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

TECHNICAL REPORT NO. 825

December 1988

**DESIGNING EXPERIMENTS FOR SELECTING
THE LARGEST NORMAL MEAN WHEN THE
VARIANCES ARE KNOWN AND UNEQUAL:
OPTIMAL SAMPLE SIZE ALLOCATION**

by

Robert E. Bechhofer

Anthony J. Hayter^{*}

Ajit C. Tamhane^{**}

Approved for public release; distribution unlimited.

Research partially supported by the U.S. Army Research Office through the Mathematical Sciences Institute of Cornell University, and the British Science and Engineering Research Council.

^{*}School of Mathematics, University of Bath, Bath BA2 7AY, ENGLAND.

^{**}Department of Statistics, Northwestern University, Evanston, Illinois 60201.

Abstract

Let Π_i denote a normal population with unknown mean μ_i and known variance σ_i^2 ($1 \leq i \leq k$). Let $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$ denote the ordered μ_i , and without loss of generality, let $\sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_k^2$. It is desired to select the population with mean $\mu_{[k]}$. We consider a natural selection procedure that takes n_i observations from Π_i ($1 \leq i \leq k$) and selects the population yielding the largest sample mean as the one associated with $\mu_{[k]}$. For fixed $N = \sum_{i=1}^k n_i$ we wish to find the optimal allocation of the n_i which is defined to be the one that maximizes the infimum of the probability of a correct selection ($P(\text{CS})$) over that part of the parameter space where $\mu_{[k]} - \mu_{[k-1]} \geq \delta^*$; here $\delta^* > 0$ is the preassigned constant employed in the indifference-zone approach to this selection problem.

Let $\lambda = \delta^* \sqrt{N/\bar{\sigma}}$ where $\bar{\sigma}^2 = \sum_{i=1}^k \sigma_i^2/k$. For $k \geq 3$ we prove that the "convenient" (easy to implement) allocation that satisfies $\sigma_1^2/n_1 = \dots = \sigma_k^2/n_k$ (which is nonoptimal for $k = 2$) is optimal if and only if $\lambda \leq \lambda_L$ or $\lambda \geq \lambda_U$ where λ_L and λ_U can be explicitly determined as functions only of the largest and the smallest relative variances, $\beta_k = \sigma_k^2/\bar{\sigma}^2$ and $\beta_1 = \sigma_1^2/\bar{\sigma}^2$, respectively. If instead of fixing N one fixes P^* , the preassigned lower bound on the $P(\text{CS})$ employed in the indifference-zone approach, then the above necessary and sufficient condition becomes $P^* \leq P_L$ or $P^* \geq P_U$.

For $k = 3$ and selected $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ -configurations numerical searches for the optimal allocation were carried out when $P_L < P^* < P_U$. These results suggest that in this case the optimal allocation for $k \geq 3$ satisfies $\sigma_1^2/n_1 \geq \dots \geq \sigma_k^2/n_k$, and that substantial savings can be achieved by using the optimal allocation instead of the "convenient" allocation, particularly when $\sigma_{\max}^2/\sigma_{\min}^2$ is large. The allocation that satisfies $\sigma_1/n_1 = \dots = \sigma_k/n_k$ (which is optimal for $k = 2$) is found to achieve most of the gains of the optimal allocation when $P_L < P^* < P_U$, and hence is suggested as a practical alternative in this case. All of

these results are derived by using a continuous approximation to the underlying exact discrete optimization problem, the approximation becoming more accurate as N gets large.

1. Introduction and Summary

Let Π_i denote a normal population with unknown mean μ_i and known variance σ_i^2 ($1 \leq i \leq k$). Without loss of generality we label the populations so that $\sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_k^2$. To avoid trivialities, we will assume that at least one of these inequalities is strict. Let Ω denote the parameter space of all parameter points $\omega = (\underline{\mu}, \underline{\sigma}^2)$ where $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_k)$ and $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$. Let $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$ denote the ordered values of the μ_i . We assume that the correct pairing of Π_i (and hence of σ_i^2) with $\mu_{[j]}$ ($1 \leq i, j \leq k$) is completely unknown. The experimenter's goal is to select the population with mean $\mu_{[k]}$ (referred to as the "best" population and assumed to be unique). If the decision procedure selects this population then a correct selection (CS) is said to have been made.

We adopt the indifference-zone approach of Bechhofer (1954) for this selection problem. In this approach, consideration is restricted to those procedures which guarantee the probability requirement:

$$\inf_{\Omega(\delta^*)} P(\text{CS}) \geq P^* \quad (1.1)$$

where

$$\Omega(\delta^*) = \{\omega \in \Omega \mid \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\} \quad (1.2)$$

is the so-called preference zone (complement of the indifference zone in Ω), and $\delta^* > 0$ and $P^* \in (1/k, 1)$ are prespecified constants.

Throughout this article we consider only the "natural" single-stage selection procedure $\mathcal{R}$, which takes independent random samples $\{X_{ij} \mid (1 \leq j \leq n_i)\}$ from the Π_i ($1 \leq i \leq k$) and selects the population that yields the largest sample mean, $\max_{1 \leq i \leq k} \bar{X}_i$, where $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i$ ($1 \leq i \leq k$). The optimal choice of the sample sizes n_i to guarantee the specified probability requirement (1.1) is the problem considered in the present article.

We define the optimal allocation $\hat{n} = (\hat{n}_1, \hat{n}_2, \dots, \hat{n}_k)$, where the $\hat{n}_i \geq 0$ are integers, as the one which when used in procedure $\mathcal{R}$, guarantees (1.1) with the smallest possible total sample size $\hat{N} = \sum_{i=1}^k \hat{n}_i$. It is easy to see that $\hat{n}$ also solves the dual of this optimization problem, namely, for a fixed positive integer N , $\hat{n}$ maximizes $\inf_{\Omega(\delta^*)} P(\text{CS})$ among all allocations $n = (n_1, n_2, \dots, n_k)$ where the $n_i \geq 0$ are integers with $\sum_{i=1}^k n_i = N$. In fact, we primarily address the dual problem in the present article.

A convenient choice of the n_i (ignoring the integer restriction on them) is one that makes $\text{var}(\bar{X}_i)$ ($1 \leq i \leq k$) equal, i.e.,

$$\frac{\sigma_1^2}{n_1} = \frac{\sigma_2^2}{n_2} = \dots = \frac{\sigma_k^2}{n_k}. \quad (1.3)$$

This allocation has the advantage that standard tables such as Table I in Bechhofer (1954) or Table A1 in Gibbons, Olkin and Sobel (1977) can be used to determine the n_i necessary to guarantee (1.1) using $\mathcal{R}$; see Remark 1. In Bechhofer (1954) it was pointed out that the allocation (1.3) is not optimal for $k = 2$, the optimal allocation (again ignoring the integer restriction on the n_i) being

$$\frac{\sigma_1}{n_1} = \frac{\sigma_2}{n_2}. \quad (1.4)$$

Dudewicz and Dalal (1975) have studied for $k = 2$ the relative efficiency of the optimal allocation (1.4) with respect to the allocation (1.3). They have shown that if σ_1^2/σ_2^2 approaches zero, then the allocation (1.3) requires twice as many observations as that required by the allocation (1.4) to guarantee the same probability requirement.

For $k \geq 3$ the optimal allocation has not been previously determined. Tong and Wetzell (1984) have given some asymptotic results but their emphasis is on

the sequential setting. Gupta and Miescke (1988) have considered this problem in the decision theoretic framework. In Bechhofer (1954, p. 24), in Hall (1959, p. 965), and in Dudewicz and Dalal (1975, p. 34) it is stated that for $k \geq 3$ the optimal allocation appears to be too complicated for practical application, while Gibbons et al. (1977, p. 68) remark that (1.3) may not be optimal for $k \geq 3$. In this article we show that for $k \geq 3$, the allocation (1.3) is in fact optimal for certain ranges of values of the parameters of the problem. More precisely, let

$$\lambda = \frac{\delta^* \sqrt{N}}{\bar{\sigma}} \quad (1.5)$$

where

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^k \sigma_i^2}{k} \quad (1.6)$$

We show that for given variances $\sigma_1^2, \dots, \sigma_k^2$, the allocation (1.3) is optimal except for $\lambda_L < \lambda < \lambda_U$ where λ_L and λ_U are two critical constants which can be determined explicitly by solving a simple equation for each. Since $\inf_{\Omega(\delta^*)} P(\text{CS})$ for the allocation (1.3) is a strictly increasing function of λ , the above limits on λ imply corresponding limits on P^* , namely, $P_L < P^* < P_U$. Furthermore, letting

$$\beta_i = \frac{\sigma_i^2}{\bar{\sigma}^2} \quad (1 \leq i \leq k), \quad (1.7)$$

we show that $\lambda_L(P_L)$ depends only on β_k while $\lambda_U(P_U)$ depends only on β_1 . Thus the determination of P_L and P_U requires only the specification of the largest and smallest relative variances (with respect to the average variance), respectively. In most practical cases of interest, P_L is quite small (.30 ~ .50), and so it is only P_U that needs to be determined. We show that

$\lambda_L < \infty$ ($P_L < 1$) always, while $\lambda_U < \infty$ ($P_U < 1$) only when

$$\beta_1 > \frac{k}{2(k-1)}, \quad (1.8)$$

i.e., when σ_1^2 is "sufficiently" large with respect to $\bar{\sigma}^2$. Otherwise λ_U doesn't exist and we may take $\lambda_U = \infty$ ($P_U = 1$). Thus the allocation (1.3) is always optimal for sufficiently small λ (P^*). If the condition (1.8) is satisfied then the allocation (1.3) is also optimal for sufficiently large λ (P^*) but not if (1.8) is violated. All of these results are derived by solving a continuous approximation to the exact discrete optimization problem that ignores the integer restrictions on the sample sizes.

The outline of the paper is as follows. Section 2 gives a mathematical formulation of the optimization problem. Section 3 gives the main theoretical results of the paper. The special case $k = 2$ is discussed in Section 3.1. The new results for $k \geq 3$ are summarized in Theorems 1-3 in Section 3.2. The proofs of all of the theorems are given in the Appendix. Section 3.3 gives a table of critical values of β_1 for selected values of k and P^* which is useful in determining whether allocation (1.3) is optimal or not. Section 4 gives the results of numerical searches for the optimal allocation that we carried out for $k = 3$ and for selected σ^2 configurations when allocation (1.3) is not optimal.

2. Problem Formulation

Let

$$\Omega_i(\delta^*) = \{\omega \in \Omega(\delta^*) \mid \mu_i = \mu_{[k]}\} \quad (1 \leq i \leq k); \quad (2.1)$$

i.e., $\Omega_i(\delta^*)$ is that part of the preference zone $\Omega(\delta^*)$ where the Π_i having the variance σ_i^2 is the best population. It was shown in Bechhofer (1954) that for procedure $\mathcal{R}$ with any choice of $\underline{n}$ and for any fixed known $\underline{\sigma}^2$,

$$\inf_{\Omega_i(\delta^*)} P(\text{CS}) = P_{\mu_i(\delta^*)}(\text{CS}) \quad (2.2)$$

where $\mu_i(\delta^*)$ is any $\underline{\mu}$ satisfying

$$\mu_1 = \dots = \mu_{i-1} = \mu_{i+1} = \dots = \mu_k = \mu_i - \delta^*,$$

i.e., $\mu_i(\delta^*)$ is the so-called slippage configuration with $\mu_i = \mu_{[k]} \quad (1 \leq i \leq k)$.

Denoting $P_{\mu_i(\delta^*)}(\text{CS})$ by $P_i \quad (1 \leq i \leq k)$ we see that

$$\inf_{\Omega(\delta^*)} P(\text{CS}) = \min_{1 \leq i \leq k} P_i. \quad (2.3)$$

If we let

$$\gamma_i = \frac{n_i}{N} \quad (1 \leq i \leq k) \quad (2.4)$$

then it is easy to show that for fixed $\underline{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)$

$$P_i = P_i(\underline{\gamma} \mid \lambda, \underline{\beta}) = \int_{-\infty}^{\infty} \prod_{\substack{j=1 \\ j \neq i}}^k \Phi \left[\sqrt{\gamma_j / \beta_j} (x \sqrt{\beta_i / \gamma_i} + \lambda) \right] \phi(x) dx \quad (1 \leq i \leq k) \quad (2.5)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cdf and pdf, respectively, and λ is given by (1.5).

For given N and $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_k)$ and specified δ^* , each P_i is a function

of the discrete valued argument $\underline{\gamma}$ since each $\gamma_j \geq 0$ is a multiple of $1/N$ with $\sum_{j=1}^k \gamma_j = 1$. For any given k, σ^2, δ^* and N , the exact integer-valued optimal allocation $\hat{\underline{n}} = (\hat{n}_1, \hat{n}_2, \dots, \hat{n}_k)$ that maximizes (2.3) where $\sum_{i=1}^k \hat{n}_i = N$ (or equivalently $\hat{\underline{\gamma}} = (\hat{n}_1/N, \hat{n}_2/N, \dots, \hat{n}_k/N)$) can be found by enumeration. However, this is only feasible for small values of N . Moreover, the integer solution has the disadvantage that a separate answer is needed for each δ^*/σ and N . We seek an approximation to this integer programming problem that does not depend on δ^*/σ and N separately but rather only on $\lambda = \delta^* \sqrt{N}/\sigma$. Toward this end we will henceforth regard the $\gamma_j \geq 0$ as continuous variables summing to unity and ignore their dependence on N . This continuous approximation obviously will become more accurate as N gets large. This same device was employed in Bechhofer (1969). Thus $P_i = P_i(\underline{\gamma}|\lambda, \underline{\beta})$ can be regarded as a continuous function of $\underline{\gamma}$ for given $\underline{\beta}$ and specified $\lambda \geq 0$. We refer to any $\underline{\gamma}$ lying in the $(k-1)$ -simplex

$$\Gamma = \{ \underline{\gamma} : \gamma_i \geq 0, \sum_{i=1}^k \gamma_i = 1 \}$$

as an allocation.

We now can state the continuous optimization problem (which is an approximation to the exact discrete optimization problem) as follows:

For given $\underline{\beta}$ and specified $\lambda \geq 0$ find

$$\max_{\underline{\gamma} \in \Gamma} \min_{1 \leq i \leq k} P_i(\underline{\gamma}|\lambda, \underline{\beta}). \quad (2.6)$$

We denote the solution to (2.6) by $\hat{\underline{\gamma}} = \hat{\underline{\gamma}}(\lambda, \underline{\beta})$ and refer to it as the optimal allocation; we denote the corresponding max-min probability by $\hat{P} = \hat{P}(\lambda, \underline{\beta})$. For fixed given $\underline{\beta}$ we will be interested in studying the behavior of $\hat{\underline{\gamma}}$ and $\hat{P}$ as functions of λ .

Remark 1: If we denote the allocation (1.3) by $\underline{\gamma}^0 = (\gamma_1^0, \gamma_2^0, \dots, \gamma_k^0)$ where

$$\frac{\gamma_1^0}{\beta_1} = \frac{\gamma_2^0}{\beta_2} = \dots = \frac{\gamma_k^0}{\beta_k} = \frac{1}{\frac{k}{\sum_{i=1}^k \beta_i}} = \frac{1}{k}, \quad (2.8)$$

then it is interesting to note from (2.5) that

$$P_1(\gamma^0 | \lambda, \beta) = P_2(\gamma^0 | \lambda, \beta) = \dots = P_k(\gamma^0 | \lambda, \beta) = P^0(\lambda, \beta) \quad (\text{say}) \quad (2.9)$$

where

$$P^0(\lambda, \beta) = \int_{-\infty}^{\infty} \phi^{k-1} \left(x + \frac{\lambda}{\sqrt{k}} \right) \phi(x) dx = \int_{-\infty}^{\infty} \phi^{k-1} \left(x + \frac{\delta^* \sqrt{N}}{\sigma \sqrt{k}} \right) \phi(x) dx. \quad (2.10)$$

If $c(k, P^*)$ denotes the solution in c to the equation

$$\int_{-\infty}^{\infty} \phi^{k-1}(x+c) \phi(x) dx = P^* \quad (2.11)$$

then the total sample size N^0 required to guarantee the probability requirement (1.1) when using the allocation (1.3) (or equivalently (2.8)) is given (ignoring the integer restriction on the sample sizes) by

$$N^0 = \left\{ \frac{c(k, P^*)}{\delta^*} \right\}^2 \sum_{i=1}^k \sigma_i^2. \quad (2.12)$$

The corresponding n_i 's (denoted by n_i^0 's) are given by

$$n_i^0 = \left\{ \frac{c(k, P^*)}{\delta^*} \right\}^2 \sigma_i^2 \quad (1 \leq i \leq k). \quad (2.13)$$

The critical constant $c(k, P^*)$ is tabulated in the references mentioned following (1.3).

3. Optimal Allocation for $k \geq 2$

3.1 Special Case $k = 2$

For $k = 2$, we see that (2.5) reduces to

$$P_1 = P_2 = \Phi[\lambda / \sqrt{(\beta_1/\gamma_1) + (\beta_2/\gamma_2)}],$$

and, as noted in (1.4), the optimal allocation for all $\lambda \geq 0$ is given by

$$\hat{\gamma}_i = \frac{\sqrt{\beta_i}}{\sqrt{\beta_1} + \sqrt{\beta_2}} = \frac{\sigma_i}{\sigma_1 + \sigma_2} \quad (i = 1, 2).$$

The case $k = 2$ has several special simplifying features, which do not extend to the cases $k \geq 3$. These features are as follows:

- i) We have $P_1 = P_2$ for all $\lambda \geq 0$. Thus for fixed $\mu_{[2]} - \mu_{[1]} = \delta^*$ (and, in fact, for any fixed $\underline{\mu}$), σ^2 and γ , ^{the} $P(\text{CS})$ is the same regardless of the association between $\mu_{[i]}$ and σ_j^2 ($i, j = 1, 2$).
- ii) Next note that $P_1 = P_2$ can be expressed as a univariate normal cdf, which for given N is maximized by minimizing $\text{var}(\bar{X}_1 - \bar{X}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$ subject to $n_1 + n_2 = N$.

For $k \geq 3$ the P_i are in general different. Furthermore, each P_i is a multivariate normal probability, which depends not only on the $\text{var}(\bar{X}_i - \bar{X}_j) = \sigma_i^2/n_i + \sigma_j^2/n_j$ ($j \neq i, 1 \leq j \leq k$), but also on the

$$\text{corr}(\bar{X}_i - \bar{X}_j, \bar{X}_i - \bar{X}_{j'}) = \frac{\sigma_i^2/n_i}{\{(\sigma_i^2/n_i + \sigma_j^2/n_j)(\sigma_i^2/n_i + \sigma_{j'}^2/n_{j'})\}^{1/2}} \quad (j \neq j' \neq i, 1 \leq j, j' \leq k).$$

3.2 General Case $k \geq 3$

Our objective in this section is to determine the range of values of λ for which the allocation γ^0 given by (2.8) is optimal when $k \geq 3$. The principal results of this section are summarized in the following theorems.

Theorem 1: Define

$$A(\lambda) = \int_{-\infty}^{\infty} x \phi^{k-2}(x) \phi(x) \phi(x - \frac{\lambda}{\sqrt{k}}) dx, \quad (3.1)$$

$$B(\lambda) = \int_{-\infty}^{\infty} (x - \frac{\lambda}{\sqrt{k}}) \phi^{k-2}(x) \phi(x) \phi(x - \frac{\lambda}{\sqrt{k}}) dx, \quad (3.2)$$

and

$$G(\lambda) = \left(\frac{k}{k-1}\right) \left[\frac{A(\lambda)}{A(\lambda)-B(\lambda)}\right]. \quad (3.3)$$

Then the allocation (2.8) is locally optimal iff

$$G(\lambda) \geq \beta_k \quad \text{or} \quad G(\lambda) \leq \beta_1. \quad \square \quad (3.4)$$

Corollary: For $\lambda = G^{-1}(1)$ (that G^{-1} exists and is unique follows from Theorem 2 below) the allocation (2.8) is locally optimal iff $\sigma_1^2 = \dots = \sigma_k^2$. $\square$

Remark 2: Note that only local optimality is claimed for γ^0 . However, extensive numerical studies carried out for $k = 3$ strongly indicate that γ^0 is in fact globally optimal whenever $G(\lambda) \geq \beta_k$ or $G(\lambda) \leq \beta_1$. Also for $\lambda = 0$ it can be shown analytically that γ^0 is globally optimal; see Theorem 4.1 in Bechhofer and Tamhane (1982).

Theorem 2: For $\lambda > 0$ the function $G(\lambda)$ is continuous and strictly decreasing in λ with $\lim_{\lambda \rightarrow 0} G(\lambda) = \infty$ and $\lim_{\lambda \rightarrow \infty} G(\lambda) = k/2(k-1)$. Hence the condition (3.4) is equivalent to the condition

$$\lambda \leq \lambda_L \quad \text{or} \quad \lambda \geq \lambda_U. \quad (3.5)$$

Here λ_L is the unique finite solution in λ of the equation

$$G(\lambda) = \beta_k, \quad (3.6)$$

and if (1.8) is satisfied then λ_U is the unique finite solution in λ of the

equation

$$G(\lambda) = \beta_1. \quad (3.7)$$

If (1.8) is not satisfied then (3.7) does not have a solution and in that case

we define $\lambda_U = \infty$. □

Corollary: The allocation (2.8) is locally optimal iff the specified P^* is $\leq P_L$ or $\geq P_U$ where

$$P_L = \int_{-\infty}^{\infty} \phi^{k-1} \left(x + \frac{\lambda_L}{\sqrt{k}} \right) \phi(x) dx \quad (3.8)$$

and

$$P_U = \begin{cases} \int_{-\infty}^{\infty} \phi^{k-1} \left(x + \frac{\lambda_U}{\sqrt{k}} \right) \phi(x) dx & \text{if } \beta_1 > \frac{k}{2(k-1)} \\ 1 & \text{if } \beta_1 \leq \frac{k}{2(k-1)}. \end{cases} \quad (3.9)$$

□

We now give (in Theorem 3) an alternative representation for equations (3.6) and (3.7) which is convenient for computing. This representation involves multivariate normal cdf's for which we use the following notation: Let $X_1, X_2, \dots, X_p$ have a joint p -variate normal distribution with zero means, unit variances, and common correlation $\rho = \text{corr}(X_i, X_j)$ for $i \neq j$, ($1 \leq i, j \leq p$). Then we denote the equicoordinate multivariate normal probability

$$P(X_1 \leq x, X_2 \leq x, \dots, X_p \leq x)$$

by $\Phi_p(x|\rho)$. Of course, for $p = 1$ this probability is simply the univariate normal cdf denoted by $\Phi(x)$. For $p = 0$ we define this probability to be unity.

Theorem 3: Set $\tau = \lambda/\sqrt{6k}$. Then $\lambda_L = \tau_L \sqrt{6k}$ where τ_L is the unique solution in τ of the equation

$$\frac{\tau \phi_{k-2}(\tau|1/3)}{\phi(\tau) \phi_{k-3}(\tau/\sqrt{2}|1/4)} = \frac{k(k-2)}{3[2(k-1)\beta_k - k]} \quad (3.10)$$

Similarly if condition (1.8) is satisfied, then $\lambda_U = \tau_U \sqrt{6k}$ where τ_U is the unique solution in τ of the equation

$$\frac{\tau \phi_{k-2}(\tau|1/3)}{\phi(\tau) \phi_{k-3}(\tau/\sqrt{2}|1/4)} = \frac{k(k-2)}{3[2(k-1)\beta_1 - k]} \quad (3.11)$$

□

Remark 3: For $\tau > 0$ the l.h.s of (3.10) and (3.11) is positive. It is apparent therefore that these equations will not have solutions if their respective r.h.s. are negative. The r.h.s. of (3.10) is always positive because $\beta_k > k/2(k-1)$ while the r.h.s. of (3.11) is positive iff $\beta_1 > k/2(k-1)$ which is the condition (1.8).

Remark 4: For $k = 3$ the l.h.s. of (3.8) and (3.9) reduces to $\tau \phi(\tau)/\phi(\tau)$ which is very simple to evaluate.

3.3 Tables of Critical Values of β_1 for $k \geq 3$

To assist the experimenter in designing his experiment we have prepared a table of values of the lower bound on β_1 , say β_1^* , and the values of the associated lower bound on P_U , say P_U^* , such that for $P^* \geq P_U^*$, the allocation γ^0 given by (1.3) is (locally) optimal if $\beta_1 \geq \beta_1^*$. For selected values of (k, P_U^*) , we computed the values of β_1^* as follows: First, we calculated $\lambda_U^* = \sqrt{k} c(k, P_U^*)$ (see equation (3.9)) where $c(k, P_U^*)$, which is the solution to (2.11) when $P^* = P_U^*$, was obtained from Table I in Bechhofer (1954). Next β_1^* was calculated by setting $\tau = \tau_U^* = \lambda_U^*/\sqrt{6k}$ in (3.11) and solving for β_1 (or equivalently by setting $\lambda = \lambda_U^*$ in (3.7) and solving for β_1). Table 1 gives these values of β_1^* for $k = 3(1)8$ and $P_U^* = 0.80, 0.90, 0.95$ and 0.99 . We also have added a row for $P_U^* = 1$ in which

case $\beta_1^* = k/2(k-1)$.

Table 1
Critical Values β_1^*

P_U^*	k					
	3	4	5	6	7	8
0.80	0.907	0.854	0.826	0.809	0.798	0.791
0.90	0.838	0.810	0.748	0.731	0.720	0.712
0.95	0.806	0.741	0.708	0.690	0.678	0.670
0.99	0.775	0.670	0.664	0.643	0.629	0.620
1.00	0.750	0.667	0.625	0.600	0.583	0.571

As an illustration of the use of this table, suppose that $k = 3$ and $P^* = 0.95$. If $\beta_1 \geq \beta_1^* = 0.806$ then $\gamma_{\sim}^0$ is the optimal allocation, and the corresponding sample sizes required can be found from (2.13) once δ^* is specified. If $\beta_1 < \beta_1^*$ then the optimal allocation is not given by $\gamma_{\sim}^0$.

An analogous table could be given for β_k^* for selected values of P_L^* such that for $P^* \leq P_L^*$, the allocation $\gamma_{\sim}^0$ is locally optimal if $\beta_k \leq \beta_k^*$. However, such a table is likely to be of less practical value since large values of P^* are more commonly used. Therefore we have not given a table of (β_k^*, P_L^*) values.

4. Numerical Results for $k = 3$

In Section 3 we derived a necessary and sufficient condition for $\underline{\gamma}^0$ to be locally optimal for any $k \geq 3$. In the present section we investigate the nature of the optimal allocation when that condition does not hold, i.e., when $P_L < P^* < P_U$. We also investigate the amount of the associated saving in the total sample size in comparison to that required by the allocation $\underline{\gamma}^0$ to guarantee the same specified probability requirement (1.1).

An analytical characterization of the optimal allocation appears very difficult when $P_L < P^* < P_U$ holds. The difficulty is compounded by the fact that the optimal allocation is not fixed in this case but changes in a continuous manner with P^* (or equivalently with λ). Therefore we decided to investigate numerically the behavior of the optimal allocation as a function of P^* by performing a search in the allocation space Γ . Of course, this is a very formidable computational task for large k , so we confined our attention to $k = 3$, in which case the search is just in two dimensions.

For $k = 3$, we present the results for a total of six $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ configurations. The first three configurations have $\sigma_3^2/\sigma_1^2 = 3$, while the second three have $\sigma_3^2/\sigma_1^2 = 10$. These configurations and the associated $\underline{\beta}$ vectors are listed in Table 2. Note that the optimal allocation (and, as will be seen below, the relative saving in the total sample size) depends only on the relative magnitudes of the σ_i^2 , not their absolute magnitudes. For each configuration, we have $\beta_1 < k/2(k-1) = 0.75$ and hence $P_U = 1$. The P_L -values associated with each configuration $\underline{\beta}$ (recall that P_L depends on $\underline{\beta}$ only through β_k) are also listed in Table 2. For each configuration the optimal allocation was determined numerically for $P^* = 0.80, 0.90, 0.95$ and 0.99 ; note that this practical range of P^* -values is well in excess of P_L for each configuration.

The numerical search for the optimal allocation was carried out as follows:

Let λ^0 be the λ -value required using the allocation $\underline{\gamma}^0$ to guarantee the probability requirement (1.1) for specified P^* and for any $\delta^* > 0$; from (2.10) we see that $\lambda^0 = \sqrt{k} c(k, P^*)$. (Note that the λ -value required to guarantee (1.1) using any allocation $\underline{\gamma}$ depends only on $(\beta_1, \dots, \beta_k)$ and P^* but the total sample size N depends also on $\delta^*/\bar{\sigma}$.) Let $\hat{\lambda}$ be the corresponding λ -value required using the associated optimal allocation $\hat{\underline{\gamma}}$ where, clearly, $\hat{\lambda} \leq \lambda^0$. Starting with λ^0 we decreased λ in steps of 0.001, determining the optimal allocation $\hat{\underline{\gamma}}$ and the associated max-min probability $\hat{P}$ for each λ (note that $\hat{P}$ decreases with λ), until the smallest possible λ for which $\hat{P} \geq P^*$ is attained. This is the desired value of $\hat{\lambda}$, which is tabulated together with $\hat{\underline{\gamma}}$ in Table 2. A mesh size of at most 1/60 was used for each γ_1 in the search over the allocation space Γ .

The percentage relative saving ($\hat{RS}$) in the total sample size resulting from the use of the optimal allocation $\hat{\underline{\gamma}}$ instead of the allocation $\underline{\gamma}^0$ to guarantee the same probability requirement (1.1) is given (ignoring the integer restrictions on the sample sizes) by

$$\begin{aligned}\hat{RS} &= \left(\frac{N^0 - \hat{N}}{N^0} \right) \times 100 \\ &= \left\{ \frac{(\lambda^0)^2 - (\hat{\lambda})^2}{(\lambda^0)^2} \right\} \times 100.\end{aligned}\tag{4.1}$$

The values of $\hat{RS}$ are also listed in Table 2.

From Table 2 we first note that, as one would expect, the relative savings are substantially higher for the configurations with $\sigma_3^2/\sigma_1^2 = 10$ compared to those for the configurations with $\sigma_3^2/\sigma_1^2 = 3$. Thus the saving in the total sample size from the use of the optimal allocation $\hat{\underline{\gamma}}$ (in comparison to that required when using the allocation $\underline{\gamma}^0$) appears to increase with $\sigma_{\max}^2/\sigma_{\min}^2$. For each configuration the relative saving is highest for $P^* = 0.80$ and decreases as P^*

increases. For the configuration $\sigma^2 = (1,1,10)$, the relative saving is nearly 23% for $P^* = 0.80$. This indicates that there is much to be gained by using the optimal allocation $\hat{\gamma}$ instead of the "convenient" allocation γ^0 , particularly when $\sigma_{\max}^2/\sigma_{\min}^2$ is large and P^* is about midway between P_L and P_U . We should, however, stress that although the relative savings are small for large P^* , the absolute savings, $N^0 - \hat{N} = \{(\lambda^0)^2 - (\hat{\lambda})^2\}(\bar{\sigma}/\delta^*)^2$, can be quite large, more so when $\delta^*/\bar{\sigma}$ is small.

The virtues of the optimal allocation $\hat{\gamma}$ relative to the "convenient" allocation γ^0 have been well illustrated in this numerical study. However, in practice the numerical search for the optimal allocation can be prohibitively expensive and possibly even infeasible for large k . Therefore it would be desirable to have a simple heuristic rule that would improve upon γ^0 and possibly serve as a reasonable approximation to the optimal allocation $\hat{\gamma}$. With this in mind we now carefully examine the $\hat{\gamma}$ -vectors listed in Table 2.

In several cases we see that $\hat{\gamma}$ does not change as we vary P^* . We do not have a simple explanation for this behavior of the optimal allocation. We also observe quite unmistakably that $\hat{\gamma}_i < \gamma_i^0$ if $\beta_i > 1$, $\hat{\gamma}_i > \gamma_i^0$ if $\beta_i < 1$ and $\hat{\gamma}_i = \gamma_i^0$ if $\beta_i = 1$. (The last fact is observed for the configuration $\sigma^2 = (1,2,3)$ for which we have $\beta_2 = 1$ and $\hat{\gamma}_2 = \gamma_2^0 = 0.333$ in all the cases studied.) In other words, it appears that $\hat{\gamma}$ allocates a smaller (larger) proportion of observations (than that allocated by γ^0) to any population with larger (smaller) variance relative to $\bar{\sigma}^2$ which results in the inequality

$$\frac{\hat{\gamma}_1}{\beta_1} \geq \frac{\hat{\gamma}_2}{\beta_2} \geq \dots \geq \frac{\hat{\gamma}_k}{\beta_k}. \quad (4.2)$$

Now we know that $\beta_i \geq \sqrt{\beta_i}$ depending on whether $\beta_i \geq 1$. Therefore the

allocation $\tilde{\gamma}$ with

$$\frac{\tilde{\gamma}_2}{\sqrt{B_1}} = \frac{\tilde{\gamma}_2}{\sqrt{B_2}} = \dots = \frac{\tilde{\gamma}_k}{\sqrt{B_k}}, \quad (4.3)$$

which chooses the n_i 's in proportion to the σ_i 's, is an allocation that satisfies (4.2). Recall that this allocation is uniformly optimal for $k = 2$. It would be of interest to find out how close this allocation is to the optimum for $k \geq 3$ when $P_L < P^* < P_U$ holds and hence when γ^0 is known not to be optimal. Toward this end we determined the smallest λ -value (denoted by $\tilde{\lambda}$) for the allocation $\tilde{\gamma}$ such that the associated probability is $\geq P^*$ for the σ^2 -configurations and P^* -values listed in Table 2. We also calculated the percentage relative saving (RS) associated with $\tilde{\gamma}$ relative to γ^0 as in (4.1):

$$\begin{aligned} RS &= \left(\frac{N^0 - \tilde{N}}{N^0} \right) \times 100 \\ &= \left\{ \frac{(\lambda^0)^2 - (\tilde{\lambda})^2}{(\lambda^0)^2} \right\} \times 100. \end{aligned} \quad (4.4)$$

The results are given in Table 3.

Inspection of Tables 2 and 3 reveals that in many cases, the $\tilde{\gamma}$ allocation achieves relative savings nearly equal to those achieved by the optimal allocation $\hat{\gamma}$. The $\tilde{\gamma}$ allocation improves upon the γ^0 allocation in all of the cases studied except two (for $\sigma^2 = (1,2,3)$ and $(1,1,3)$ when $P^* = 0.99$), and in those two cases the excess sample size required by $\tilde{\gamma}$ compared to that required by γ^0 is not large, at least in relative terms.

Recognizing the computational difficulties involved in determining the optimal allocation $\hat{\gamma}$ when $P_L < P^* < P_U$ holds, we recommend the $\tilde{\gamma}$ allocation in this case with little reservation.

It should be noted that to determine $\tilde{\lambda}$ required by $\tilde{\gamma}$ for given β and specified P^* , it is necessary to search over λ by calculating $P_i(\tilde{\gamma}|\lambda, \beta)$ for $i = 1, 2, \dots, k$ and finding the smallest $\lambda = \tilde{\lambda}$ for which $\min_{1 \leq i \leq k} P_i(\tilde{\gamma}|\lambda, \beta) \geq P^*$ holds. However, since this is only a one-dimensional search over λ (the allocation $\tilde{\gamma}$ is fixed) and since $\min_{1 \leq i \leq k} P_i$ is an increasing function of λ , it is not a difficult task, although it does involve evaluating the k univariate integrals, $P_i(\tilde{\gamma}|\lambda, \beta)$ for $i = 1, 2, \dots, k$, for each λ . No fixed rule can be given to tell us which $P_i(\tilde{\gamma}|\lambda, \beta)$ ($1 \leq i \leq k$) is the minimum for given λ and β . Finally we note that a good choice for the starting value of λ is provided by λ^0 (which is readily obtainable from available tables as noted before). If it turns out that $\tilde{\lambda} > \lambda^0$ (as it did in the two cases in Table 3) then, of course, the allocation $\tilde{\gamma}$ should not be used. One may either use γ^0 in that case or try some alternative allocation γ satisfying (4.2).

Table 2

Optimum Allocation $\hat{Y}$ and Associated Relative Saving $\hat{R}\hat{S}$

$\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$	(1,3,3)	(1,2,3)	(1,1,3)	(1,10,10)	(1,4,10)	(1,1,10)
$\beta = (\beta_1, \beta_2, \beta_3)$	(.429, 1.286, 1.286)	(.500, 1.000, 1.500)	(.600, .600, 1.800)	(.143, 1.430, 1.430)	(.200, .800, 2.000)	(.250, .250, 2.500)
$\gamma^0 = (\gamma_1^0, \gamma_2^0, \gamma_3^0)$	(.143, .429, .429)	(.167, .333, .500)	(.200, .200, .600)	(.048, .476, .476)	(.067, .267, .667)	(.083, .083, .833)
P^*	λ^0	P_L	0.55	0.41	0.44	0.51
.80	2.862	$\hat{Y}$	(.233, .333, .433)	(.267, .267, .467)	(.167, .300, .533)	(.183, .183, .633)
		$\hat{\lambda}$	2.818	2.795	2.667	2.514
		$\hat{R}\hat{S}$	3.05%	4.63%	13.16%	22.84%
.90	3.863	$\hat{Y}$	(.200, .333, .467)	(.233, .233, .533)	(.150, .300, .550)	(.167, .167, .667)
		$\hat{\lambda}$	3.831	3.810	3.667	3.492
		$\hat{R}\hat{S}$	1.65%	2.73%	9.89%	18.29%
.95	4.694	$\hat{Y}$	(.200, .333, .467)	(.233, .233, .533)	(.133, .300, .567)	(.167, .167, .667)
		$\hat{\lambda}$	4.674	4.654	4.504	4.300
		$\hat{R}\hat{S}$	0.85%	1.70%	7.93%	16.08%
.99	6.265	$\hat{Y}$	(.175, .413, .413)	(.233, .233, .533)	(.117, .300, .583)	(.167, .167, .667)
		$\hat{\lambda}$	6.227	6.231	6.070	5.806
		$\hat{R}\hat{S}$	1.21%	1.08%	6.13%	14.12%

Table 3

Allocation $\tilde{Y}$ and Associated Relative Saving $\tilde{RS}$

$\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$	(1,3,3)	(1,2,3)	(1,1,3)	(1,10,10)	(1,4,10)	(1,1,10)
$\sqrt{\tilde{B}} = (\sqrt{\tilde{B}_1}, \sqrt{\tilde{B}_2}, \sqrt{\tilde{B}_3})$	(.655, 1.134, 1.134)	(.707, 1.000, 1.225)	(.755, .775, 1.342)	(.378, 1.195, 1.195)	(.447, .894, 1.414)	(.500, .500, 1.581)
$\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3)$	(.224, .388, .388)	(.241, .341, .418)	(.268, .268, .464)	(.137, .432, .432)	(.162, .325, .513)	(.194, .194, .613)
$P^* = 0.80$	$\tilde{\lambda}$	2.827	2.801	2.657	2.673	2.516
	$\tilde{RS}$	4.08%	4.22%	13.81%	12.77%	22.72%
$P^* = 0.90$	$\tilde{\lambda}$	3.812	3.853	3.834	3.653	3.502
	$\tilde{RS}$	2.62%	0.52%	1.50%	10.58%	17.82%
$P^* = 0.95$	$\tilde{\lambda}$	4.659	4.694	4.663	4.490	4.315
	$\tilde{RS}$	1.49%	0%	1.32%	8.50%	15.50%
$P^* = 0.99$	$\tilde{\lambda}$	6.259	6.315	6.296	6.110	5.850
	$\tilde{RS}$	0.19%	-1.60%	-0.99%	4.89%	12.81%

5. Concluding Remarks

In this paper we have shown that the convenient allocation $\underline{\gamma}^0$ given by (2.8) is optimal for $k \geq 3$ if and only if $P^* \leq P_L$ or $P^* \geq P_U$ where P_L and P_U can be explicitly determined given β_k and β_1 , respectively. Thus the determination of P_L and P_U requires the knowledge of only the largest and smallest relative variances. However, the determination of the optimal allocation $\hat{\underline{\gamma}}$ (whether or not it equals $\underline{\gamma}^0$) requires the knowledge of all the β_i 's, and the determination of the associated sample sizes $\hat{n}_i$ needed to guarantee (1.1) for specified $\{\delta^*, P^*\}$ requires the knowledge of all of the σ_i^2 's. The optimal allocation is difficult to determine when $P_L < P^* < P_U$ holds. In that case use of the allocation $\tilde{\underline{\gamma}}$ given by (4.3) (or some other allocation satisfying (4.2)) is suggested.

As a final remark, we raise the question of the appropriateness of the procedure R which bases its decision on the sample means $\bar{X}_i$, which have, in general, unequal variances σ_i^2/n_i . Suppose that the two largest sample means differ by a small amount but the largest sample mean has a much larger variance than the second largest sample mean. (This is possible even when the σ_i^2 's are equal but the n_i 's are not.) Intuition suggests that in this case we should select the population yielding the second largest sample mean as the "best." This is because the second largest sample mean is a much more reliable estimator of its population mean (which is thus more likely to be large and possibly the largest) than the largest sample mean is of its population mean (which is thus less likely to be the largest). Recently Berger and Deely (1988) have given a Bayesian solution to this problem which involves shrinking the sample means toward a central average, the extent of shrinkage being greater for extreme (large or small) sample means having larger variances.

We are indebted to Dr. Prakash Awate who contributed in the early stages of this work while he was a graduate student at Cornell University.

References

- 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. (1969), "Optimal allocation of observations when comparing several treatments with a control," Multivariate Analysis, II (Ed. P. R. Krishnaiah), New York: Academic Press, 465-473.
- Bechhofer, R. E. and Tamhane, A. C. (1982), "Optimal sample size allocation for selecting the best of several normal populations with known, unequal variances," Technical Report No. 567, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y. 14853.
- Berger, J. O. and Deely, J. (1988), "A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology," J. Amer. Statist. Assoc., 83, 364-373.
- Dudewicz, E. J. and Dalal, S. R. (1975), "Allocation of observations in ranking and selection with unequal variances," Sankhyā, Ser. B, 37, 28-78.
- Gibbons, J. D., Olkin, I. and Sobel, M. (1977), Selecting and Ordering Populations, New York: John Wiley.
- Gupta, S. S. and Miescke, K. J. (1988), "On the problem of finding the largest mean under heteroscedasticity," in Statistical Decision Theory and Related Topics, IV (Eds. S. S. Gupta and J. O. Berger), vol. 2, 37-49.
- Hall, W. J. (1959), "The most economical character of some Bechhofer and Sobel decision rules," Ann. Math. Statist., 30, 964-969.
- Tong, Y. L. and Wetzell, D. E. (1984), "Allocation of observations for selecting the best normal population," Design of Experiments: Ranking and Selection (eds. T. J. Santner and A. C. Tamhane), New York: Marcel-Dekker, 213-224.