0.6)	0.8200 (0.0025)	69.3 (0.2)	213.0 (0.6)	85.9 (0.3)	259.6 (0.9)
(P-H, $\delta^*/2$) $\theta = 0.18288$	0.9188 (0.0025)	55.6 (0.2)	166.3 (0.5)	0.8151 (0.0025)	68.4 (0.2)	219.0 (0.5)	80.7 (0.2)	256.3 (0.7)

TABLE 4.2

Estimates of achieved probability of a correct selection ($\bar{p}$ or $\bar{w}$), expected number of stages to terminate sampling ($\bar{n}$), and expected total number of observations to terminate sampling ($\bar{t}$) for three configurations of the population means, and selected procedures.

($k = 4$, $\delta^* = 0.2$, $P^* = 0.90$)

Procedure	Equally-spaced δ^* apart: $\mu_{[i+1]} - \mu_{[i]} = \delta^* \ (1 \leq i \leq 3)$			Least favorable: $\mu_{[1]} = \mu_{[3]} = \mu_{[4]} - \delta^*$			Equal means: $\mu_{[1]} = \mu_{[4]}$	
	$\bar{p}$	$\bar{n}$	$\bar{t}$	$\bar{p}$ or $\bar{w}$	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$
B $n=151$	0.9588	151	604	0.9008	151	604	151	604
T-B $\hat{n}_1=78, \hat{n}_2=79$ $\hat{h}=0.1932$	0.9583 (0.0018)	116.6 (0.4)	394.7 (0.8)	0.9035 (0.0019)	135.4 (0.2)	474.7 (0.7)	150.9 (0.2)	545.0 (0.8)
B-K-S	0.9354 (0.0022)	61.6 (0.3)	246.3 (1.3)	0.9124 (<0.0001)	95.8 (0.4)	383.2 (1.4)	167.8 (1.1)	671.1 (4.2)
(B-K-S) _T $n_0=205$	0.9350 (0.0023)	61.8 (0.3)	247.0 (1.3)	0.9006 (0.0001)	93.1 (0.1)	372.6 (0.4)	136.7 (0.5)	546.6 (2.2)
(P, $\delta^*/4$)	0.9745 (0.0014)	93.2 (0.4)	269.5 (1.0)	0.9333 (0.0016)	119.7 (0.3)	381.5 (0.9)	175.5 (0.6)	542.4 (1.7)
(P, $\delta^*/2$)	0.9840 (0.0011)	116.0 (0.4)	347.3 (1.0)	0.9620 (0.0012)	142.4 (0.3)	477.9 (0.9)	202.2 (0.5)	669.8 (1.5)
(P-F, $\delta^*/4$) $\theta = 0.04022$	0.9693 (0.0016)	87.1 (0.4)	252.3 (0.9)	0.9265 (0.0017)	113.1 (0.3)	358.7 (0.9)	160.8 (0.6)	495.8 (1.6)
(P-F, $\delta^*/2$)	0.9702 (0.0016)	92.3 (0.3)	276.0 (0.8)	0.9205 (0.0017)	114.0 (0.3)	376.3 (0.8)	150.7 (0.4)	492.1 (1.2)
(P-H, $\delta^*/4$) $\theta = 0.04175$	0.9720 (0.0015)	86.1 (0.4)	249.6 (0.9)	0.9247 (0.0017)	110.8 (0.3)	351.0 (0.9)	158.5 (0.6)	489.7 (1.5)
(P-H, $\delta^*/2$) $\theta = 0.06902$	0.9684 (0.0016)	90.2 (0.3)	269.8 (0.8)	0.9208 (0.0017)	112.0 (0.3)	369.8 (0.8)	147.9 (0.4)	482.8 (1.1)

TABLE 4.3

Estimates of achieved probability of a correct selection ($\bar{p}$ or $\bar{w}$), expected number of stages to terminate sampling ($\bar{n}$), and expected total number of observations to terminate sampling ($\bar{t}$) for three configurations of the population means, and selected procedures.
 $(k = 4, \delta^* = 0.2, P^* = 0.95)$

Procedure	Equally-spaced δ^* apart: $\mu_{[i+1]} - \mu_{[i]} = \delta^* (1 \leq i \leq 3)$			Least favorable: $\mu_{[1]} = \mu_{[3]} = \mu_{[4]} - \delta^*$			Equal means: $\mu_{[1]} = \mu_{[4]}$	
	$\bar{p}$	$\bar{n}$	$\bar{t}$	$\bar{p}$ or $\bar{w}$	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$
B $n=213$	0.9805	213	852	0.9502	213	852	213	852
T-B $\hat{n}_1=115, \hat{n}_2=110$ $\hat{h}=0.1358$	0.9794 (0.0013)	149.4 (0.5)	530.3 (1.0)	0.9495 (0.0014)	174.5 (0.4)	614.2 (1.0)	210.9 (0.3)	748.8 (1.2)
B-K-S	0.9656 (0.0017)	80.8 (0.5)	323.1 (1.8)	0.9566 (<0.0001)	122.4 (0.4)	489.8 (1.8)	261.2 (1.7)	1044.7 (6.6)
(B-K-S) _T $n_0=290$	0.9628 (0.0017)	81.0 (0.4)	324.2 (1.8)	0.9504 (0.0002)	120.6 (0.3)	482.6 (1.2)	202.9 (0.8)	811.7 (3.0)
(P, $\delta^*/4$)	0.9882 (0.0010)	113.1 (0.5)	327.5 (1.1)	0.9644 (0.0012)	144.2 (0.4)	467.2 (1.1)	229.3 (0.8)	718.2 (2.1)
(P, $\delta^*/2$)	0.9916 (0.0008)	140.6 (0.5)	421.5 (1.1)	0.9796 (0.0009)	169.7 (0.4)	577.9 (1.1)	255.7 (0.6)	855.8 (1.8)
(P-F, $\delta^*/4$) $\theta = 0.01925$	0.9869 (0.0010)	108.6 (0.5)	314.5 (1.1)	0.9621 (0.0012)	138.9 (0.4)	448.9 (1.1)	220.0 (0.7)	685.8 (2.0)
(P-F, $\delta^*/2$)	0.9856 (0.0011)	116.7 (0.4)	349.3 (1.0)	0.9603 (0.0013)	142.8 (0.3)	479.1 (0.9)	201.5 (0.5)	668.2 (1.5)
(P-H, $\delta^*/4$) $\theta = 0.01959$	0.9818 (0.0012)	107.2 (0.5)	310.9 (1.1)	0.9599 (0.0013)	138.8 (0.4)	447.8 (1.1)	216.3 (0.7)	675.9 (2.0)
(P-H, $\delta^*/2$) $\theta = 0.03390$	0.9854 (0.0011)	115.2 (0.4)	345.9 (1.0)	0.9604 (0.0013)	141.8 (0.3)	476.3 (0.9)	200.0 (0.5)	662.2 (1.5)

TABLE 4.4

Estimates of achieved probability of a correct selection ($\bar{p}$ or $\bar{w}$), expected number of stages to terminate sampling ($\bar{n}$), and expected total number of observations to terminate sampling ($\bar{t}$) for three configurations of the population means, and selected procedures.
 $(k = 4, \delta^* = 0.2, P^* = 0.99)$

Procedure	Equally-spaced δ^* apart: $\mu_{[i+1]} - \mu_{[i]} = \delta^* (1 \leq i \leq 3)$			Least favorable: $\mu_{[1]} = \mu_{[3]} = \mu_{[4]} - \delta^*$			Equal means: $\mu_{[1]} = \mu_{[4]}$	
	$\bar{p}$	$\bar{n}$	$\bar{t}$	$\bar{p}$ or $\bar{w}$	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$
B $n=361$	0.9964	361	1444	0.9901	361	1444	361	1444
T-B $\hat{n}_1=220, \hat{n}_2=158$ $\hat{h}=0.08243$	0.9963 (0.0006)	236.2 (0.4)	912.5 (0.9)	0.9894 (0.0007)	256.5 (0.4)	964.8 (1.0)	348.2 (0.6)	1242.5 (1.9)
B-K-S	0.9938 (0.0007)	123.6 (0.7)	494.3 (2.6)	0.9914 (<0.0001)	176.5 (0.4)	705.9 (1.6)	542.5 (3.5)	2169.8 (13.9)
(B-K-S) _T $n_0=480$	0.9921 (0.0008)	122.4 (0.6)	489.6 (2.5)	0.9902 (<0.0001)	175.8 (0.4)	703.0 (1.6)	370.8 (1.2)	1483.3 (4.7)
(P, $\delta^*/4$)	0.9969 (0.0005)	156.2 (0.6)	455.1 (1.4)	0.9916 (0.0006)	196.9 (0.5)	656.6 (1.4)	364.7 (1.1)	1163.7 (3.1)
(P, $\delta^*/2$)	0.9981 (0.0004)	194.2 (0.6)	585.3 (1.4)	0.9956 (0.0004)	231.7 (0.4)	806.7 (1.4)	383.2 (0.8)	1310.0 (2.4)
(P-F, $\delta^*/4$) $\theta = 0.003610$	0.9958 (0.0006)	153.8 (0.6)	447.9 (1.4)	0.9920 (0.0006)	194.7 (0.5)	648.0 (1.4)	356.1 (1.1)	1137.6 (3.0)
(P-F, $\delta^*/2$)	0.9977 (0.0004)	170.9 (0.5)	514.3 (1.3)	0.9919 (0.0006)	206.0 (0.4)	710.9 (1.3)	328.5 (0.7)	1115.0 (2.1)
(P-H, $\delta^*/4$) $\theta = 0.003623$	0.9974 (0.0005)	154.6 (0.6)	449.7 (1.4)	0.9925 (0.0006)	192.9 (0.5)	643.4 (1.4)	357.3 (1.1)	1139.3 (3.1)
(P-H, $\delta^*/2$) $\theta = 0.006689$	0.9970 (0.0005)	171.2 (0.5)	514.8 (1.2)	0.9912 (0.0006)	205.9 (0.4)	709.9 (1.3)	328.7 (0.7)	1112.9 (2.1)

LF— and EM—configurations were abstracted from B—G (1987). All other MC results are new and were prepared especially for this article.

The $P\{CS\}$ achieved by the B procedure for the particular sample sizes used was calculated by quadrature. The achieved $P\{CS\}$ for the other procedures was estimated by the observed proportion of correct selections ($\bar{p}$), except for B—K—S and $(B—K—S)_T$ used in the LF—configuration, where a more precise unbiased estimate ($\bar{w}$) was used. (See B—G (1987), Remark 3.3.) $E\{N\}$ and $E\{T\}$ were estimated by the observed average number of stages to terminate sampling ($\bar{n}$) and the observed average total number of observations to terminate sampling ($\bar{t}$), respectively. In each of the four tables, the entries under $\bar{n}$ for the B and T—B procedures are the numbers of "vectors" of observations taken in the single— and two—stage procedures, respectively; thus these entries are comparable to the other entries in the $\bar{n}$ columns.

The number of independent MC replications on which each estimate is based for each procedure is 12,000 for the $ES(\delta^*)$ —configuration, 24,000 for the LF—configuration and 12,000 for the EM—configuration, except that some of the estimates for $(B—K—S)_T$ in the LF—configuration were based on more than 24,000 independent replications. The number in parentheses below each estimate is the estimated standard error of the estimate above it.

5. DISCUSSION OF RESULTS

In Table 5.1 we have indicated the procedure that, based on our results, dominates in terms of minimum $\bar{n}$ or minimum $\bar{t}$ for each of the three configurations of the population means; these are given for four P^* values.

First we note that in terms of minimum $\bar{n}$, $(B—K—S)_T$ dominates in the LF—configuration for all P^* values. This is not surprising since in that configuration, B—K—S is equivalent to a sequential probability ratio test for the corresponding

identification problem (see B-K-S (1968), Section 4.3.1), and the effect of truncation is effectively to eliminate the "excess," i.e., achieved $P\{CS\} - P^*$. The domination of $(B-K-S)_T$ in terms of minimum $\bar{n}$ extends here to the $ES(\delta^*)$ -configuration (and we conjecture that it would continue to hold for even more favorable configurations). $(B-K-S)_T$ also dominates in all three configurations for both $\bar{n}$ and $\bar{t}$ when $P^* = 0.75$; it is at this (the lowest) P^* -value that the excess is greatest in the LF-configuration, and as noted above, truncation essentially eliminates that excess.

TABLE 5.1

Procedure having minimum $\bar{n}$ or minimum $\bar{t}$ for
each of three configurations of the population means
($k = 4$, $\delta^* = 0.2$, selected P^*)

P^*	ES(δ^*)-configuration		LF-configuration		EM-configuration	
	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$	$\bar{n}$	$\bar{t}$
0.99	$(B-K-S)_T$	$(P-H, \delta^*/4)$	$(B-K-S)_T$	$(P-H, \delta^*/4)$	$(P-H, \delta^*/2)$	$(P-H, \delta^*/2)$
0.95	$(B-K-S)_T$	$(P-H, \delta^*/4)$	$(B-K-S)_T$	$(P-H, \delta^*/4)$	$(P-H, \delta^*/2)$	$(P-H, \delta^*/2)$
0.90	$(B-K-S)_T$	$(B-K-S)_T$	$(B-K-S)_T$	$(P-H, \delta^*/4)$	$(B-K-S)_T$	$(P-H, \delta^*/2)$
0.75	$(B-K-S)_T$	$(B-K-S)_T$	$(B-K-S)_T$	$(B-K-S)_T$	$(B-K-S)_T$	$(B-K-S)_T$

Note: In Table 4.4 the recorded value of $\bar{t}$ in the $ES(\delta^*)$ -configuration for $P^* = 0.99$ is smaller for $(P-F, \delta^*/4)$ than for $(P-H, \delta^*/4)$ although the difference is not statistically significant. However, $E\{T\}$ of which $\bar{t}$ is an estimate must be smaller for $(P-H, \delta^*/4)$ than for $(P-F, \delta^*/4)$ since the continuation region for the former is

contained in the continuation region for the latter. The observed difference in $\bar{t}$ means is due to sampling errors. Thus we have "credited" $(P-H, \delta^*/4)$ with the smaller value of $\bar{t}$ for $P^* = 0.99$ in the $ES(\delta^*)$ -configuration and have similarly "credited" $(P-H, \delta^*/2)$ with the smaller value of $\bar{n}$ for $P^* = 0.99$ in the EM-configuration. A similar reversal for $\bar{n}$ occurs in the $ES(\delta^*)$ -configuration between $(B-K-S)$ and $(B-K-S)_T$ for $P^* = 0.95$ and 0.90 , and for $\bar{t}$ in the $ES(\delta^*)$ -configuration for $P^* = 0.90$. In each of these reversals we have "credited" $(B-K-S)_T$ with the smaller $\bar{n}$ or $\bar{t}$ value.

Also noteworthy is the fact that in terms of minimum $\bar{t}$, $(P-H, \delta^*/2)$ dominates in the EM-configuration for the largest P^* -values, namely, $P^* = 0.90, 0.95, 0.99$. This is a consequence of the fact that for each P^* , $(P-H, \delta^*/2)$ has the smallest predetermined upper bound on the maximum number of stages to terminate sampling. In terms of minimum $\bar{t}$, $(P-H, \delta^*/4)$ dominates in the LF-configuration for the largest P^* -values, namely, $P^* = 0.90, 0.95, 0.99$, and in the $ES(\delta^*)$ -configuration for $P^* = 0.95, 0.99$; elimination of populations is easier in the earlier stages with $(P-H, \delta^*/4)$ than with $(P-H, \delta^*/2)$.

Further MC studies carried out for $k = 5$, $\delta^* = 0.4$ established that $(B-K-S)_T$, $(P-H, \delta^*/4)$ and $(P-H, \delta^*/2)$ dominated in terms of minimum $(\bar{n}, \bar{t})$ in essentially the same regions as those described in Table 5.1. The results of these studies are available from the authors on request.

In our analysis of the performance of the various procedures we have not focused on the achieved $P\{CS\}$ provided only that (2.2) is guaranteed. Thus, for example, $(P-H, \delta^*/2)$ achieved a $P\{CS\}$ of 0.815 (approximately) in the LF-configuration for $P^* = 0.75$. This large overprotection could be eliminated by truncation with consequent substantial reductions in $E\{N\}$ and $E\{T\}$ in all configurations. However, the preparation of a table of such truncation numbers would be excessively expensive (and

in particular, much more costly than the corresponding tables for $(B-K-S)_T$ in B-G (1987, 1989)).

Remark 5.1: In Table 4.1, the MC estimate of the achieved $P\{CS\}$ for the T-B procedure, i.e., $\bar{p} = 0.7361$, is significantly lower (statistically) than the specified $P^* = 0.75$. It has been brought to our attention by Professor Ajit Tamhane that this result is explained by the fact that the constants $(\hat{c}_1, \hat{c}_2, \hat{d})$ for $k = 4$, $P^* = 0.75$ in Table II of T-B (1979) on which $(\hat{n}_1, \hat{n}_2, \hat{h})$ are based are in error because they do not satisfy the condition $(\hat{c}_1)^2 + (\hat{c}_2)^2 > (c_{k,P^*})^2$; this condition is necessary (but not sufficient) in order for the two-stage procedure T-B to guarantee the same $\{\delta^*, P^*\}$ condition that the corresponding single-stage procedure B guarantees. The violation of this condition results in $\hat{n}_1$ and/or $\hat{n}_2$ being too small, i.e., $\hat{n}_1 + \hat{n}_2 < n$. Thus for $k = 4$, $P^* = 0.75$, $\delta^* = 0.2$, we have $\hat{n}_1 + \hat{n}_2 = 34 + 32 < n = 71$ which results in the specified P^* value not being attained in the LF-configuration. Several other entries in Table II of T-B (1979) for $P^* = 0.75$ suffer from the same error (although the discrepancies are quite small).

6. CONCLUDING REMARKS

We note from Table 5.1 that neither the single-stage procedure (B) nor the two-stage procedure (T-B) dominates either in terms of $\bar{n}$ or $\bar{t}$ for any configuration or any P^* -value. However, both of these procedures play important roles in real-life situations in which truly sequential experimentation is not a feasible possibility. This is the case, for example, in agriculture where a stage is usually a growing season. Most such experiments would thus require a single-stage procedure; if a two-stage procedure is feasible, then T-B might well be preferable to B, since the former uniformly dominates the latter in terms of $E\{T\}$ while still guaranteeing (2.2).

We mention that at the time that T-B (1979) was written, the LF-configuration for that two-stage procedure had not been determined for $k > 2$. As a consequence a conservative lower bound on the $P\{CS\}$ was used which, although it guaranteed (2.2), resulted in a larger $E\{T\}$ than would have been the case if the LF-configuration had been known, and corresponding values of $(\hat{n}_1, \hat{n}_2, \hat{h})$ based on that bound had been calculated. However, Miescke and Sehr (1980) established the LF-configuration for $k = 3$, and more recently, Sehr (1988), and independently Bhandari and Chaudhuri (1987) have established the LF-configuration for all k . Calculations based on the now-known (hitherto conjectured) LF-configuration could be expected to yield substantial decreases in $E\{T\}$ (relative to the present $E\{T\}$ -values), especially for large k and/or low P^* . But calculations based on the exact $P\{CS|LF\}$ given by equation (5.6) of T-B (1977) with δ replaced by δ^* would be prohibitively expensive, at least at the present time; this calculation involves the numerical evaluation of a four-fold integral.

For minimizing $E\{N\}$ when sequential experimentation is feasible, $(B-K-S)_T$ would appear to be preferable except for configurations which are suspected as being close to the EM-configuration when P^* is high, in which situation $(P-H, \delta^*/2)$ should be used; however, tables of n_0 -values for $(B-K-S)_T$ have been prepared only for $k = 2(1)5$. For $k > 5$ we recommend the use of $(P-H, \delta^*/4)$ for all but the EM-configuration, and $(P-H, \delta^*/2)$ for the latter configuration. The beauty of $(P-H, \delta^*/4)$ or $(P-H, \delta^*/2)$ is that in order to implement the procedure it is only necessary to determine the appropriate θ -value for the particular (k, P^*) of interest; this is a trivial calculation.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. Army Research office through the Mathematical Sciences Institute of Cornell University. The authors wish to thank Mr. Mark Hartmann for providing us with a copy of his article on an improvement of the Paulson–Fabian procedure; he has approved our including his improved procedure among our comparisons. We also acknowledge the assistance of Professor Tom Santner who calculated for us the exact $P\{\text{CS}\}$ –values achieved by the single–stage procedure, and made many constructive suggestions. We are indebted to Professor Ajit Tamhane who solved the mystery reported on in Remark 5.1, and for helpful comments. Finally, we wish to thank Ms. Kathy King for typing our manuscript; as always she has done a beautiful job.

BIBLIOGRAPHY

- 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.
- Bechhofer, R.E. and Goldsman, D.M. (1987). Truncation of the Bechhofer–Kiefer–Sobel sequential procedure for selecting the normal population which has the largest mean. Comm. Statist.—Simula. Computa., B16(4), 1067–1091.
- Bechhofer, R.E. and Goldsman, D.M. (1989). Truncation of the Bechhofer–Kiefer–Sobel sequential procedure for selecting the normal population which has the largest mean (III): supplementary truncation numbers and resulting performance characteristics. Comm. Statist.—Simula. Computa., B18(1), 63–81.
- Bechhofer, R.E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures (with special reference to Koopman–Darmois populations), Chicago, The University of Chicago Press.
- Bhandari, S.K. and Chaudhuri, A.R. (1987). On two conjectures about two-stage selection procedures. To appear in Sankhyā, Ser. B.
- Fabian, V. (1974). Note on Anderson's sequential procedures with triangular boundary. Ann. Statist. 2, 170–176.
- Hartmann, M. (1988). An improvement on Paulson's sequential ranking procedure. Comm. Statist.—Sequential Analysis 7 (4), 363–372.
- Miescke, K.–J. and Sehr, J. (1980). On a conjecture concerning the least favorable configuration of certain two-stage selection procedures. Comm. Statist.—Theor. Meth. A9, 1609–1617.
- Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. Ann. Math. Statist. 35, 174–180.
- Sehr, J. (1988). On a conjecture concerning the least favorable configuration of a two-stage selection procedure. Comm. Statist.—Theor. Meth. A17 (10), 3221–3235.
- Tamhane, A.C. and Bechhofer, R.E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. Comm. Statist.—Theor. Meth. A6, 1003–1033.
- Tamhane, A.C. and Bechhofer, R.E. (1979). A two-stage minimax procedure with screening for selecting the largest normal mean (II): a new PCS lower bound and associated tables. Comm. Statist.—Theor. Meth. A8, 337–358.