

AD-A185 614

AN ELIMINATION TYPE TWO-STAGE PROCEDURE
FOR SELECTING THE POPULATION WITH
THE LARGEST MEAN FROM K LOGISTIC POPULATIONS

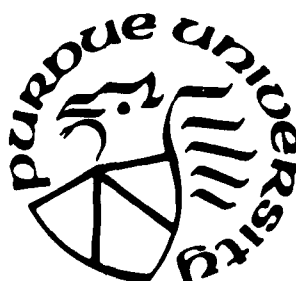
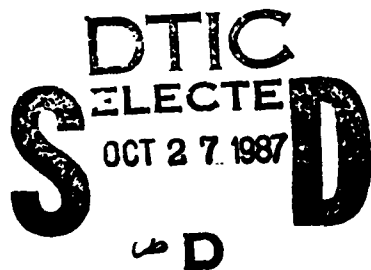
by

Shanti S. Gupta
Purdue University

ShangHyun Han
Purdue University

Technical Report #87-39

PURDUE UNIVERSITY



DEPARTMENT OF STATISTICS

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

1

AN ELIMINATION TYPE TWO-STAGE PROCEDURE
FOR SELECTING THE POPULATION WITH
THE LARGEST MEAN FROM K LOGISTIC POPULATIONS

by

Shanti S. Gupta
Purdue University

ShangHyun Han
Purdue University

Technical Report #87-39

Department of Statistics
Purdue University

August 1987

DTIC
ELECTE
OCT 27 1987
S D
RD

*This research was supported by the Office of Naval Research Contract
N00014-84-C-0167 at Purdue University. Reproduction in whole or in
part is permitted for any purpose of the United States Government.

DISTRIBUTION STATEMENT 1
Approved for public release
Distribution Unlimited

**AN ELIMINATION TYPE TWO-STAGE PROCEDURE
FOR SELECTING THE POPULATION WITH
THE LARGEST MEAN FROM k LOGISTIC POPULATIONS ¹**

by

Shanti S. Gupta and SangHyun Han

Purdue University

Abstract

A formula for an approximation to the distribution of the sample means of a logistic population is derived by using the Edgeworth series expansions. Using this approximation, ^{the authors} we consider an elimination type two-stage procedure based on the sample means for selecting the population with the largest mean from k logistic populations when the common variance is known. A table of the constants needed to implement this procedure is provided and the efficiency of this procedure relative to the single-stage procedure is investigated.

random variables; tables (data). Keywords:
KEY WORDS: Two-Stage Procedure, Selection Procedure, Largest Mean, Subset Selection, Logistic Populations.

1 Introduction

For the problem of selecting the population having the largest mean from normal populations with a common known variance σ^2 , Cohen (1959), Alam (1970) and Tamhane and Bechhofer (1977, 1979) have studied two-stage elimination type procedures, in which they used Gupta's (1956, 1965) subset selection procedure in the first stage to screen out non-contending populations and Bechhofer's (1954) indifference zone approach to select the best from among the populations in the second stage.

¹This research was partially supported by the Office of Naval Research Contract N00014-84-C-0167 and NSF Grant DMS-8606964 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.



| |
|--------------------------|
| ↓ |
| <input type="checkbox"/> |
| <input type="checkbox"/> |
| y Copies |
| Date received for |
| SECRET |
| A-1 |

Tamhane and Bechhofer (1977, 1979) studied in depth a two-stage elimination type procedure (\mathcal{P}'_2) for selecting the largest normal mean when the common variance is known. In order to determine a set of constants necessary to implement \mathcal{P}'_2 , they proposed a criterion of minimizing the maximum over the entire parameter space of the expected total sample size required by \mathcal{P}'_2 subject to the procedure's guaranteeing a specified probability of a correct selection. As a consequence, \mathcal{P}'_2 based on this unrestricted minimax design criterion possesses the highly desirable property that the expected total sample size required by \mathcal{P}'_2 is always less than or equal to the total sample size required by the best competing single-stage procedure of Bechhofer (1954), regardless of the true configuration of the population means.

The logistic distribution has been widely used by Berkson (1944,1951,1957) as a model for analyzing experiments involving quantal response. Pearl and Reed (1920) used this in studies connected with population growth. Plackett (1958,1959) has considered the use of this distribution with life test data. Gupta (1962) has studied this distribution as a model in life testing problems.

The importance of the logistic distribution in the modeling of stochastic phenomena has resulted in numerous other studies involving probabilistic and statistical aspects of the distribution. For example, Gumbel (1944), Gumbel and Keeney (1950) and Talacko (1956) show that it arises as a limiting distribution in various situations; Birnbaum and Dudman (1963), Gupta and Shah (1965) study its order statistics. Many other authors, for example, Antle, Klimko and Harkness (1970), Gupta and Gnanadesikan (1966) and Tarter and Clark (1965), investigate inference questions about its parameters.

In this paper we consider an elimination type two-stage procedure for selecting the logistic population with the largest population mean when the populations have a common known variance. Using an approximation to the distribution of the sample means from a logistic population, we propose a two-stage elimination type procedure \mathcal{P}_2 and a non-linear optimization problem by using a minimax criterion to find a set of constants needed to implement \mathcal{P}_2 . We derive lower bounds of the probability of a

correct selection and the infimum over the preference zone of the lower bounds. We determine the supremum of the expected total sample size needed for \mathcal{P}_2 over the whole parameter space. We provide tables of constants to implement \mathcal{P}_2 and of the efficiency of \mathcal{P}_2 relative to the corresponding single-stage procedure \mathcal{P}_1 for the two special cases of the equally spaced and slippage configurations.

2 Distribution of logistic sample means

Because of the similarity between the logistic and the normal distributions, the sample mean and variance, the moment estimators of the logistic population parameters, are effective tools for statistical decisions involving the logistic distribution. Antle, Klimko and Harkness (1970) give a function of the sample mean as a confidence interval estimate of the population mean when the population variance is known. Schafer and Sheffield (1973) show that in terms of the mean squared error the moment estimators of the logistic population parameters are as good as their maximum likelihood estimators. The fact that the distribution of a sample mean has monotone likelihood ratio (MLR) with respect to the population mean when the variance is known is used by Goel (1975) to obtain a uniformly most accurate confidence interval for the population mean and a uniformly most powerful test for one-sided hypotheses involving the population mean. The sampling distribution of the mean is a primary requirement for these statistical purposes. The papers by Antle, Klimko and Harkness (1970) and Tarter and Clark (1965) used a Monte Carlo method for this distribution.

Goel (1975) obtained an expression for the distribution function of the sum of independent and identically distributed (*iid*) logistic variates by using the Laplace transform inverse method for convolutions of Pólya type functions, a technique developed by Schoenberg (1953) and Hirschman and Widder (1955). He provides a table of the cumulative distribution function (cdf) of the sum of *iid* logistic variates for the sample size $n = 2(1)12$, $x = 0(0.01)3.99$ and $n = 13(1)15$, $x = 1.20(0.01)3.99$. George

and Mudholkar (1983) obtained an expression for the distribution of a convolution of the *iid* logistic variables by directly inverting the characteristic function. However, since both formulas of Goel (1975) and George and Mudholkar (1983) contain a term $(1 - e^x)^{-k}$, $k = 1, \dots, n$, a problem of precision of the computation at the values of x near zero arises when n is large. George and Mudholkar (1983) also show that a standardized Student's t distribution provides a very good approximation for the distribution of a convolution of the *iid* logistic random variables.

In this section, we consider an approximation problem for the distribution of a standardized mean of samples from a logistic population by using Edgeworth series expansions. It can be shown that this is a far better approximation than the Student's t distribution as suggested in Goerge and Mudholkar (1983) and hence this approximation will be used henceforth.

2.1 Logistic distribution

A random variable X has the logistic distribution with mean μ and variance σ^2 , sometimes denoted by $L(\mu, \sigma^2)$, if the probability density function (pdf) of X is given by

$$f(x) = (g/\sigma)[\exp\{-g(x - \mu)/\sigma\}][1 + \exp\{-g(x - \mu)/\sigma\}]^{-2} \quad (1)$$

and the cdf of X is defined by

$$F(x) = [1 + \exp\{-g(x - \mu)/\sigma\}]^{-1}, \quad (2)$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ and $g = \pi/\sqrt{3}$. This distribution is symmetrical about the mean μ .

Letting $Y = (X - \mu)g/\sigma$, the random variable Y has the logistic distribution with mean zero and variance $\pi^2/3$. The pdf and cdf of the random variable Y are given by

$$f(y) = [\exp\{-y\}][1 + \exp\{-y\}]^{-2} \quad (3)$$

and

$$F(y) = [1 + \exp\{-y\}]^{-1} \quad (4)$$

respectively, where $-\infty < y < \infty$. (3) may be written in terms of $F(y)$ as

$$f(y) = F(y)(1 - F(y)). \quad (5)$$

The moment generating function (*mgf*) of Y is given by

$$\begin{aligned} M_Y(t) &= \Gamma(1+t)\Gamma(1-t) \\ &= \pi t / \sin \pi t, \quad |t| < 1. \end{aligned} \quad (6)$$

We can also express (6) as

$$M_Y(t) = \sum_{j=0}^{\infty} (-1)^{j-1} [2(2^{2j-1} - 1)/(2j)!] B_{2j} (\pi t)^{2j}, \quad (7)$$

where B_ν 's are Bernoulli numbers defined as

$$x/(\exp(x) - 1) = \sum_{\nu=0}^{\infty} B_\nu x^\nu / (\nu!). \quad (8)$$

The ν^{th} central moments of Y , denoted by $\mu_\nu(y)$, can be obtained as

$$\begin{aligned} \mu_\nu(y) &= E(Y^\nu) \\ &= \begin{cases} (-1)^{\nu/2-1} [2(2^{\nu-1} - 1)] B_\nu \pi^\nu; & \text{if } \nu = 2j, j = 1, 2, \dots, \\ 0; & \text{otherwise,} \end{cases} \end{aligned}$$

by using (7). Then the ν^{th} central moments of X , denoted by $\mu_\nu(x)$, are given by

$$\begin{aligned} \mu_\nu(x) &= E(X - \mu)^\nu \\ &= (\sigma/g)^\nu E(Y^\nu) \\ &= \begin{cases} (-1)^{\nu/2-1} (\sqrt{3}\sigma)^\nu [2(2^{\nu-1} - 1)] B_\nu; & \text{if } \nu = 2j, j = 1, 2, \dots, \\ 0; & \text{otherwise.} \end{cases} \end{aligned}$$

The ν^{th} cumulant of X , denoted by $K_\nu(x)$, $\nu = 1, 2, \dots$ is defined by

$$\log \varphi_X(t) = \sum_{\nu=1}^{\infty} K_\nu(x) (it)^\nu / (\nu!),$$

where $\varphi_X(t)$ is the characteristic function of the random variable X and the ν^{th} relative cumulant of X , $\lambda_\nu(x)$, is defined by

$$\lambda_\nu(x) = K_\nu(x)(K_2(x))^{-\nu/2}.$$

Using the moments of X and the definition, the first few of the ν^{th} relative cumulants of X are given by

$$\begin{aligned} \lambda_1(x) &= \mu/\sigma, \\ \lambda_2(x) &= 1, \\ \lambda_4(x) &= 6/5, \\ \lambda_6(x) &= 48/7, \\ \lambda_8(x) &= 432/5, \\ \lambda_{10}(x) &= 145152/77, \\ &\vdots \\ \lambda_{2j+1}(x) &= 0, \quad j = 1, 2, \dots \end{aligned} \tag{9}$$

2.2 Edgeworth series expansions for the distribution of the mean of samples from a logistic population

Let X_1, X_2, \dots, X_n be a random sample of size n from a logistic population $L(\mu, \sigma^2)$ with mean μ and variance σ^2 whose cdf and pdf are given in (1) and (2) respectively. Define a standardized mean of samples of size n from $L(\mu, \sigma^2)$, Z say, as

$$\begin{aligned} Z &= \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^n (X_i - \mu) \\ &= \frac{\sqrt{n}}{\sigma} (\bar{X} - \mu), \end{aligned} \tag{10}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ is the sample mean.

Let $f_n(z)$ and $F_n(z)$ denote the pdf and cdf of the standardized mean of samples of size n from $L(\mu, \sigma^2)$. Then the Edgeworth series expansions of the $f_n(z)$ and $F_n(z)$

are given symbolically as

$$f_n(z) = \phi(z) + \phi(z) \sum_{j=1}^{\nu} p_j(z) n^{-j/2} + O(n^{-(\nu+1)/2})$$

and

$$F_n(z) = \Phi(z) - \phi(z) \sum_{j=1}^{\nu} P_j(z) n^{-j/2} + O(n^{-(\nu+1)/2})$$

respectively, where $\phi(z)$ and $\Phi(z)$ are the standard normal pdf and cdf respectively and $p_j(z)$ and $P_j(z)$ are polynomials in z , which are obtained up to $\nu = 10$ in Draper and Tierney (1973).

Using $p_j(z)$ and $P_j(z)$ from TABLE II of Draper and Tierney (1973) and the relative cumulants of X given in (9), the Edgeworth series expansions of the $f_n(z)$ and $F_n(z)$ correct to order n^{-3} are given by

$$\begin{aligned} f_n(z, \nu = 6) &= \phi(z) \left\{ 1 + \left[\left(\frac{1}{4!} \right) \left(\frac{6}{5} \right) H_4(z) \right] n^{-1} \right. \\ &\quad + \left[\left(\frac{1}{6!} \right) \left(\frac{48}{7} \right) H_6(z) + \left(\frac{35}{8!} \right) \left(\frac{6}{5} \right)^2 H_8(z) \right] n^{-2} \\ &\quad + \left[\left(\frac{1}{8!} \right) \left(\frac{432}{5} \right) H_8(z) \right. \\ &\quad \left. + \left(\frac{210}{10!} \right) \left(\frac{48}{7} \right) \left(\frac{6}{5} \right) H_{10}(z) + \left(\frac{5775}{12!} \right) \left(\frac{6}{5} \right)^3 H_{12}(z) \right] n^{-3} \left. \right\} + O(n^{-7/2}) \end{aligned} \quad (11)$$

and

$$\begin{aligned} F_n(z, \nu = 6) &= \Phi(z) - \phi(z) \left\{ \left[\left(\frac{1}{4!} \right) \left(\frac{6}{5} \right) H_3(z) \right] n^{-1} \right. \\ &\quad + \left[\left(\frac{1}{6!} \right) \left(\frac{48}{7} \right) H_5(z) + \left(\frac{35}{8!} \right) \left(\frac{6}{5} \right)^2 H_7(z) \right] n^{-2} \\ &\quad + \left[\left(\frac{1}{8!} \right) \left(\frac{432}{5} \right) H_7(z) \right. \\ &\quad \left. + \left(\frac{210}{10!} \right) \left(\frac{48}{7} \right) \left(\frac{6}{5} \right) H_9(z) + \left(\frac{5775}{12!} \right) \left(\frac{6}{5} \right)^3 H_{11}(z) \right] n^{-3} \left. \right\} + O(n^{-7/2}) \end{aligned} \quad (12)$$

respectively, where $H_j(x)$'s are the Hermite polynomials of degree j , which are defined by

$$\left(\frac{d}{dx} \right)^j \exp(-x^2/2) = (-1)^j H_j(x) \exp(-x^2/2), \quad j = 0, 1, \dots$$

The Hermite polynomials follow the recurrence relation

$$H_j(x) = xH_{j-1}(x) - (j-1)H_{j-2}(x), \quad j = 2, 3, \dots,$$

and are given in TABLE III in Draper and Tierney (1973) up to $j = 1(1)30$.

3 An elimination type two-stage procedure for selecting the best population

3.1 Preliminaries

Let π_i , $i = 1, \dots, k$, denote k logistic populations with unknown means μ_i and a common known variance σ^2 , and let

$$\Omega = \{\vec{\mu} = (\mu_1, \dots, \mu_k); -\infty < \mu_i < \infty, i = 1, \dots, k\}$$

be the parameter space. Denote the ranked values of the μ_i by

$$\mu_{[1]} \leq \dots \leq \mu_{[k]}$$

and let

$$\delta_{ij} = \mu_{[i]} - \mu_{[j]}.$$

We assume that the experimenter has no prior knowledge concerning the pairing of the π_i with the $\mu_{[j]}$, $i = 1, \dots, k$, $j = 1, \dots, k$. Let $\pi_{(j)}$ denote the population associated with $\mu_{[j]}$.

The goal of the experimenter is to select the 'best' population which is defined as the population with the largest mean. This event is referred to as a correct selection (CS). The experimenter restricts consideration to procedures (\mathcal{P}) which guarantee the probability requirement

$$P_{\vec{\mu}}[CS|\mathcal{P}] \geq P^*, \quad \forall \vec{\mu} \in \Omega(\delta), \quad (13)$$

where $\delta > 0$ and $1/k < P^* < 1$ are specified prior to the start of experimentation and

$$\Omega(\delta) = \{\bar{\mu} \in \Omega \mid (\mu_{[k]} - \mu_{[k-1]}) \geq \delta\}$$

which is defined as the preference zone for a correct selection.

Here we propose an elimination type two-stage procedure $\mathcal{P}_2 = \mathcal{P}_2(n_1, n_2, h)$ which depends on non-negative integers n_1, n_2 and a real constant $h > 0$ which are determined prior to the start of experimentation. The constants (n_1, n_2, h) depends on k, δ and P^* and they are chosen so that \mathcal{P}_2 guarantees the probability requirement (13) and possess a certain minimax property.

Procedure \mathcal{P}_2 ;

Stage 1: Take n_1 independent observations

$$X_{ij}^{(1)}, \quad j = 1, \dots, n_1,$$

from $\pi_i, i = 1, \dots, k$, and compute the k sample means

$$\bar{X}_i^{(1)} = \frac{1}{n_1} \sum_{j=1}^{n_1} X_{ij}^{(1)}, \quad i = 1, \dots, k.$$

Let $\bar{X}_{[k]}^{(1)} = \max_{1 \leq j \leq k} \bar{X}_j^{(1)}$. Determine the subset I of $\{1, \dots, k\}$ where

$$I = \{i \mid \bar{X}_i^{(1)} \geq \bar{X}_{[k]}^{(1)} - h\sigma/\sqrt{n_1}\},$$

and let π_I denote the associated subset of $\{\pi_1, \dots, \pi_k\}$.

1. If π_I consists of one population, stop sampling and assert that the population associated with $\bar{X}_{[k]}^{(1)}$ is best.
2. If π_I consists of more than one population, proceed to the second stage.

Stage 2: Take n_2 additional independent observations $X_{ij}^{(2)}, j = 1, \dots, n_2$, from each population in π_I , and compute the cumulative sample means

$$\begin{aligned} \bar{X}_i &= \frac{1}{n_1 + n_2} \left(\sum_{j=1}^{n_1} X_{ij}^{(1)} + \sum_{j=1}^{n_2} X_{ij}^{(2)} \right) \\ &= \frac{1}{n_1 + n_2} (n_1 \bar{X}_i^{(1)} + n_2 \bar{X}_i^{(2)}) \end{aligned}$$

for $i \in I$, where

$$\bar{X}_i^{(2)} = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{ij}^{(2)}.$$

Assert that the population associated with $\max_{i \in I} \bar{X}_i$ is the best.

There are an infinite number of combinations of (n_1, n_2, h) for given k, δ and P^* , which will exactly guarantee the probability requirement given by (13), and different design criteria lead to different choices. We will consider one of these criteria.

Let S' denote the cardinality of the set I in stage one and let

$$S = \begin{cases} 0; & \text{if } S' = 1 \\ S'; & \text{if } S' > 1. \end{cases} \quad (14)$$

Then the total sample size required by \mathcal{P}_2 , TSS say, is

$$TSS = kn_1 + Sn_2.$$

Let $E_{\bar{\mu}}[TSS|\mathcal{P}_2]$ denote the expected total sample size for \mathcal{P}_2 under $\bar{\mu}$.

We adopt the following unrestricted minimax criterion to make a choice of (n_1, n_2, h) as well as to have the total sample size TSS small. For given k and specified δ and P^* , choose (n_1, n_2, h) to

$$\begin{aligned} & \text{minimize} && \sup_{\bar{\mu} \in \Omega} E_{\bar{\mu}}[TSS|\mathcal{P}_2] \\ & \text{subject to} && \inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2] \geq P^*, \end{aligned} \quad (15)$$

where (n_1, n_2) are non-negative integers and $h \geq 0$.

For any population whose sample mean has the MLR property, Bhandari and Chaudhuri (1987) proved that the least favorable configuration (LFC) of the two-stage population means problem is a slippage configuration. However, the problem of evaluating the exact probability of a correct selection in the LFC associated with \mathcal{P}_2 is complicated and still remains to be solved. Here we will consider lower bounds for $P_{\bar{\mu}}[CS|\mathcal{P}_2]$ and construct conservative two-stage procedures.

3.2 Lower bounds for the probability of a correct selection for \mathcal{P}_2

In this section we derive lower bounds for $P_{\bar{\mu}}[CS|\mathcal{P}_2]$. These lower bounds will prove to be particularly useful since we will prove that they achieve their infimum over $\Omega(\delta)$ at $\mu(\delta)$ which has components

$$\mu = \mu_{[1]} = \cdots = \mu_{[k-1]} = \mu_{[k]} - \delta, \delta \geq 0.$$

This result will permit us to construct a conservative two-stage procedure which guarantees the probability requirement (13).

The next theorem gives one of these lower bounds for $P_{\bar{\mu}}[CS|\mathcal{P}_2]$.

Theorem 3.1 *For any $\bar{\mu} \in \Omega$ we have*

$$\begin{aligned} P_{\bar{\mu}}[CS|\mathcal{P}_2] &\geq \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} F_{n_1}(x + \delta_{ki}\sqrt{n_1}/\sigma + h) dF_{n_1}(x) \\ &\quad + \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} F_{n_1+n_2}(x + \delta_{ki}\sqrt{n_1+n_2}/\sigma) dF_{n_1+n_2}(x) - 1, \end{aligned} \quad (16)$$

where $F_n(x)$ is the cdf of the standardized sample means of size n from $L(\mu, \sigma^2)$.

Proof

For any $\bar{\mu} \in \Omega$ we have

$$\begin{aligned} P_{\bar{\mu}}[CS|\mathcal{P}_2] &= P_{\bar{\mu}}[\bar{X}_{(k)}^{(1)} \geq \bar{X}_{(i)}^{(1)} - h\sigma/\sqrt{n_1}, i \neq k, \bar{X}_{(k)} \geq \max_{i \in I} \bar{X}_{(i)}] \\ &\geq P_{\bar{\mu}}[\bar{X}_{(k)}^{(1)} \geq \bar{X}_{(i)}^{(1)} - h\sigma/\sqrt{n_1}, \bar{X}_{(k)} \geq \bar{X}_{(i)}, \forall i \neq k] \\ &\geq P_{\bar{\mu}}[\bar{X}_{(k)}^{(1)} \geq \bar{X}_i^{(1)} - h\sigma/\sqrt{n_1}, \forall i \neq k] \\ &\quad + P_{\bar{\mu}}[\bar{X}_{(k)} \geq \bar{X}_{(i)}, \forall i \neq k] - 1, \end{aligned} \quad (17)$$

since $P(A \cap B) \geq P(A) + P(B) - 1$ for any two events A and B. Then a straightforward computation leads to the conclusion of this theorem. \square

Corollary 3.1 For all $\vec{\mu} \in \Omega(\delta)$ we have

$$\inf_{\vec{\mu} \in \Omega(\delta)} P_{\vec{\mu}}[CS|P_2] \geq \quad (18)$$

$$\int_{-\infty}^{\infty} \{F_{n_1}(x + \delta\sqrt{n_1}/\sigma + h)\}^{k-1} dF_{n_1}(x) \\ + \int_{-\infty}^{\infty} \{F_{n_1+n_2}(x + \delta\sqrt{n_1+n_2}/\sigma)\}^{k-1} dF_{n_1+n_2}(x) - 1. \quad (19)$$

Proof

The proof follows immediately on noting that the right hand side of (16) is non-decreasing in each δ_{k_i} for $i = 1, \dots, k-1$. \square

Remark 3.1 Since the right hand side of (19) is strictly increasing in each of n_1 , $n_1 + n_2$ and h and tends to one as n_1 , or n_2 and h tend to ∞ , we see that the probability requirement (19) can be guaranteed if one (or more) of these constants is chosen sufficiently large.

Remark 3.2 If we let $h \rightarrow \infty$ on the right hand side of (16) we obtain

$$\int_{-\infty}^{\infty} \prod_{i=1}^{k-1} F_{n_1+n_2}(x + \delta_{k_i}\sqrt{n_1+n_2}/\sigma) dF_{n_1+n_2}(x)$$

which is an expression for $P_{\vec{\mu}}[CS|P_1]$ where P_1 uses a common single-stage sample size $n = n_1 + n_2$ per population. Thus P_1 is a special case of P_2 based on a conservative lower bound and hence $E_{\vec{\mu}}[TSS|P_2] \leq kn$ for all $\vec{\mu} \in \Omega$.

Remark 3.3 The distribution of the mean of samples from logistic population has the monotone likelihood ratio (MLR) property with respect to the location parameter (Goel (1975)) and hence the distributions of the $\bar{X}_i^{(1)}$ and $\bar{X}_i^{(2)}$ are stochastically increasing (SI) families in μ_i , $i = 1, \dots, k$.

Remark 3.4 The cumulative sample means

$$\bar{X}_i = \frac{n_1}{n_1 + n_2} \bar{X}_i^{(1)} + \frac{n_2}{n_1 + n_2} \bar{X}_i^{(2)}$$

are strictly increasing in each $\bar{X}_i^{(j)}$, $j = 1, 2$, $i = 1, \dots, k$.

We can now find another lower bound to the $P_{\bar{\mu}}[CS|\mathcal{P}_2]$ given in the following theorem by noting the facts mentioned in Remark 3.3 and Remark 3.4. This lower bound can be shown to be uniformly superior to the one given in Theorem 3.1. It is also straightforward to determine the LFC of the population means relative to this new lower bound.

Theorem 3.2 *For any $\bar{\mu} \in \Omega$ we have*

$$\begin{aligned} & \inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2] \\ & \geq \int_{-\infty}^{\infty} \{F_{n_1}(x + \delta\sqrt{n_1}/\sigma + h)\}^{k-1} dF_{n_1}(x) \\ & \quad \cdot \int_{-\infty}^{\infty} \{F_{n_1+n_2}(x + \delta\sqrt{n_1+n_2}/\sigma)\}^{k-1} dF_{n_1+n_2}(x), \end{aligned} \quad (20)$$

where $F_n(x)$ is the cdf of the standardized sample mean of size n from $L(\mu, \sigma^2)$.

Proof

Let $F(\cdot|\mu_i)$ and $G(\cdot|\mu_i)$ denote the cdf's of the $\bar{X}_i^{(1)}$ and \bar{X}_i , respectively and let $H(\cdot, \cdot|\mu_i)$ denote the joint cdf of the $\bar{X}_i^{(1)}$ and \bar{X}_i . Then $F(\cdot|\mu_i)$, $G(\cdot|\mu_i)$ and $H(\cdot, \cdot|\mu_i)$ are non-increasing in μ_i , $i = 1, \dots, k$, from Remark 3.3 and Remark 3.4. Without loss of generality we may assume that $\mu_1 \leq \dots \leq \mu_k$. Then for all $\bar{\mu} \in \Omega(\delta)$,

$$\begin{aligned} & P_{\bar{\mu}}[CS|\mathcal{P}_2] \\ & = P_{\bar{\mu}}[\bar{X}_{(k)}^{(1)} \geq \max_{1 \leq j \leq k} \bar{X}_{(j)}^{(1)} - h\sigma/\sqrt{n_1}, \bar{X}_{(k)} = \max_{j \in I} \bar{X}_{(j)}] \\ & \geq P_{\bar{\mu}}[\bar{X}_{(k)}^{(1)} \geq \bar{X}_{(j)}^{(1)} - h\sigma/\sqrt{n_1}, \bar{X}_{(k)} \geq \bar{X}_{(j)}, \forall j = 1, \dots, k-1] \\ & = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} H(x + h\sigma/\sqrt{n_1}, y|\mu_i) dH(x, y|\mu_k) \\ & \geq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} H(x + h\sigma/\sqrt{n_1}, y|\mu_k - \delta) dH(x, y|\mu_k) \\ & = E_{\mu_k}[H^{k-1}\{\bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1}, \bar{X}_{(k)}|\mu_k - \delta\}], \end{aligned}$$

where the expectation is with respect to the joint distribution of $\bar{X}_{(k)}^{(1)}$ and $\bar{X}_{(k)}$. Hence

$$\inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2] \geq \inf_{\bar{\mu} \in \Omega(\delta)} E_{\mu_k}[H^{k-1}\{\bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1}, \bar{X}_{(k)}|\mu_k - \delta\}]$$

and it is enough to show that for all $\bar{\mu} \in \Omega(\delta)$,

$$\begin{aligned} & E_{\mu_k} [H^{k-1} \{ \bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1}, \bar{X}_{(k)} | \mu_k - \delta \}] \\ & \geq E_{\mu_k} [F^{k-1} (\bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1} | \mu_k - \delta)] E_{\mu_k} [G^{k-1} (\bar{X}_{(k)} | \mu_k - \delta)]. \end{aligned}$$

By Remark 3.4, for all a, b and μ ,

$$\begin{aligned} & P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq a, \bar{X}_{(j)} \leq b \} \\ & = P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq a, \bar{X}_{(j)}^{(1)} \leq \frac{n_1+n_2}{n_1} (b - \frac{n_2}{n_1+n_2} \bar{X}_{(j)}^{(2)}) \} \\ & = E_{\mu} [P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq a, \bar{X}_{(j)}^{(1)} \leq \frac{n_1+n_2}{n_1} (b - \frac{n_2}{n_1+n_2} \bar{X}_{(j)}^{(2)}) | \bar{X}_{(j)}^{(2)} \}] \\ & \geq E_{\mu} [P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq a | \bar{X}_{(j)}^{(2)} \}] \\ & \quad \cdot P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq \frac{n_1+n_2}{n_1} (b - \frac{n_2}{n_1+n_2} \bar{X}_{(j)}^{(2)}) | \bar{X}_{(j)}^{(2)} \}] \\ & = P_{\mu} \{ \bar{X}_{(j)}^{(1)} \leq a \} P_{\mu} \{ \bar{X}_{(j)} \leq b \} \end{aligned}$$

and hence

$$\begin{aligned} & E_{\mu_k} [H^{k-1} \{ \bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1}, \bar{X}_{(k)} | \mu_k - \delta \}] \\ & \geq E_{\mu_k} [F^{k-1} \{ \bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1} | \mu_k - \delta \}] \\ & \quad \cdot G^{k-1} \{ \bar{X}_{(k)} | \mu_k - \delta \}] \\ & \geq E_{\mu_k} [F^{k-1} \{ \bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1} | \mu_k - \delta \}] \\ & \quad \cdot E_{\mu_k} [G^{k-1} \{ \bar{X}_{(k)} | \mu_k - \delta \}] \end{aligned}$$

by the Chebyshev's inequality (Hardy, Littlewood and Pólya (1934)), since

$$F \{ \bar{X}_{(k)}^{(1)} + h\sigma/\sqrt{n_1} | \mu_k - \delta \}$$

and

$$G \{ \bar{X}_{(k)} | \mu_k - \delta \}$$

are non-decreasing in $\bar{X}_k^{(1)}$. \square

Remark 3.5 *If we let*

$$a = \int_{-\infty}^{\infty} \{ F_{n_1}(x + \delta\sqrt{n_1}/\sigma + h) \}^{k-1} dF_{n_1}(x)$$

and

$$b = \int_{-\infty}^{\infty} \{F_{n_1+n_2}(x + \delta\sqrt{n_1+n_2}/\sigma)\}^{k-1} dF_{n_1+n_2}(x),$$

then (19) states that

$$\inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2] \geq a + b - 1$$

and (20) states that

$$\inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2] \geq ab.$$

By noting that $a + b - 1 < ab$ for all $a, b \in (0, 1)$, the lower bound (20) is uniformly superior to the lower bound (19), and hence we will use the lower bound (20) henceforth.

3.3 Expected total sample size for \mathcal{P}_2

In order to solve the optimization problem (15) we first find an analytical expression for the $E_{\bar{\mu}}[TSS|\mathcal{P}_2]$ and then determine the $\sup_{\bar{\mu} \in \Omega} E_{\bar{\mu}}[TSS|\mathcal{P}_2]$ and the sets of μ_i -values at which this supremum occurs.

Theorem 3.3 For any $\bar{\mu} \in \Omega$ we have

$$\begin{aligned} E_{\bar{\mu}}[TSS|\mathcal{P}_2] &= kn_1 + n_2 \sum_{i=1}^k \int_{-\infty}^{\infty} \left\{ \prod_{\substack{j=1 \\ j \neq i}}^k F_{n_1}(x + \delta_{ij}\sqrt{n_1}/\sigma + h) \right. \\ &\quad \left. - \prod_{\substack{j=1 \\ j \neq i}}^k F_{n_1}(x + \delta_{ij}\sqrt{n_1}/\sigma - h) \right\} dF_{n_1}(x), \end{aligned} \quad (21)$$

where $F_n(x)$ is the cdf of the standardized sample means of size n from $L(\mu, \sigma^2)$.

Proof

For any $\bar{\mu} \in \Omega$ we have

$$E_{\bar{\mu}}[TSS|\mathcal{P}_2] = kn_1 + n_2 E_{\bar{\mu}}[S|\mathcal{P}_2],$$

where S is defined as in (14). Now

$$E_{\bar{\mu}}[S|\mathcal{P}_2] = E_{\bar{\mu}}[S'|\mathcal{P}_2] - P_{\bar{\mu}}[S' = 1|\mathcal{P}_2]$$

$$\begin{aligned}
&= \sum_{i=1}^k P_{\bar{\mu}}[\bar{X}_{(i)}^{(1)} \geq \bar{X}_{(j)}^{(1)} - h\sigma/\sqrt{n_1}, \forall j \neq i] \\
&\quad - \sum_{i=1}^k P_{\bar{\mu}}[\bar{X}_{(i)}^{(1)} \geq \bar{X}_{(j)}^{(1)} + h\sigma/\sqrt{n_1}, \forall j \neq i] \quad (22)
\end{aligned}$$

and hence Theorem 3.3 follows immediately. \square

The following theorem summarizes the result concerning the supremum of the $E_{\bar{\mu}}[TSS|\mathcal{P}_2]$ for $\bar{\mu} \in \Omega$.

Theorem 3.4 For any $\bar{\mu} \in \Omega$, fixed k and (n_1, n_2, h) we have

$$\begin{aligned}
&\sup_{\bar{\mu} \in \Omega} E_{\bar{\mu}}[TSS|\mathcal{P}_2] \\
&= kn_1 + n_2 \int_{-\infty}^{\infty} [\{F_{n_1}(x+h)\}^{k-1} - \{F_{n_1}(x-h)\}^{k-1}] dF_{n_1}(x) \quad (23)
\end{aligned}$$

which occurs when $\mu_{[1]} = \dots = \mu_{[k]}$, where $F_n(x)$ is the cdf of the standardized sample means of size n from $L(\mu, \sigma^2)$.

Proof

Noting Remark 3.3 and Remark 3.4 we can use the results of Gupta (1965) which show that $E_{\bar{\mu}}[S'|\mathcal{P}_2]$ achieves its supremum for $\bar{\mu} \in \Omega$ when $\mu_{[1]} = \dots = \mu_{[k]}$. By a similar argument $P_{\bar{\mu}}[S' = 1|\mathcal{P}_2]$ achieves its infimum when $\mu_{[1]} = \dots = \mu_{[k]}$. Hence the result follows immediately from Theorem 3.3. \square

3.4 Optimization problem yielding conservative solutions

In this section we consider the optimization problem (15) which one must solve in order to determine the constants (n_1, n_2, h) which are necessary to implement \mathcal{P}_2 . As we noted earlier, the problem of evaluating the exact probability of a correct selection in the LFC associated with \mathcal{P}_2 is very complicated. Thus we replace the exact $\inf_{\bar{\mu} \in \Omega(\delta)} P_{\bar{\mu}}[CS|\mathcal{P}_2]$ by the conservative lower bound given by the right hand side of (20), and consider the following optimization problem.

For the given k, δ and P^* choose the constants (n_1, n_2, h) to

$$\text{minimize} \quad kn_1 + n_2 \int_{-\infty}^{\infty} [\{F_{n_1}(x+h)\}^{k-1} - \{F_{n_1}(x-h)\}^{k-1}] dF_{n_1}(x)$$

$$\text{subject to } \int_{-\infty}^{\infty} \{F_{n_1}(x + \delta\sqrt{n_1}/\sigma + h)\}^{k-1} \cdot \int_{-\infty}^{\infty} \{F_{n_1+n_2}(x + \delta\sqrt{n_1+n_2}/\sigma)\}^{k-1} dF_{n_1+n_2}(x) \geq P^*, \quad (24)$$

where n_1 and n_2 are non-negative integers and $h \geq 0$.

Let us denote by $(\hat{n}_1, \hat{n}_2, \hat{h})$ the solution to the optimization problem (24). Then we can use the approximate design constants

$$n_1 = [\hat{n}_1 + 1], \quad n_2 = [\hat{n}_2 + 1], \quad h = \hat{h},$$

where $[z]$ denotes the greatest integer which is less than z , to implement \mathcal{P}_2 .

Table 1, Table 2, Table 3 and Table 4 contain the constants $(\hat{n}_1, \hat{n}_2, \hat{h})$ necessary to approximate (n_1, n_2, h) and the values of the expected total sample size (ETSS) for $k = 2, 3, 4, 5, 10, 15$, $P^* = 0.75, 0.90, 0.95, 0.99$ and $\delta/\sigma = 0.1, 0.5, 1.0, 2.0, 4.0$. All computations were carried out in double-precision arithmetic on a Vax-11/780. The SUMT (Sequential Unconstrained Minimization Techniques: Fiacco and McCormick (1968)) algorithm is used to solve the non-linear optimization problem. A source program in Fortran for the SUMT algorithm is given by Kuester and Mize (1973).

3.5 The performance of the two-stage procedure relative to the single-stage procedure

As a measure of efficiency of the two-stage procedure \mathcal{P}_2 relative to that of the single-stage procedure \mathcal{P}_1 when both guarantee the same basic probability requirement (13), we consider the ratio termed relative efficiency (RE) $E_{\bar{\mu}}[TSS|\mathcal{P}_2]/k\hat{n}_s$, where \hat{n}_s is the estimate of the minimum sample size n_s needed in the single-stage procedure \mathcal{P}_1 . Clearly RE depends on $\bar{\mu}$, δ and P^* . Values of the RE less than unity favor \mathcal{P}_2 over \mathcal{P}_1 .

Now the RE is given by

$$RE = \frac{1}{k\hat{n}_s} [k\hat{n}_1 + \hat{n}_2 \sum_{i=1}^k \int_{-\infty}^{\infty} \{ \prod_{\substack{j=1 \\ j \neq i}}^k F_{\hat{n}_1}(t + \delta_{ij}\sqrt{\hat{n}_1}/\sigma + \hat{h})$$

$$- \prod_{\substack{j=1 \\ j \neq i}}^k F_{\hat{n}_1}(t + \delta_{ij} \sqrt{\hat{n}_1}/\sigma - \hat{h}) \} dF_{\hat{n}_1}(t)]. \quad (25)$$

where \hat{n}_s is the solution of

$$\int_{-\infty}^{\infty} \{F_{\hat{n}_s}(t + (\sqrt{\hat{n}_s}/\sigma)\delta)\}^{k-1} dF_{\hat{n}_s}(t) = P^*. \quad (26)$$

We consider the relative efficiency for two special cases, namely, the equally spaced and the slippage configurations. First, for the equally spaced configuration, we assume that the unknown means of the k populations are $\mu, \mu + \delta, \dots, \mu + (k-1)\delta$ which have ranks 1, 2, \dots, k , respectively. Let RE_{eq} denote the relative efficiency with respect to the above configuration. Then, since $\delta_{ij} = \mu_{[i]} - \mu_{[j]} = (i-j)\delta$,

$$RE_{eq} = \frac{1}{k\hat{n}_s} [k\hat{n}_1 + \hat{n}_2 \sum_{i=1}^k \int_{-\infty}^{\infty} \{ \prod_{\substack{j=1 \\ j \neq i}}^k F_{\hat{n}_1}(t + \sqrt{\hat{n}_1}(i-j)\delta/\sigma + \hat{h}) \\ - \prod_{\substack{j=1 \\ j \neq i}}^k F_{\hat{n}_1}(t + \sqrt{\hat{n}_1}(i-j)\delta/\sigma - \hat{h}) \} dF_{\hat{n}_1}(t)]. \quad (27)$$

Next, for the slippage configuration, we assume that the unknown means of the k populations are $\mu_{[j]} = \mu, j = 1, \dots, k-1$, and $\mu_{[k]} = \mu + \delta, \delta \geq 0$. Then the relative efficiency with respect to the above configuration, RE_{sp} , is given by

$$RE_{sp} = \frac{1}{k\hat{n}_s} [k\hat{n}_1 + \hat{n}_2 \{ (k-1) \int_{-\infty}^{\infty} (F_{\hat{n}_1}(t + \hat{h}) - F_{\hat{n}_1}(t - \hat{h}))^{k-2} \\ \cdot (F_{\hat{n}_1}(t - \sqrt{\hat{n}_1}\delta/\sigma + \hat{h}) - F_{\hat{n}_1}(t - \sqrt{\hat{n}_1}\delta/\sigma - \hat{h})) dF_{\hat{n}_1}(t) \\ + \int_{-\infty}^{\infty} (F_{\hat{n}_1}(t + \sqrt{\hat{n}_1}\delta/\sigma + \hat{h}) - F_{\hat{n}_1}(t + \sqrt{\hat{n}_1}\delta/\sigma - \hat{h}))^{k-1} dF_{\hat{n}_1}(t) \}]. \quad (28)$$

Table 5 and Table 6 give the values of the RE_{eq} and RE_{sp} for given values of $P^* = 0.75, 0.90, 0.95, 0.99, k = 2, 3, 4, 5, 10, 15$ and $\delta/\sigma = 0.1, 0.5, 1.0, 2.0, 4.0$.

For any values of P^*, k and δ , $RE_{eq} \leq 1$ and $RE_{sp} \leq 1$ and hence the two-stage procedure is more efficient than the single-stage procedure in terms of the expected total sample sizes. Furthermore, the effectiveness of \mathcal{P}_2 appears to be increasing in k since the values of RE_{eq} and RE_{sp} are decreasing in k .

Table 1: Constants to implement the two-stage procedure \mathcal{P}_2 for selecting the largest logistic population: $P^* = 0.75$.

| $P^* = 0.75$ | | | | | |
|--------------|-----------------|-------------|-------------|------------|--------------|
| k | δ/σ | \hat{a}_1 | \hat{a}_2 | \hat{h} | ETSS |
| 2 | 0.10 | 0.4548e+02 | 0.4539e+02 | 0.5530e+01 | 0.181720e+03 |
| | 0.50 | 0.2620e+01 | 0.8929e+00 | 0.7323e+01 | 0.702648e+01 |
| | 1.00 | 0.3121e+00 | 0.4983e+00 | 0.7097e+01 | 0.162056e+01 |
| | 2.00 | 0.9556e-01 | 0.1070e+00 | 0.6698e+01 | 0.405153e+00 |
| | 4.00 | 0.2834e-01 | 0.2232e-01 | 0.6127e+01 | 0.101300e+00 |
| 3 | 0.10 | 0.1018e+03 | 0.1044e+03 | 0.3688e+01 | 0.615938e+03 |
| | 0.50 | 0.4778e+01 | 0.3340e+01 | 0.4516e+01 | 0.243344e+02 |
| | 1.00 | 0.8971e+00 | 0.1050e+01 | 0.7309e+01 | 0.584085e+01 |
| | 2.00 | 0.1876e+00 | 0.2774e+00 | 0.6606e+01 | 0.139489e+01 |
| | 4.00 | 0.8039e-01 | 0.3586e-01 | 0.6112e+01 | 0.348738e+00 |
| 4 | 0.10 | 0.1392e+03 | 0.1515e+03 | 0.2751e+01 | 0.112241e+04 |
| | 0.50 | 0.5900e+01 | 0.5554e+01 | 0.2947e+01 | 0.447121e+02 |
| | 1.00 | 0.1711e+01 | 0.1037e+01 | 0.4159e+01 | 0.109639e+02 |
| | 2.00 | 0.3255e+00 | 0.3270e+00 | 0.5857e+01 | 0.260899e+01 |
| | 4.00 | 0.8061e-01 | 0.8252e-01 | 0.5737e+01 | 0.652252e+00 |
| 5 | 0.10 | 0.1631e+03 | 0.2013e+03 | 0.2278e+01 | 0.166485e+04 |
| | 0.50 | 0.6766e+01 | 0.7657e+01 | 0.2341e+01 | 0.666234e+02 |
| | 1.00 | 0.1826e+01 | 0.1630e+01 | 0.2746e+01 | 0.165961e+02 |
| | 2.00 | 0.3864e+00 | 0.4182e+00 | 0.4050e+01 | 0.399414e+01 |
| | 4.00 | 0.9599e-01 | 0.1051e+00 | 0.4087e+01 | 0.998538e+00 |
| 10 | 0.10 | 0.2357e+03 | 0.4304e+03 | 0.1494e+01 | 0.451824e+04 |
| | 0.50 | 0.9587e+01 | 0.1738e+02 | 0.1468e+01 | 0.181456e+03 |
| | 1.00 | 0.2504e+01 | 0.4455e+01 | 0.1398e+01 | 0.458878e+02 |
| | 2.00 | 0.6367e+00 | 0.1178e+01 | 0.1361e+01 | 0.117515e+02 |
| | 4.00 | 0.1582e+00 | 0.2758e+00 | 0.1470e+01 | 0.295260e+01 |
| 15 | 0.10 | 0.2714e+03 | 0.5855e+03 | 0.1369e+01 | 0.744887e+04 |
| | 0.50 | 0.1100e+02 | 0.2372e+02 | 0.1352e+01 | 0.299197e+03 |
| | 1.00 | 0.2858e+01 | 0.6119e+01 | 0.1308e+01 | 0.757369e+02 |
| | 2.00 | 0.7466e+00 | 0.1668e+01 | 0.1255e+01 | 0.195906e+02 |
| | 4.00 | 0.1906e+00 | 0.4032e+00 | 0.1313e+01 | 0.501004e+01 |

Table 2: Constants to implement the two-stage procedure \mathcal{P}_2 for selecting the largest logistic population: $P^* = 0.90$.

| $P^* = 0.90$ | | | | | |
|--------------|-----------------|-------------|-------------|------------|--------------|
| k | δ/σ | \hat{n}_1 | \hat{n}_2 | \hat{h} | ETSS |
| 2 | 0.10 | 0.1668e+03 | 0.1728e+03 | 0.2446e+01 | 0.650194e+03 |
| | 0.50 | 0.7013e+01 | 0.6404e+01 | 0.2591e+01 | 0.259726e+02 |
| | 1.00 | 0.1932e+01 | 0.1311e+01 | 0.3369e+01 | 0.643201e+01 |
| | 2.00 | 0.4011e+00 | 0.3724e+00 | 0.5331e+01 | 0.154620e+01 |
| | 4.00 | 0.1044e+00 | 0.8907e-01 | 0.5026e+01 | 0.386564e+00 |
| 3 | 0.10 | 0.2745e+03 | 0.2513e+03 | 0.2017e+01 | 0.146152e+04 |
| | 0.50 | 0.1126e+02 | 0.9634e+01 | 0.2071e+01 | 0.585665e+02 |
| | 1.00 | 0.2971e+01 | 0.2135e+01 | 0.2332e+01 | 0.146860e+02 |
| | 2.00 | 0.6894e+00 | 0.5189e+00 | 0.5004e+01 | 0.362197e+01 |
| | 4.00 | 0.1693e+00 | 0.1310e+00 | 0.4955e+01 | 0.900049e+00 |
| 4 | 0.10 | 0.3298e+03 | 0.3318e+03 | 0.1713e+01 | 0.229940e+04 |
| | 0.50 | 0.1340e+02 | 0.1300e+02 | 0.1728e+01 | 0.922982e+02 |
| | 1.00 | 0.3489e+01 | 0.3048e+01 | 0.1796e+01 | 0.232917e+02 |
| | 2.00 | 0.8374e+00 | 0.7008e+00 | 0.2643e+01 | 0.592662e+01 |
| | 4.00 | 0.2090e+00 | 0.1704e+00 | 0.2831e+01 | 0.147462e+01 |
| 5 | 0.10 | 0.3664e+03 | 0.4034e+03 | 0.1556e+01 | 0.315013e+04 |
| | 0.50 | 0.1488e+02 | 0.1595e+02 | 0.1553e+01 | 0.126542e+03 |
| | 1.00 | 0.3863e+01 | 0.3858e+01 | 0.1559e+01 | 0.320185e+02 |
| | 2.00 | 0.9610e+00 | 0.9217e+00 | 0.1867e+01 | 0.829954e+01 |
| | 4.00 | 0.2403e+00 | 0.2184e+00 | 0.2071e+01 | 0.208230e+01 |
| 10 | 0.10 | 0.4549e+03 | 0.6465e+03 | 0.1367e+01 | 0.750100e+04 |
| | 0.50 | 0.1844e+02 | 0.2588e+02 | 0.1357e+01 | 0.301614e+03 |
| | 1.00 | 0.4784e+01 | 0.6497e+01 | 0.1328e+01 | 0.765662e+02 |
| | 2.00 | 0.1328e+01 | 0.1644e+01 | 0.1257e+01 | 0.201395e+02 |
| | 4.00 | 0.3335e+00 | 0.4262e+00 | 0.1362e+01 | 0.528481e+01 |
| 15 | 0.10 | 0.4934e+03 | 0.7911e+03 | 0.1366e+01 | 0.119540e+05 |
| | 0.50 | 0.1999e+02 | 0.3177e+02 | 0.1358e+01 | 0.480822e+03 |
| | 1.00 | 0.5180e+01 | 0.8022e+01 | 0.1335e+01 | 0.122187e+03 |
| | 2.00 | 0.1433e+01 | 0.2074e+01 | 0.1280e+01 | 0.322460e+02 |
| | 4.00 | 0.3751e+00 | 0.5593e+00 | 0.1328e+01 | 0.865274e+01 |

Table 3: Constants to implement the two-stage procedure \mathcal{P}_2 for selecting the largest logistic population: $P^* = 0.95$.

| $P^* = 0.95$ | | | | | |
|--------------|-----------------|-------------|-------------|------------|--------------|
| k | δ/σ | \hat{a}_1 | \hat{a}_2 | \hat{h} | ETSS |
| 2 | 0.10 | 0.3008e+03 | 0.2827e+03 | 0.1781e+01 | 0.104953e+04 |
| | 0.50 | 0.1227e+02 | 0.1098e+02 | 0.1810e+01 | 0.421247e+02 |
| | 1.00 | 0.3215e+01 | 0.2504e+01 | 0.1958e+01 | 0.106222e+02 |
| | 2.00 | 0.7631e+00 | 0.5883e+00 | 0.3556e+01 | 0.268233e+01 |
| | 4.00 | 0.1899e+00 | 0.1457e+00 | 0.3785e+01 | 0.667647e+00 |
| 3 | 0.10 | 0.4362e+03 | 0.3657e+03 | 0.1574e+01 | 0.211419e+04 |
| | 0.50 | 0.1768e+02 | 0.1436e+02 | 0.1589e+01 | 0.849214e+02 |
| | 1.00 | 0.4579e+01 | 0.3388e+01 | 0.1654e+01 | 0.214801e+02 |
| | 2.00 | 0.1223e+01 | 0.6952e+00 | 0.2269e+01 | 0.553339e+01 |
| | 4.00 | 0.2858e+00 | 0.1853e+00 | 0.3237e+01 | 0.139794e+01 |
| 4 | 0.10 | 0.4991e+03 | 0.4519e+03 | 0.1452e+01 | 0.318364e+04 |
| | 0.50 | 0.2023e+02 | 0.1787e+02 | 0.1453e+01 | 0.127954e+03 |
| | 1.00 | 0.5232e+01 | 0.4325e+01 | 0.1464e+01 | 0.324315e+02 |
| | 2.00 | 0.1420e+01 | 0.9417e+00 | 0.1675e+01 | 0.846044e+01 |
| | 4.00 | 0.3393e+00 | 0.2423e+00 | 0.2163e+01 | 0.218343e+01 |
| 5 | 0.10 | 0.5381e+03 | 0.5259e+03 | 0.1392e+01 | 0.426098e+04 |
| | 0.50 | 0.2182e+02 | 0.2086e+02 | 0.1388e+01 | 0.171314e+03 |
| | 1.00 | 0.5649e+01 | 0.5112e+01 | 0.1379e+01 | 0.434710e+02 |
| | 2.00 | 0.1546e+01 | 0.1182e+01 | 0.1430e+01 | 0.114045e+02 |
| | 4.00 | 0.3809e+00 | 0.3002e+00 | 0.1751e+01 | 0.299628e+01 |
| 10 | 0.10 | 0.6279e+03 | 0.7682e+03 | 0.1349e+01 | 0.973702e+04 |
| | 0.50 | 0.2544e+02 | 0.3070e+02 | 0.1342e+01 | 0.391770e+03 |
| | 1.00 | 0.6592e+01 | 0.7667e+01 | 0.1321e+01 | 0.996400e+02 |
| | 2.00 | 0.1827e+01 | 0.1923e+01 | 0.1269e+01 | 0.263641e+02 |
| | 4.00 | 0.4897e+00 | 0.5216e+00 | 0.1344e+01 | 0.724983e+01 |
| 15 | 0.10 | 0.6674e+03 | 0.9126e+03 | 0.1377e+01 | 0.153152e+05 |
| | 0.50 | 0.2703e+02 | 0.3659e+02 | 0.1370e+01 | 0.616396e+03 |
| | 1.00 | 0.7002e+01 | 0.9178e+01 | 0.1354e+01 | 0.156917e+03 |
| | 2.00 | 0.1942e+01 | 0.2339e+01 | 0.1310e+01 | 0.416523e+02 |
| | 4.00 | 0.5293e+00 | 0.6784e+00 | 0.1300e+01 | 0.115109e+02 |

Table 4: Constants to implement the two-stage procedure P_2 for selecting the largest logistic population: $P^* = 0.99$.

| $P^* = 0.99$ | | | | | |
|--------------|-----------------|-------------|-------------|------------|--------------|
| k | δ/σ | \hat{n}_1 | \hat{n}_2 | \hat{h} | ETSS |
| 2 | 0.10 | 0.6892e+03 | 0.5071e+03 | 0.1295e+01 | 0.202774e+04 |
| | 0.50 | 0.2784e+02 | 0.2014e+02 | 0.1300e+01 | 0.815644e+02 |
| | 1.00 | 0.7189e+01 | 0.4895e+01 | 0.1309e+01 | 0.207286e+02 |
| | 2.00 | 0.1966e+01 | 0.1107e+01 | 0.1414e+01 | 0.546768e+01 |
| | 4.00 | 0.5013e+00 | 0.2771e+00 | 0.2143e+01 | 0.148773e+01 |
| 3 | 0.10 | 0.8588e+03 | 0.5804e+03 | 0.1248e+01 | 0.366047e+04 |
| | 0.50 | 0.3475e+02 | 0.2299e+02 | 0.1249e+01 | 0.147249e+03 |
| | 1.00 | 0.8954e+01 | 0.5614e+01 | 0.1254e+01 | 0.374304e+02 |
| | 2.00 | 0.2442e+01 | 0.1300e+01 | 0.1314e+01 | 0.988866e+01 |
| | 4.00 | 0.6518e+00 | 0.3209e+00 | 0.2052e+01 | 0.278376e+01 |
| 4 | 0.10 | 0.9268e+03 | 0.6663e+03 | 0.1254e+01 | 0.526885e+04 |
| | 0.50 | 0.3750e+02 | 0.2646e+02 | 0.1253e+01 | 0.211992e+03 |
| | 1.00 | 0.9668e+01 | 0.6501e+01 | 0.1250e+01 | 0.539214e+02 |
| | 2.00 | 0.2647e+01 | 0.1542e+01 | 0.1266e+01 | 0.142783e+02 |
| | 4.00 | 0.7350e+00 | 0.4650e+00 | 0.1230e+01 | 0.404965e+01 |
| 5 | 0.10 | 0.9639e+03 | 0.7432e+03 | 0.1271e+01 | 0.687580e+04 |
| | 0.50 | 0.3903e+02 | 0.2955e+02 | 0.1266e+01 | 0.276700e+03 |
| | 1.00 | 0.1008e+02 | 0.7283e+01 | 0.1258e+01 | 0.704195e+02 |
| | 2.00 | 0.2770e+01 | 0.1751e+01 | 0.1252e+01 | 0.186821e+02 |
| | 4.00 | 0.7668e+00 | 0.4714e+00 | 0.1451e+01 | 0.532902e+01 |
| 10 | 0.10 | 0.1049e+04 | 0.9971e+03 | 0.1343e+01 | 0.149575e+05 |
| | 0.50 | 0.4246e+02 | 0.3976e+02 | 0.1340e+01 | 0.602284e+03 |
| | 1.00 | 0.1099e+02 | 0.9879e+01 | 0.1327e+01 | 0.153541e+03 |
| | 2.00 | 0.3045e+01 | 0.2443e+01 | 0.1295e+01 | 0.409579e+02 |
| | 4.00 | 0.8789e+00 | 0.6787e+00 | 0.1394e+01 | 0.119728e+02 |
| 15 | 0.10 | 0.1088e+04 | 0.1147e+04 | 0.1400e+01 | 0.231194e+05 |
| | 0.50 | 0.4405e+02 | 0.4583e+02 | 0.1396e+01 | 0.931205e+03 |
| | 1.00 | 0.1140e+02 | 0.1143e+02 | 0.1384e+01 | 0.237593e+03 |
| | 2.00 | 0.3166e+01 | 0.2855e+01 | 0.1356e+01 | 0.635515e+02 |
| | 4.00 | 0.9277e+00 | 0.8022e+00 | 0.1455e+01 | 0.187987e+02 |

Table 5: Relative efficiency of the two-stage procedure \mathcal{P}_2 : Equally spaced configuration.

| Equally Spaced Configuration | | | | | | |
|------------------------------|-----|-----------------|-------|-------|-------|-------|
| P^* | k | δ/σ | | | | |
| | | 0.1 | 0.5 | 1.0 | 2.0 | 4.0 |
| 0.750 | 2 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 3 | 0.973 | 0.991 | 1.000 | 0.999 | 0.999 |
| | 4 | 0.836 | 0.857 | 0.941 | 0.993 | 0.991 |
| | 5 | 0.720 | 0.732 | 0.776 | 0.858 | 0.861 |
| | 10 | 0.551 | 0.556 | 0.572 | 0.581 | 0.582 |
| | 15 | 0.503 | 0.507 | 0.520 | 0.530 | 0.536 |
| 0.900 | 2 | 0.922 | 0.935 | 0.975 | 0.998 | 0.998 |
| | 3 | 0.782 | 0.792 | 0.824 | 0.981 | 0.981 |
| | 4 | 0.689 | 0.695 | 0.713 | 0.762 | 0.776 |
| | 5 | 0.642 | 0.648 | 0.663 | 0.677 | 0.687 |
| | 10 | 0.554 | 0.559 | 0.573 | 0.612 | 0.603 |
| | 15 | 0.518 | 0.523 | 0.535 | 0.570 | 0.569 |
| 0.950 | 2 | 0.820 | 0.826 | 0.847 | 0.965 | 0.974 |
| | 3 | 0.716 | 0.721 | 0.736 | 0.795 | 0.852 |
| | 4 | 0.664 | 0.669 | 0.683 | 0.721 | 0.718 |
| | 5 | 0.634 | 0.639 | 0.653 | 0.690 | 0.678 |
| | 10 | 0.564 | 0.569 | 0.582 | 0.620 | 0.625 |
| | 15 | 0.533 | 0.537 | 0.550 | 0.586 | 0.585 |
| 0.990 | 2 | 0.716 | 0.719 | 0.729 | 0.758 | 0.782 |
| | 3 | 0.690 | 0.695 | 0.706 | 0.737 | 0.741 |
| | 4 | 0.666 | 0.670 | 0.682 | 0.715 | 0.713 |
| | 5 | 0.646 | 0.650 | 0.663 | 0.698 | 0.687 |
| | 10 | 0.591 | 0.596 | 0.609 | 0.646 | 0.668 |
| | 15 | 0.565 | 0.570 | 0.582 | 0.619 | 0.645 |

Table 6: Relative efficiency of the two-stage procedure P_2 : Slippage configuration.

| Slippage Configuration | | | | | | |
|------------------------|-----|-----------------|-------|-------|-------|-------|
| P^* | k | δ/σ | | | | |
| | | 0.1 | 0.5 | 1.0 | 2.0 | 4.0 |
| 0.750 | 2 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| | 3 | 0.982 | 0.995 | 1.000 | 0.999 | 0.999 |
| | 4 | 0.907 | 0.930 | 0.988 | 0.999 | 0.998 |
| | 5 | 0.811 | 0.829 | 0.898 | 0.980 | 0.981 |
| | 10 | 0.516 | 0.520 | 0.531 | 0.540 | 0.557 |
| | 15 | 0.457 | 0.461 | 0.474 | 0.485 | 0.495 |
| 0.900 | 2 | 0.922 | 0.935 | 0.975 | 0.998 | 0.998 |
| | 3 | 0.796 | 0.809 | 0.852 | 0.994 | 0.993 |
| | 4 | 0.698 | 0.706 | 0.730 | 0.854 | 0.877 |
| | 5 | 0.636 | 0.642 | 0.658 | 0.711 | 0.745 |
| | 10 | 0.527 | 0.532 | 0.546 | 0.590 | 0.583 |
| | 15 | 0.494 | 0.499 | 0.513 | 0.551 | 0.551 |
| 0.950 | 2 | 0.820 | 0.826 | 0.847 | 0.965 | 0.974 |
| | 3 | 0.709 | 0.715 | 0.734 | 0.818 | 0.908 |
| | 4 | 0.651 | 0.656 | 0.671 | 0.722 | 0.754 |
| | 5 | 0.616 | 0.621 | 0.636 | 0.678 | 0.685 |
| | 10 | 0.545 | 0.550 | 0.564 | 0.606 | 0.612 |
| | 15 | 0.517 | 0.522 | 0.535 | 0.574 | 0.575 |
| 0.990 | 2 | 0.716 | 0.719 | 0.729 | 0.758 | 0.782 |
| | 3 | 0.678 | 0.683 | 0.695 | 0.729 | 0.741 |
| | 4 | 0.654 | 0.659 | 0.671 | 0.707 | 0.706 |
| | 5 | 0.634 | 0.640 | 0.653 | 0.690 | 0.682 |
| | 10 | 0.583 | 0.588 | 0.602 | 0.641 | 0.664 |
| | 15 | 0.559 | 0.564 | 0.577 | 0.615 | 0.643 |

References

- [1] ALAM, K. (1970). A two-sample procedure for selecting the population with the largest mean from k normal populations. *Ann. Inst. Statist. Math.* **22**, 127-136.
- [2] ANTLE, C. E., KLIMKO, L. and HARKNESS, W. (1970). Confidence intervals for the parameters of the logistic distribution. *Biometrika* **57**, 397-402.
- [3] BECHHOFFER, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16-39.
- [4] BERKSON, J. (1944). Application of the Logistic Function to Bio-assay. *J. Amer. Statist. Assoc.* **39**, 357-365.
- [5] BERKSON, J. (1951). Why I prefer logits to probits. *Biometrics* **7**, 327-339.
- [6] BERKSON, J. (1957). Tables for the maximum likelihood estimate of the logistic function. *Biometrics* **13**, 28-34.
- [7] BIRNBAUM, A. and DUDMAN, J. (1963). Logistic Order Statistics. *Ann. Math. Statist.* **34**, 653-663.
- [8] BHANDARI, S. K. and CHAUDHURI, A. R. (1987). On two conjectures about two-stage selection problem. (to appear in *Sankhyā Ser. B*) .
- [9] COHEN, D. S. (1959). A Two-Sample Decision Procedure for Ranking Means of Normal Populations with a Common Known Variance. M.S. Thesis, Dept. of Operations Research, Cornell Univ., Ithaca, New York.
- [10] DRAPER, N. R. and TIERNEY, D. E. (1973). Exact Formulas for Additional Terms in Some Important Series Expansions. *Comm. Statist.* **1**, 495-524.
- [11] FIACCO, A. V. and MCCORMICK, G. P. (1968). *Nonlinear Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, Inc., New York.

- [12] GOEL, P. K. (1975). On the Distribution of Standardized Mean of Samples from the Logistic Population. *Sankhyā Ser. B* **2**, 165-172.
- [13] GEORGE, E. O. and MUDHOLKAR, G. S. (1983). On the Convolution of Logistic Random Variables. *Metrika* **30**, 1-13.
- [14] GUMBEL, E. J. (1944). Ranges and midranges. *Ann. Math. Statist.* **15**, 414-422.
- [15] GUMBEL, E. J. and Keeney, R. D. (1950). The extremal quotient. *Ann. Math. Statist.* **21**, 523-538.
- [16] GUPTA, S. S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill.
- [17] GUPTA, S. S. (1962). Life Test Sampling Plans for Normal and Lognormal Distributions. *Technometrics* **4**, 151-175.
- [18] GUPTA, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7**, 225-245.
- [19] GUPTA, S. S. and GNANADESIKAN, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika* **53**, 565-570.
- [20] GUPTA, S. S. and SHAH, B. K. (1965). Exact Moments and Percentage Points of the Order Statistics and the Distribution of the Range from the Logistic Distribution. *Ann. Math. Statist.* **36**, 907-920.
- [21] HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934). *Inequalities*. Cambridge Univ. Press, Cambridge.
- [22] HIRSCHMAN, I. I. and WIDDER, D. V. (1955). *The Convolution Transform*. Princeton, N.J.

- [23] KUESTER, J. L. and MIZE, J. H. (1973). *Optimization Techniques with Fortran*, McGraw-Hill, Inc.
- [24] PEARL, R. and REED, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Nat. Acad. Sci.* **6**, 275-288.
- [25] PLACKETT, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist.* **29**, 131-142.
- [26] PLACKETT, R. L. (1959). The analysis of life test data. *Technometrics* **1**, 9-19.
- [27] SCHAFER, R. E. and SHEFFIELD, T. S. (1973). Inferences on the Parameters of the Logistic Distribution. *Biometrics* **29**, 449-455.
- [28] SCHOENBERG, I. J. (1953). On Pólya frequency functions and their Laplace transformations. *J. d'Annal. Math.* **1**, 331-374.
- [29] TALACKO, J. (1956). Perk's distributions and their role in the theory of Wiener's stochastic variates. *Trabajos de Estadística* **7**, 159-174.
- [30] TAMHANE, A. C. and BECHHOFER, R. E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. *Commun. Statist.-Theor. Meth.* **A6**, 1003-1033.
- [31] TAMHANE, A. C. and BECHHOFER, R. E. (1979). A two-stage minimax procedure with screening for selecting the largest normal mean (II): an improved PCS lower bound and associated tables. *Commun. Statist.-Theor. Meth.* **A8**, 337-358.

AD-A185 614

1a. REPORT SECURITY CLASSIFICATION

2a. SECURITY CLASSIFICATION AUTHORITY
Unclassified

DISTRIBUTION / AVAILABILITY OF REPORT

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

Technical Report #87-39

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION

Purdue University

6b. OFFICE SYMBOL
(if applicable)

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code)

Department of Statistics
West Lafayette, IN 47907

7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

Office of Naval Research

8b. OFFICE SYMBOL
(if applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

N00014-84-C-0167

8c. ADDRESS (City, State, and ZIP Code)

Arlington, VA 22217-5000

10. SOURCE OF FUNDING NUMBERS

| PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT ACCESSION NO. |
|---------------------|-------------|----------|-------------------------|
| | | | |

11. TITLE (Include Security Classification)

An Elimination type two-stage procedure for selecting the population with the largest mean from k logistic populations

12. PERSONAL AUTHOR(S)

Shanti S. Gupta and SangHyun Han

13a. TYPE OF REPORT

Technical

13b. TIME COVERED

FROM _____ TO _____

14. DATE OF REPORT (Year, Month, Day)

August 26, 1987

15. PAGE COUNT

28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

| FIELD | GROUP | SUB-GROUP |
|-------|-------|-----------|
| | | |
| | | |
| | | |

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Two-stage Procedure, Selection Procedure, Largest Mean, Subset Selection, Logistic Populations

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A formula for an approximation to the distribution of the sample means of a logistic population is derived by using the Edgeworth series expansions. Using this approximation, we consider an elimination type two-stage procedure based on the sample means for selecting the population with the largest mean from k logistic populations when the common variance is known. A table of the constants needed to implement this procedure is provided and the efficiency of this procedure relative to the single-stage procedure is investigated.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT

UNCLASSIFIED/UNLIMITED SAME AS RPT DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION

Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL

Shanti S. Gupta

22b. TELEPHONE (Include Area Code)

317-494-6031

22c. OFFICE SYMBOL

END

12-87

DTIC