SELECTION AND RANKING PROCEDURES FOR LOGISTIC POPULATIONS*

by

Shanti S. Gupta**
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

and

SangHyun Han
Korea Military Academy
Wharang Dae
Seoul, KOREA

Technical Report # 91-34C

Department of Statistics
Purdue University

July, 1991

*Research supported in part by NSF Grants DMS–8923071 and DMS–8717799 at Purdue University.

**To be presented as an invited paper by Professor S.S. Gupta at the International Conference on "Order Statistics and Nonparametrics" in Alexandria, Egypt on Sept. 18-20, 1991.

△

SELECTION AND RANKING PROCEDURES FOR LOGISTIC POPULATIONS

by

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, IN 47907-1399

and

SangHyun Han
Korea Military Academy
Wharang Dae
Seoul, KOREA

Abstract

In this paper we investigate the problem of selecting a population associated with the largest location parameter among k logistic populations. We propose and study a single-stage procedure using the indifference zone approach which selects the population associated with the largest sample mean. Also two subset selection rules based on the largest order statistics are investigated; these rules select the population Π_i if and only if $T_i(\underset{\sim}{X}) - \max_j T_j(\underset{\sim}{X}) > d$, where d is an appropriate constant chosen to satisfy the usual probability requirement of a correct selection. In the preceding $T_i(\underset{\sim}{X})$ is a suitable statistic based on the sample $\underset{\sim}{X}$ from ith population. An approximation for the distribution of the sample mean from the logistic population is derived by using Edgeworth series expansions. Using this approximation the procedures are compared when $T_i(\underset{\sim}{X})$ is either a sample mean or a sample median.

Key words: Selection and ranking, Subset selection, Logistic populations, Largest order statistic.

1. Introduction

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 shape of the logistic distribution is similar to the normal distribution. The simple explicit relationship between the logistic random variate, its probability density function (pdf) and its cumulative distribution function (cdf) render much of the analysis of the logistic distribution attractively simple and many authors, for example, Berkson (1951) prefer it to the normal distribution. 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), Tarter and Clark (1965), Gupta and Balakrishnan (1991), Balakrishnan, Gupta and Panchapakesan (1991), and Panchapakesan (1991), investigate inference questions about its parameters.

As might be expected, because of the similarity beetween 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) obtains 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 cdf of the sum of *iid* logistic variates for the sample size $n = 2(1)12, x = 0(0.01)3.99$ and $n = 13(1)15, x = 1.20(0.01)3.99$. He also gives a table of the quantiels for $n = 2(1)15, \alpha = 0.90, 0.95, 0.975, 0.99, 0.995$. George and Mudholkar (1983) obtain 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 $(1 - e^x)^{-k}, k = 1, \ldots, n$, 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 stnadardized Student's t distribution provides a very good approximation for the distribution of a convolution of the *iid* logistic random variables.

This paper considers approximation problems to the distribution and quantiles of a standardized mean of samples from a logistic population by using Edgeworth and Cornish-Fisher series expansions respectively. Tables of the cdf and quantiles are provided and it is shown that these are far better approximations than the Student's t distributions as suggested in George and Mudholkar (1983) and hence these approximations will be used henceforth.

In the rest of this paper a single-stage procedure $\mathcal{P}_1$ using an indifference zone formulation for selecting the best among several logistic populations with unknown means and a common known variance based on sample means is proposed and studied. A table of the smallest sample size n needed to implement $\mathcal{P}_1$ subject to the basic probability requirement is provided.

Two subset selection rules R_1 and R_2 based on sample means and sample medians respectively for selecting the best among several logistic populations are proposed and tables of constants to implement the rules are provided. We also compare the two rules with respect to their performance characteristics.

2. Distribution of logistic sample means

2.1. Logistic distribution

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

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

and the cdf of X is defined by

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

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

The standard logistic distribution with mean zero and variance unity, denoted by $L(0,1)$, has the pdf and cdf defined as

$$f(x) = g[\exp\{-gx\}][1+\exp\{-gx\}]^{-2} \tag{3}$$

and

$$F(x) = [1+\exp\{-gx\}]^{-1} \tag{4}$$

respectively, where $-\infty < x < \infty$. The standard logistic distribution has a shape similar to the standard normal distribution. The curve of the logistic distribution crosses the density curve of the normal between 0.68 and 0.69. The inflection points of the pdf of the standard logistic distribution are $\pm$ 0.53 (approx.).

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

$$f(y) = [\exp\{-y\}][1+\exp\{-y\}]^{-2} \tag{5}$$

and

$$F(y) = [1+\exp\{-y\}]^{-1} \tag{6}$$

respectively, where $-\infty < y < \infty$. (5) may be written in terms of $F(y)$ as

$$f(y) = F(y)(1 - F(y)). \tag{7}$$

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

$$\begin{aligned} M_Y(t) &= \Gamma(1+t)\Gamma(1-t) \\ &= \pi t/\sin(\pi t), \;\; |t| < 1. \end{aligned} \tag{8}$$

We can also express (8) as

$$M_Y(t) = \sum_{j=0}^{\infty}(-1)^{j-1}[2(2^{2j-1}-1)/(2j)!]B_{2j}(\pi t)^{2j}, \tag{9}$$

where B_ν's are Bernoulli numbers defined as

$$x/(\exp(x)-1) = \sum_{\nu=0}^{\infty} B_\nu x^\nu/(\nu!). \tag{10}$$

The relationships between B_ν's are given by

$$\begin{gathered} B_0 = 1 \\ 1 + \binom{k}{1}B_1 + \binom{k}{2}B_2 + \ldots + \binom{k}{k-1}B_{k-1} = 0, k = 1, \ldots, \end{gathered} \tag{11}$$

and hence the first few values of B_ν are

$$\begin{aligned} B_0 &= 1, \\ B_1 &= -1/2, \\ B_2 &= 1/6, \\ B_4 &= -1/30, \\ B_6 &= 1/42, \\ B_8 &= -1/30, \\ B_{10} &= 5/66, \\ &\vdots \\ B_{2m+1} &= 0, m = 1, 2, \ldots \end{aligned} \tag{12}$$

The ν^{th} central moments of Y, denoted by μ_ν, can be obtained as

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

by using (9).

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

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

In terms of the central moments $\mu_\nu(x)$ of X, first few of the ν^{th} cumulants of X, denoted by $K_\nu(x), \nu = 1,2,\ldots$, which are 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, are given by

$$\begin{aligned} K_1(x) &= \mu_1(x) = \mu, \\ K_2(x) &= \mu_2(x) = \sigma^2, \\ K_4(x) &= \mu_4(x) - 3(\mu_2(x))^2 = \tfrac{6}{5}\sigma^4, \\ K_6(x) &= \mu_6(x) - 15\mu_2(x)\mu_4(x) + 30(\mu_2(x))^3 = \tfrac{48}{7}\sigma^6, \\ K_8(x) &= \mu_8(x) - 28\mu_2(x)\mu_6(x) - 35(\mu_4(x))^2 + 420(\mu_2(x))^2\mu_4(x) \\ &\quad - 630(\mu_2(x))^4 = \tfrac{432}{5}\sigma^8, \\ K_{10}(x) &= \mu_{10}(x) - 45\mu_2(x)\mu_8(x) - 210\mu_4(x)\mu_6(x) \\ &\quad + 1260(\mu_2(x))^2\mu_6(x) + 3150\mu_2(x)(\mu_4(x))^2 \\ &\quad - 18900(\mu_2(x))^3\mu_4(x) + 22680(\mu_2(x))^5 = \tfrac{145152}{77}\sigma^{10}, \\ &\vdots \\ K_{2j+1}(x) &= 0, j = 1,2,\ldots. \end{aligned}$$

The first few relative cumulants of $X, \lambda_\nu(x)$, where $\lambda_\nu(x)$ is defined as

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

are given by

$$\begin{aligned}
\lambda_1(x) &= \mu/\sigma, \\
\lambda_2(x) &= 1, \\
\lambda_4(x) &= 6/5, \\
\lambda_6(x) &= 48/7, \\
\lambda_8(x) &= 432/5, \\
\lambda_{10}(x) &= 145152/77, \\
&\vdots \\
\lambda_{2j+1}(x) &= 0, j = 1, 2, \ldots. \qquad (13)
\end{aligned}$$

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

Let $X_1, X_2, \ldots, X_n$ 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

$$\begin{aligned}
Z &= \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} (X_i - \mu) \\
&= \frac{\sqrt{n}}{\sigma} (\overline{X} - \mu), \qquad (14)
\end{aligned}$$

where $\overline{X} = \frac{1}{n} \Sigma_{i=1}^{n} X_i$ 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 (13), the Edgeworth series expansions of the $f_n(z)$ and $F_n(z)$ correct to order $n^{-\nu/2}, \nu = 4, 6, 8$, are given by

$$\begin{aligned}
&f_n(z, \nu = 4) \\
&= \phi(z)\{1 + [(\tfrac{1}{4!})(\tfrac{6}{5})H_4(z)]n^{-1} \\
&\quad + [(\tfrac{1}{6!})(\tfrac{48}{7})H_6(z) + (\tfrac{35}{8!})(\tfrac{6}{5})^2 H_8(z)]n^{-2}\} + O(n^{-5/2}), \\
&F_n(z, \nu = 4) \\
&= \Phi(z) - \phi(z)\{[(\tfrac{1}{4!})(\tfrac{6}{5})H_3(z)]n^{-1} \\
&\quad + [(\tfrac{1}{6!})(\tfrac{48}{7})H_5(z) + (\tfrac{35}{8!})(\tfrac{6}{5})^2 H_7(z)]n^{-2}\} + O(n^{-5/2}), \\
&f_n(z, \nu = 6) \\
&= f_n(z, \nu = 4) + \phi(z)[(\tfrac{1}{8!})(\tfrac{432}{5})H_8(z) \\
&\quad + (\tfrac{210}{10!})(\tfrac{48}{7})(\tfrac{6}{5})H_{10}(z) + (\tfrac{5775}{12!})(\tfrac{6}{5})^3 H_{12}(z)]n^{-3} + O(n^{-7/2}), \qquad (15) \\
&F_n(z, \nu = 6) \\
&= F_n(z, \nu = 4) - \phi(z)[(\tfrac{1}{8!})(\tfrac{432}{5})H_7(z) \\
&\quad + (\tfrac{210}{10!})(\tfrac{48}{7})(\tfrac{6}{5})H_9(z) + (\tfrac{5775}{12!})(\tfrac{6}{5})^3 H_{11}(z)]n^{-3} + O(n^{-7/2}), \qquad (16) \\
&f_n(z, \nu = 8) \\
&= f_n(z, \nu = 6) + \phi(z)[(\tfrac{1}{10!})(\tfrac{145152}{77})H_{10}(z) \\
&\quad + (\tfrac{495}{12!})(\tfrac{432}{5})(\tfrac{6}{5})H_{12}(z) + (\tfrac{462}{12!})(\tfrac{48}{7})^2 H_{12}(z) \\
&\quad + (\tfrac{105105}{14!})(\tfrac{6}{5})^2(\tfrac{48}{7})H_{14}(z) + (\tfrac{2627625}{16!})(\tfrac{6}{5})^4 H_{16}(z)]n^{-4} + O(n^{-9/2})
\end{aligned}$$

and

$$
\begin{aligned}
&F_n(z,\nu=8)\\
&= F_n(z,\nu=6) - \phi(z)[(\tfrac{1}{10!})(\tfrac{145152}{77})H_9(z)\\
&+ (\tfrac{495}{12!})(\tfrac{432}{5})(\tfrac{6}{5})H_{11}(z) + (\tfrac{462}{12!})(\tfrac{48}{7})^2 H_{11}(z)\\
&+ (\tfrac{105105}{14!})(\tfrac{6}{5})^2(\tfrac{48}{7})H_{13}(z) + (\tfrac{2627625}{16!})(\tfrac{6}{5})^4 H_{15}(z)]n^{-4} + O(n^{-9/2}),
\end{aligned}
$$

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

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

The first thirty Hermite polynomials which follow the recurrence relation

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

are given in Table III in Draper and Tierney (1973).

Table 1 contains the values of the cdf of the standardized mean of samples of size n from a logistic population with mean μ and variance σ^2 for $n = 3, 10, 15$ and $z = 0.00(0.10)1.00(0.20)3.00(0.40)3.80$ using the Edgeworth series expansion correct to order n^{-3} given in (16). Entries in the tables were computed by using double-precision arithmetic on a Vax-11/780.

2.3. Cornish-Fisher series expansions for the quantiles of the mean of samples from a logistic population

The representation of a quantile of one distribution in terms of the corresponding quantile of another is widely used as a technique for obtaining approximations for the percentage points. One of the most popular of such quantile representations was introduced by Cornish and Fisher (1937) and later reformulated by Fisher and Cornish (1960) and is referred to as the Cornish-Fisher expansion.

By means of formal substitutions, Taylor expansions and identification of powers of n, the Cornish-Fisher expansion of a quantile z of $F_n(z)$ which is the cdf of the standardized mean of samples of size n from $L(\mu, \sigma^2)$, in terms of the corresponding normal quantile y, is of the form

$$z = y + \sum_{j=1}^{\nu} Q_j(y) n^{-j/2} + O(n^{-(\nu+1)/2}),$$

where $Q_j(y)$'s are polynomials of y, which are obtained up to $\nu = 8$ in Draper and Tierney (1973).

Using $Q_j(y)$ from Table VII of Draper and Tierney (1973) and $\lambda_j(x)$ in (13), we obtain the Cornish-Fisher series expansions for the quantiles z of $F_n(z)$ up to order $\nu = 4, 6, 8$ as follows:

$$\begin{aligned}
z(\nu = 4) =& y + [(\tfrac{1}{4!})(\tfrac{6}{5})(y^3 - 3y)]n^{-1} \\
& + [(\tfrac{1}{6!})(\tfrac{48}{7})(y^5 - 10y^3 + 15y) \\
& + (\tfrac{35}{8!})(\tfrac{6}{5})^2(-9y^5 + 72y^3 - 87y)]n^{-2} + O(n^{-5/2}), \\
z(\nu = 6) =& z(\nu = 4) + [(\tfrac{1}{8!})(\tfrac{432}{5})(y^7 - 21y^5 + 105y^3 - 105y) \\
& + (\tfrac{210}{10!})(\tfrac{48}{7})(\tfrac{6}{5})(-15y^7 + 255y^5 - 1035y^3 + 855y) \\
& + (\tfrac{5775}{12!})(\tfrac{6}{5})^3(243y^7 - 3537y^5 + 12177y^3 - 8667y)]n^{-3} \\
& + O(n^{-7/2})
\end{aligned} \tag{17}$$

and

$$\begin{aligned}
& z(\nu = 8) \\
&= z(\nu = 6) + [(\tfrac{1}{10!})(\tfrac{145152}{77})(y^9 - 36y^7 + 378y^5 - 1260y^3 + 945y) \\
&\quad + (\tfrac{495}{12!})(\tfrac{432}{5})(\tfrac{6}{5})(-21y^9 + 630y^7 - 5502y^5 + 15330y^3 - 9765y) \\
&\quad + (\tfrac{462}{12!})(\tfrac{48}{7})^2(-25y^9 + 700y^7 - 5850y^5 + 15900y^3 - 9945y) \\
&\quad + (\tfrac{105105}{14!})(\tfrac{6}{5})^2(\tfrac{48}{7})(495y^9 - 12510y^7 + 92370y^5 - 219810y^3 + 121455y) \\
&\quad + (\tfrac{2627625}{16!})(\tfrac{6}{5})^4(-11583y^9 + 259848y^7 - 1686906y^5 + 3539376y^3 \\
&\qquad - 1743471y)]n^{-4} + O(n^{-9/2}).
\end{aligned}$$

Table 2 provides the quantiles of the distribution of the standardized mean of samples from the logistic population for sample sizes $n = 3, 5, 10, 15, 25$ and probability levels $\alpha = 0.900, 0.950, 0.975, 0.990, 0.995$ using the Cornish-Fisher series expansions correct to order $\nu = 8$. Entries of the table were calculated by using double-precision arithmetic on a Vax-11/780.

2.4. Legitimacy of using the Edgeworth and Cornish-Fisher series expansions

Noting the similarity of the distribution of Z in (14), the standardized mean of samples from $L(\mu, \sigma^2)$, to the normal distribution in shape except its relatively longer tails, George and Mudholkar (1983) compare the three approximations, that is, the standard normal distribution, the Edgeworth series expansion correct to order n^{-1} and the standardized Student's t distribution to the exact distribution of Z. In using the standardized Student's t distribution, they use the degree of freedom $\xi = 5n+4$ which can be obtained by equating the coefficients of kurtosis. They show that the Student's t distribution provides a very good approximation.

We show here that the Edgeworth and Cornish-Fisher series expansions correct to order n^{-3}, which are given in the (16) and (17) respectively, are far better approximations than even the Student's t distribution in George and Mudholkar (1983).

Table 3 illustrates the quality of the four approximations. In Table 3 the four approximations, that is, the standard normal, the Edgeworth series expansion correct to n^{-1}, the standardized Student's t and the Edgeworth series expansion correct to order n^{-3} are compared to the exact distribution given in Goel (1975). The approximation using the Edgeworth series expansion correct to order n^{-3} appears to be superior to the other three by noting that the maximum error is about 0.0001 as shown in the last column of Table 3. The exact quantiles for $n = 2, 3, \ldots, 15$ tabled by Goel (1975) were compared with the corresponding approximations obtained form the Student' t distribution and the Cornish-Fisher series expansion correct to order n^{-3}, and it was found that for sample size 7 or more the Edgeworth series expansion correct to order n^{-3} provides an excellent approximation for the standardized mean of samples from the logistic distribution. Consequently, we will use the Edgeworth series expansion correct to order n^{-3} as an approximation to the distribution of the standardized mean of the samples from the logistic distribution henceforth.

3. A single-stage procedure $\mathcal{P}_1$ for selecting the population with the largest mean from k logistic populations

Bechhofer (1954), in introducing the indifference zone formulation, considered the problem of ranking means of normal populations with a common known variance. Here we consider a single-stage procedure using an indifference zone approach for selecting the population with the largest mean from k logistic populations when they have a common known variance.

3.1. Statement of the problem

Let $\pi_1, \ldots, \pi_k$ be k independent logistic populations with unknown means μ_i and a common known variance σ^2. let $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ be the ranked μ_i. We assume that it is not known which population is associated with $\mu_{[i]}, i = 1, \ldots, k$. We further assume that a population is characterized by its population mean and the 'best' population is the one having the largest mean.

Our procedure will be based on the sample means. Let $\overline{X}_i, i = 1, \ldots, k$, denote the means of independent samples of size n from i^{th} population. The sample mean associated with population having population mean $\mu_{[i]}$ will be denoted by $\overline{X}_{(i)}$, that is, the expected value of $\overline{X}_{(i)}$ is $\mu_{[i]}$. Let $\overline{X}_{[1]} \leq \ldots \leq \overline{X}_{[k]}$ be ranked $\overline{X}_i$. If $\overline{X}_i = \overline{X}_j$ for $i \neq j$, due to the limitations of the measuring instrument, the tied means should be 'ranked' using a randomized procedure which assigns equal probability to each ordering.

Assuming that the goal of the experimenter is to select the best among the k populations, we propose a single-stage procedure $\mathcal{P}_1$ as follows.

Procedure $\mathcal{P}_1$; Take n observations from the i^{th} population for each $i = 1, \ldots, k$. Compute the k sample means $\overline{X}_1, \ldots, \overline{X}_k$. Select the population associated with $\overline{X}_{[k]}$ as the best one.

Defining the event of the experimenter's selection of the best population with $\mathcal{P}_1$ as $[CS|\mathcal{P}_1]$, the probability of a correct selection with the procedure $\mathcal{P}_1, P\{CS|\mathcal{P}_1\}$ can be written as

$$\begin{aligned}
P\{CS|\mathcal{P}_1\} &= P_{\vec{\mu}}[\overline{X}_{(k)} \geq \overline{X}_{(j)}, j = 1, \ldots, k-1] \\
&= P_{\vec{\mu}}[(\sqrt{n}/\sigma)(\overline{X}_{(j)} - \mu_{[j]}) \leq (\sqrt{n}/\sigma)(\overline{X}_{(k)} - \mu_{[k]}) \\
&\quad + (\sqrt{n}/\sigma)(\mu_{[k]} - \mu_{[j]}), j = 1, \ldots, k-1] \\
&= \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F_n(z + (\sqrt{n}/\sigma)(\mu_{[k]} - \mu_{[j]}))dF_n(z), \qquad (18)
\end{aligned}$$

where $F_n(z)$ is the cdf of the standardized mean of samples from a logistic population.

For the fixed values of the μ_i and σ^2 the probability of a correct selection will depend only on the sample size n. We propose to design the experiment in such a way, that is, choose the n in such a way that under specified conditions the probability of a correct selection with procedure $\mathcal{P}_1$ will be equal to or greater than some preassigned value P^*.

3.2. Determination of the sample sizes

Now for the problem to be meaningful P^* lies between $1/k$ and 1. Since the true values of the μ_i are not known, we need the probability of a correct selection to be at least P^* whatever be the values of the μ_i. Thus we are interested in the configuration of the μ_i for which the probability in (18) is a minimum. Such a configuration will be called a least favorable configuration (LFC). It is obvious that the LFC is given by $\mu_{[1]} = \ldots = \mu_{[k]}$. But unfortunately the minimum value of the probability in this LFC case is $1/k$. So we cannot achieve the probability requirement whatever be the sample size unless some modification is made in the probability requirement.

A natural modification is to insist on the minimum probability P^* of selecting the best population whenever the best is sufficiently far apart from the next best. In other words, the experimenter specifies a positive constant δ and requires that the probability of selecting the best population is at least P^* whenever $(\mu_{[k]} - \mu_{[k-1]}) \geq \delta$. The specification of δ provides a partition of the parameter space Ω where

$$\Omega = \{\vec{\mu} = (\mu_1, \ldots, \mu_k) | -\infty < \mu_i < \infty, i = 1, \ldots, k\} \tag{19}$$

into two parts, namely $\Omega(\delta)$ where

$$\Omega(\delta) = \{\vec{\mu} \in \Omega | (\mu_{[k]} - \mu_{[k-1]}) \geq \delta\} \tag{20}$$

and the compliment $\Omega^c(\delta)$ of $\Omega(\delta)$. The minimization of the probability of selecting the best population is over $\Omega(\delta)$. For an obvious reason, $\Omega^c(\delta)$ was called the indifference zone by Bechhofer (1954). Subsequent authors have termed $\Omega(\delta)$ the preference zone.

It is now easy to see that the LFC in $\Omega(\delta)$ is given by

$$\Omega^0(\delta) = \{\vec{\mu} \in \Omega(\delta) | \mu_{[1]} = \mu_{[k-1]} = \mu_{[k]} - \delta\} \tag{21}$$

and the minimum sample size required is the smallest integer n for which

$$\inf_{\vec{\mu}\in\Omega(\delta)} P_{\vec{\mu}}[CS|\mathcal{P}_1] = \int_{-\infty}^{\infty} (F_n(z+(\sqrt{n}/\sigma)\delta))^{k-1} dF_n(z) \geq P^*. \tag{22}$$

A table has been prepared to assist the experimenter in designing the experiments to meet the above goal. Table 4 is to be used for designing experiments involving k logistic populations to decide which has the largest population mean. The table provides the estimates $\hat{n}$ of the values of minimum sample size n associated with the probability $P^* = 0.75, 0.90, 0.95, 0.99$ for $k = 2, 3, 4, 5, 10, 15$, and $\delta/\sigma = 0.5, 1.0, 2.0$. These were computed by setting the left hand side of (22) equal to P^*. The minimum sample size n can be obtained by $n = [\hat{n}+1]$ where $[t]$ denotes the greatest integer which is less than t. All computations were carried out in double-precision arithmetic on a Vax-11/780.

4. Subset selection procedures

Gupta (1956, 1965) introduced a subset selection formulation as a multiple decision problem, where the investigation was carried out for the case of normal means. Here we consider the subset selection rules for selecting the population with the largest mean from k logistic populations. We propose two subset selection rules R_1 and R_2 based on sample means and sample medians respectively, provide tables for implementing these rules, consider the performance characteristics of each rule, and we compare the two rules to each other.

4.1. Statement of the problem

Let $\pi_i, i = 1, \ldots, k$, be k independent logistic populations with unknown means μ_i and a common known variance σ^2. Let $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$ be ranked μ_i and $\pi_{(i)}$ be the population with mean $\mu_{[i]}$. We assume that it is not known which population is associated with $\mu_{[i]}, i = 1, \ldots, k$. We further assume that a population is characterized by its population mean and the 'best' population is the one having the largest mean, that is, $\pi_{(k)}$.

Let $X_{ij}, j = 1, \ldots, n$, denote a random sample from $\pi_i, i = 1, \ldots, k$, where the observations within and between populations are all independent. Let $\overline{X}_i$ and $X_{i:l}, i = 1, \ldots, k, n = 2l-1$, denote the means and medians of samples of size n from π_i respectively. The sample mean and the sample median associated with the population having population mean $\mu_{[i]}$ will be denoted by $\overline{X}_{(i)}$ and $X_{(i):l}, i = 1, \ldots, k$, respectively. Let $\overline{X}_{[1]} \leq \ldots \leq \overline{X}_{[k]}$ and $X_{[1]:l} \leq \ldots \leq X_{[k]:l}$ be ranked $\overline{X}_i$ and $X_{i:l}$ respectively.

The goal is to select a small but non-empty subset S of the k populations so that the selected subset includes with a high probability P^* the 'best' population. The size of the selected subset S is an integer-valued random variable taking on values $1, \ldots, k$.

Let us define the two subset selection ruels R_1 and R_2 based on the sample means and sample medians, respectively, as follows;

<u>*Rule R_1*</u>*: select π_i iff*

$$\overline{X}_i \geq \max_{1 \leq j \leq k} \overline{X}_j - h_1 \sigma / \sqrt{n}, i = 1, \ldots, k, \tag{23}$$

and

<u>*Rule R_2*</u>*: select π_i iff*

$$X_{i:l} \geq \max_{1 \leq j \leq k} X_{j:l} - h_2 \sigma / \sqrt{n}, i = 1, \ldots, k, \tag{24}$$

where h_1 and h_2 are nonnegative constants.

By defining the events $[CS|R_i], i = 1, 2$, as selections of any non-empty subset of k populations which includes the best population using $R_i, i = 1, 2$, respectively, it is required that for any $\vec{\mu} \in \Omega$

$$P_{\vec{\mu}}[CS|R_i] \geq P^*, \tag{25}$$

where $P^* \in (1/k, 1)$ and Ω is the parameter space given by (19).

The requirement of (25) will be called as the basic probability requirement or the P^*-condition.

Remark 2.1 *Lorenzen and McDonald (1981) used a subset selection rule R based on sample medians defined as*

<u>*Rule R*</u>*: select π_i iff*

$$X_{i:l} \geq \max_{1 \leq j \leq k} X_{j:l} - d, \; d \geq 0, i = 1, \ldots, k,$$

where $X_{i:l}$ is defined as above. Here we use R_2 instead of R only for the purpose of comparing R_1 to R_2 easily. Basically the rule R_2 is the same as Lorenzen and McDonald's rule R.

4.2 Probability of a correct selection

- Probability of a correct selection for the means rule R_1

Using (23) we can write the probability of a correct selection for the rule R_1 as follows. For $\vec{\mu} \in \Omega$,

$$\begin{aligned} & P_{\vec{\mu}}[CS|R_1] \\ & \quad = P_{\vec{\mu}}[\overline{X}_{(k)} \geq \overline{X}_{(j)} - h_1\sigma/\sqrt{n}, \quad \forall\, j = 1, \ldots, k-1] \\ & \quad = P_{\vec{\mu}}[(\sqrt{n}/\sigma)(\overline{X}_{(j)} - \mu_{[j]}) \leq (\sqrt{n}/\sigma)(\overline{X}_{(k)} - \mu_{[k]}) + h_1 \\ & \qquad + (\sqrt{n}/\sigma)(\mu_{[k]} - \mu_{[j]}); \quad \forall\, j = 1, \ldots, k-1] \\ & \quad = \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F_n(z + h_1 + (\sqrt{n}/\sigma)(\mu_{[k]} - \mu_{[j]}))dF_n(z). \end{aligned} \tag{26}$$

We see from (26) that the infimum over the parameter space of the probability of a correct selection for the rule R_1 takes place when $\mu_1 = \ldots = \mu_k$ and so

$$\inf_{\vec{\mu} \in \Omega} P_{\vec{\mu}}[CS|R_1] = \int_{-\infty}^{\infty} \{F_n(z + h_1)\}^{k-1} dF_n(z). \tag{27}$$

That is, the LFC for the rule R_1 is Ω^0 where

$$\Omega^0 = \{\vec{\mu} \in \Omega | \mu_1 = \ldots = \mu_k = \mu\} \tag{28}$$

and the $P_{\vec{\mu}}[CS|R_1]$ in the LFC does not depend on this common μ. Hence, if we choose h_1 to satisfy

$$\int_{-\infty}^{\infty} \{F_n(z + h_1)\}^{k-1} dF_n(z) = P^* \tag{29}$$

then we have determined the smallest h_1 for which

$$\inf_{\vec{\mu} \in \Omega} P_{\vec{\mu}}[CS|R_1] \geq P^*. \tag{30}$$

It should be noted that $h_1 = h_1(n, k, P^*)$ depends on n as well as k and P^*.

Table 5 gives the values of $h_1 = h_1(n, k, P^*)$ which satisfy (29) for $n = 6(1)10, k = 2(1)10$ and $P^* = 0.75, 0.90, 0.95, 0.975, 0.99$. We use the Edgeworth series expansions correct to order n^{-3} for $F_n(x)$ and $f_n(x)$, the Gauss-Hermite quadrature algorithm with sixty nodes for the evaluation of the integrals and the modified regular falsi algorithm for solving the non-linear equation.

- Probability of a correct selection for the medians rule R_2

Let $Z_{i:1}, \ldots, Z_{i:n}$ be a random sample of size n, where n is an odd integer, drawn from the i^{th} standard logistic population. Then it is well known that the sample median, denoted by $Z_{i:l}, (n = 2l - 1)$, has the pdf

$$g_n(z) = \frac{\Gamma(2l)}{\Gamma(l)^2}[F(z)]^{l-1}[1 - F(z)]^{l-1} f(z)$$

and the cdf

$$G_n(z) = I\{F(z); l, l\}, \tag{31}$$

where $f(z)$ and $F(z)$ are the pdf and cdf of the standard logistic population given by (3) and (4) respectively and $I\{y; a, b\}$ is the incomplete beta function with parameters a and b, which is given by

$$I\{y; a, b\} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^y w^{a-1}(1-w)^{b-1} dw. \tag{32}$$

Now the probability of a correct selection for the medians rule R_2 can be written as follows. For $\vec{\mu} \in \Omega$

$$\begin{aligned} P_{\vec{\mu}}[CS|R_2] &= P_{\vec{\mu}}[X_{(k):l} \geq X_{(j):l} - h_2\sigma/\sqrt{n}, \quad \forall\, j = 1, \ldots, k-1] \\ &= P_{\vec{\mu}}[Z_{(j):l} \leq Z_{(k):l} + h_2/\sqrt{n} \\ &\qquad + (\mu_{[k]} - \mu_{[j]})/\sigma, \quad \forall\, j = 1, \ldots, k-1] \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{k-1} G_n(t + h_2/\sqrt{n} + (\mu_{[k]} - \mu_{[j]})/\sigma) dG_n(t). \end{aligned} \tag{33}$$

We see that the infimum over Ω of the probability of a correct selection for the rule R_2 takes place when $\mu_1 = \ldots = \mu_k = \mu$ and so

$$\inf_{\vec{\mu} \in \Omega} P_{\vec{\mu}}[CS|R_2] = \int_{-\infty}^{\infty} \{G_n(t + h_2/\sqrt{n})\}^{k-1} dG_n(t). \tag{34}$$

Hence, if we choose h_2 to satisfy

$$\int_{-\infty}^{\infty} \{G_n(t + h_2/\sqrt{n})\}^{k-1} dG_n(t) = P^*, \tag{35}$$

then we have determined the smallest h_2 for which

$$\inf_{\vec{\mu}\in\Omega} P_{\vec{\mu}}[CS|R_2] \geq P^*. \tag{36}$$

The values of $h_2/\sqrt{n} = h_2(n,k,P^*)/\sqrt{n}$ which satisfy (35) for $n = 1(2)19, k = 2(1)10$ and $P^* = 0.75, 0.90, 0.95, 0.975, 0.99$ were given in Table I of Lorenzen and McDonald (1981).

4.3. Performance characteristics

In this section some performance characteristics of the subset selection procedures R_1 and R_2 are studied.

Let$P_{\vec{\mu}}[\pi_{(i)}|R_j], i = 1,\ldots,k, j = 1,2$, denote the probabilities of including in the subset the population $\pi_{(i)}$, that is, the i^{th} ranked population, using the rule R_j for the $\vec{\mu} \in \Omega$, then for $i = 1,\ldots,k$,

$$\begin{aligned} P_{\vec{\mu}}[\pi_{(i)}|R_1] &= P_{\vec{\mu}}[\overline{X}_{(i)} \geq \max\nolimits_{1\leq j\leq k}\overline{X}_j - h_1\sigma/\sqrt{n}, h_1 \geq 0] \\ &= \int_{-\infty}^{\infty} \prod_{\substack{j=1\\ j\neq i}}^{k} F_n(t + h_1 + (\sqrt{n}/\sigma)(\mu_{[i]} - \mu_{[j]}))dF_n(t), \end{aligned} \tag{37}$$

and

$$\begin{aligned} P_{\vec{\mu}}[\pi_{(i)}|R_2] &= P_{\vec{\mu}}[X_{(i):l} \geq \max\nolimits_{1\leq j\leq k}X_{(j):l} - h_2\sigma/\sqrt{n}, h_2 \geq 0] \\ &= \int_{-\infty}^{\infty} \prod_{\substack{j=1\\ j\neq i}}^{k} G_n(t + h_2/\sqrt{n} + (\mu_{[i]} - \mu_{[j]})/\sigma)dG_n(t). \end{aligned} \tag{38}$$

It is easy to see that the expected sizes of the selected subset using the rule R_j for $\vec{\mu} \in \Omega$, denoted by $E_{\vec{\mu}}[S|R_j], j = 1,2$, are given as follows:

$$E_{\vec{\mu}}[S|R_j] = \sum_{i=1}^{k} P_{\vec{\mu}}[\pi_{(i)}|R_j]. \tag{39}$$

Consistent with the basic probability requirement, we would like the size of the selected subset to be small.

The expected numbers of non-best populations selected by rule R_j for $\vec{\mu} \in \Omega$, denoted by $E_{\vec{\mu}}[S^*|R_j], j = 1,2$, are defined as

$$E_{\vec{\mu}}[S^*|R_j] = \sum_{i=1}^{k-1} P_{\vec{\mu}}[\pi_{(i)}|R_j] \tag{40}$$

and also we would like the value of the $E_{\vec{\mu}}[S^*|R_j]$ to be small.

In using the rule $R_j, j = 1, 2$, the ranks of the selected populations are random variables and one may want to evaluate the expected sum of ranks of the selected populations. Let the population with parameter $\mu_{[i]}$ be assigned rank $i, i = 1, \ldots, k$. Then the expected sums of ranks of the selected populations by rule R_j for $\vec{\mu} \in \Omega$, denoted by $E_{\vec{\mu}}[SR|R_j], j = 1, 2$, are

$$E_{\vec{\mu}}[SR|R_j] = \sum_{i=1}^{k} iP_{\vec{\mu}}[\pi_{(i)}|R_j]. \tag{41}$$

For given $\vec{\mu} \in \Omega$, the expected proportions of the selected populations by the rule R_j, denoted by $E_{\vec{\mu}}[P|R_j], j = 1, 2$, are given by

$$E_{\vec{\mu}}[P|R_j] = E_{\vec{\mu}}[S|R_j]/k. \tag{42}$$

Since the values of $P_{\vec{\mu}}[\pi_{(i)}|R_j], j = 1, 2$, depend on $\vec{\mu} \in \Omega$, we consider them for the special case, namely the slippage configuration.

For the slippage configuration, we assume that the unknown means of k populations are $\mu_{[j]} = \mu, j = 1, \ldots, k-1$, and $\mu_{[k]} = \mu + \delta\sigma$ for $\delta > 0$. Then the probabilities of including in the selected subset the population $\pi_{(i)}$ using the rule R_j, denoted by $P_{sp}[\pi_{(i)}|R_j], j = 1, 2$, are given by

$$\begin{aligned} &P_{sp}[\pi_{(i)}|R_1] \\ &\quad = \int_{-\infty}^{\infty} \{F_n(t + h_1)\}^{k-2} F_n(t + h_1 - \delta\sqrt{n}) dF_n(t), i = 1, \ldots, k-1 \\ &\quad P_{sp}[\pi_{(k)}|R_1] = \int_{-\infty}^{\infty} \{F_n(t + h_1 + \delta\sqrt{n})\}^{k-1} dF_n(t), \\ &P_{sp}[\pi_{(i)}|R_2] \\ &\quad = \int_{-\infty}^{\infty} \{G_n(t + h_2/\sqrt{n})\}^{k-2} G_n(t + h_1/\sqrt{n} - \delta) dG_n(t), i = 1, \ldots, k-1, \end{aligned}$$

and

$$P_{sp}[\pi_{(k)}|R_2] = \int_{-\infty}^{\infty} \{G_n(t + h_2/\sqrt{n} + \delta)\}^{k-1} dG_n(t).$$

Now we can compute the performance characteristics $E_{\vec{\mu}}[S|R_j], E_{\vec{\mu}}[S^*|R_j], E_{\vec{\mu}}[SR|R_j]$ and $E_{\vec{\mu}}[P|R_j]$ for the slippage configurations by substituting $P_{sp}[\pi_{(i)}|R_j]$ for $P_{\vec{\mu}}[\pi_{(i)}|R_j]$ in (39), (40), (41) and (42) respectively.

Table 6 gives the values of the performance characteristics of the mean rule R_1 and Table 7 gives the same values of the medians rule R_2 for the slippage configuration for the given values of $k = 2, 3, 5, 10, P^* = 0.90, n = 3$ and $\sqrt{n}\delta = 0.5, 1.0, 2.0, 3.0, 5.0$.

For instance, from Table 6 for $P^* = 0.90, n = 3, k = 5$ and $\delta\sqrt{n} = 1.5$, the probability of a correct selection by using the means rule R_1 is 0.991. The expected size of the selected subset is 4.006 and the expected number of the non-best populations selected is 3.015. The expected sum of the ranks in the selected subset is 12.49 and the expected proportion of the selected population is 0.801. It should be noted that the expected sum of ranks by itself is not a good criterion of the performance of a selection rule. It should be looked at together with the expected values of S and S^* to make a more meaningful performance characteristic.

Note that, for both rules R_1 and R_2 and for the fixed values of P^*, n, k and $i = 1, 2, \ldots, k-1(k)$, the probability of selecting the i^{th} ranked population in the slippage configuration can be proved to be monotonically decreasing (increasing) with $\delta\sqrt{n}$ and hence monotonically decreasing (increasing) with δ and n separately. Also for $i = 1(k)$, the probability of selecting the i^{th} ranked population in the equally spaced configuration can be proved to be monotonically decreasing (increasing) with $\delta\sqrt{n}$. A look at the table values seems to indicate that, for both rules R_1 and R_2 and for the fixed values of P^*, n, k and $i = 2, \ldots, k-1$, the probability of selecting the i^{th} ranked population in the equally spaced configuration is also monotonically decreasing with $\delta\sqrt{n}$. For fixed P^*, i, n and $\delta\sqrt{n}$, the probability of selecting the i^{th} ranked population is monotonically decreasing with the values of k for all $i, i = 1, \ldots, k$.

4.4. Comparison between the means rule R_1 and the medians rule R_2

In this section we compare the efficiency of the means rule R_1 to that of the medians rule R_2. Lorenzen and McDonald (1981) have studied the problem of large sample comparisons between the two rules R_1 and R_2. They computed the asymptotic relative efficiency (ARE) of R_1 relative to R_2 defined by, for $\epsilon \in (0, 1)$ and $\vec{\mu} \in \Omega$,

$$ARE(R_1, R_2; \vec{\mu}) = \lim_{\epsilon \downarrow 0} \frac{N_{R_1}}{N_{R_2}},$$

where $N_{R_j}, j = 1, 2$, are the numbers of observations needed so that

$$\inf P_{\vec{\mu}}[CS|R_j] = P^*$$

and

$$E_{\vec{\mu}}[S^*|R_j] = \epsilon$$

by assuming a slippage configuration, that is,

$$\mu_{[1]} = \ldots = \mu_{[k-1]} = 0, \mu_{[k]} = \delta > 0.$$

Their value of the $ARE(R_1, R_2; \vec{\mu})$ is 0.822. Thus, under a slippage configuration, asymptotically the means procedure requires about 82% of the sample size required by the medians rule to achieve the same expected number of non-best populations in the selected subset.

Now we consider the small sample comparisons between the rules R_1 and R_2 by using the performance characteristics of each rule given in the previous section. In Table 8, we compute the values of the probability of a correct selection ($P(CS)$), the expected sizes of the selected subset ($E(S)$), the expected numbers of non-best populations in the selected subset ($E(S^*)$), the expected sums of the ranks of the populations selected in the subset ($E(SR)$) and the expected proportions of the populations selected in the subset ($E(P)$) for each rule R_1 and R_2 and the ratio of those values of the rules when the unknown means have the slippage configuration for the selected values of $P^* = 0.90, 0.95, n = 3, 5, k = 4$, and $\delta\sqrt{n} = 1.5, 3.0$.

It was found that

- $P(CS|R_1)/P(CS|R_2) \geq 0.991$ for all cases, the values of $P(CS)$'s are not much different for all cases.
- $E(S|R_1)/E(S|R_2) \leq 1$ for all cases, the values of $E(S), E(S^*), E(SR)$ and $E(P)$ for the rule R_1 are less than or equal to the same values for the rule R_2 for all cases.
- The values of the ratio of the rules R_1 and R_2 for all characteristics are decreasing as the values of n are increasing.

Hence, as expected, the means rule R_1 is definitely better than the medians rule R_2 in the sense of their performance characteristics and the performance of the rule R_1 relative to the rule R_2 improves as sample sizes increase.

References

Antle, C.E., Klimko, L., and Harkness, W. (1970). Confidence intervals for the parameters of the logistic distribution. *Biometrika*, **57**, 397–402.

Balakrishnan, N., Gupta, S.S., and Panchapakesan, S. (1991). Estimation of the Mean and Standard Deviation of the Logistic Distribution Based on Multiply Type-II Censored Samples. Technical Report #91-36C, Department of Statistics, Purdue University.

Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics*, **25**, 16–39.

Berkson, J. (1944). Application of the Logistic Function to Bio-assay. *Journal of the American Statistical Association*, **39**, 357–365.

Berkson, J. (1951). Why I prefer logits to probits. *Biometrics*, **7**, 327–339.

Berkson, J. (1957). Tables for the maximum likelihood estimate of the logistic function. *Biometrics*, **13**, 28–34.

Birnbaum, A. and Dudman, J. (1963). Logistic Order Statistics. *Annals of Mathematical Statistics*, **34**, 658–663.

Cornish, E.A. and Fisher, R.A. (1937). Moment and cumulants in the specification of distribution. *Rev. de l'Inst. Int. de Stat.*, **5**, 307–320.

Draper, N.R. and Tierney, D.E. (1973). Exact formulas for additional terms in some important series expansions. *Communications in Statistics*, **1**, 495-524.

Fisher, R.A. and Cornish E.A. (1960). The percentile points of distributions having known cumulants. *Technometrics*, **2**, 209–225 and *Errata*, **2**, 523.

George, E.O. and Mudholkar, G.S. (1983). On the convolution of logistic random variables. *Metrika*, **30**, 1–13.

Goel, P.K. (1975). On the distribution of standardized mean of samples from the logistic population. *Sankhyā*, 2B, 165–172.

Gumbel, E.J. (1944). Ranges and midranges. *Annals of Mathematical Statistics*, **15**, 414–422.

Gumbel, E.J. and Keeney, R.D. (1950). The extremal quotient. *Annals of Mathematical Statistics*, **21**, 523–538.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis (Mimeograph Series No. 150). Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Gupta, S.S. (1962). Life test sampling plans for normal and lognormal distributions. *Technometrics*, **4**, 151–175.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, **7**, 225–245.

Gupta, S.S. and Balakrishnan, N. (1991). Logistic Order Statistics and Their Properties. In *Handbook of the Logistic Distribution* (Ed., N. Balakrishnan), Marcel Dekker, New York (To appear).

Gupta, S.S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika*, **53**, 565–570.

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. *Annals of Mathematical Statistics*, **36**, 907–920.

Hirschman, I.I. and Widder, D.V. (1955). *The Convolution Transform.* Princeton, N.J.

Lorenzen T.L. and McDonald, G.C. (1981). Selecting logistic populations using the sample medians. *Commun. Statist,-Theor. Meth.* A10, 101-124.

Panchapakesan, S. (1991). Ranking and Selection Procedures for Logistic Populations. In *Handbook of the Logistic Distribution* (Ed. N. Balakrishnan), Marcel Dekker, New York (To appear).

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. *Proceedings of National Academy of Science*, **6**, 275–288.

Plackett, R.L. (1958). Linear estimation from censored data. *Annals of Mathematical Statistics*, **29**, 131–142.

Plackett, R.L. (1959). The analysis of life test data. *Technometrics*, **1**, 9–19.

Schafer, R.E. and Sheffield, T.S. (1973). Inferences on the parameters of the logistic distribution. *Biometrics*, **29**, 449–455.

Schoenberg, I.J. (1953). On Pólya frequency functions and their Laplace transformations. *Journal d'Analyse Mathematique*, **1**, 331–374.

Talacko, J. (1956). Perk's distributions and their role in the theory of Wiener's stochastic variates. *Trabajos de Estadistica*, **7**, 159–174.

Tarter, M.E. and Clark, V.A. (1965). Properties of the Median and other order statistics of logistic variates. *Ann. Math. Statist.*, **36**, 1779–1786.

Table 1. Approximate cdf of the standardized mean of samples from a logistic population: Sample sizes $n = 3,\ 10,\ 15$.

$x \backslash n$	3	10	15
0.	0.500	0.500	0.500
0.10	0.542	0.540	0.540
0.20	0.583	0.580	0.580
0.30	0.623	0.620	0.619
0.40	0.662	0.658	0.657
0.50	0.699	0.694	0.693
0.60	0.734	0.728	0.728
0.70	0.767	0.761	0.760
0.80	0.797	0.791	0.790
0.90	0.824	0.819	0.818
1.00	0.849	0.844	0.843
1.20	0.890	0.887	0.886
1.40	0.922	0.920	0.920
1.60	0.946	0.946	0.946
1.80	0.964	0.964	0.964
2.00	0.976	0.977	0.977
2.20	0.984	0.985	0.986
2.40	0.990	0.991	0.991
2.60	0.994	0.995	0.995
2.80	0.996	0.997	0.997
3.00	0.998	0.998	0.998
3.40	0.999	1.000	1.000
3.80	1.000	1.000	1.000

Table 2. Approximate quantiles of the standardized mean of samples from a logistic population using Cornish-Fisher expansion for $\nu = 8$.

n	Probability level				
	0.900	0.950	0.975	0.990	0.995
3	1.2550	1.6377	1.9850	2.4100	2.7136
5	1.2651	1.6403	1.9755	2.3786	2.6624
10	1.2731	1.6425	1.9680	2.3534	2.6208
15	1.2759	1.6433	1.9654	2.3446	2.6062
25	1.2781	1.6439	1.9632	2.3374	2.5942

Table 3. A comparison of four approximations for the cdf of standardized mean of samples of size 3 from a logistic population.

x	$F_3^*(x)$	$F_3^*(x) - \Phi(x)$	$F_3^*(x) - G_3(x)$	$F_3^*(x) - T_3(x)$	$F_3^*(x) - G_3'(x)$
0.05	0.5209	0.0010	0.0000	0.0001	0.0000
0.15	0.5625	0.0029	0.0000	0.0003	0.0000
0.25	0.6033	0.0046	0.0008	0.0005	0.0000
0.45	0.6809	0.0073	−0.0017	0.0007	0.0001
0.65	0.7506	0.0084	−0.0006	0.0007	0.0000
0.85	0.8106	0.0083	−0.0007	0.0007	0.0000
1.00	0.8486	0.0073	−0.0008	0.0004	0.0000
1.20	0.8903	0.0054	−0.0007	0.0002	0.0000
1.45	0.9291	0.0026	−0.0004	0.0000	0.0000
1.75	0.9598	−0.0001	0.0001	−0.0002	0.0000
2.50	0.9918	−0.0020	0.0004	0.0002	0.0000
3.00	0.9975	−0.0012	0.0001	0.0001	0.0000

$F_3^*(x)$ = cdf of the standardized mean of 3 iid logistic r.v.'s.
$\Phi(x)$ = cdf of the standard normal r.v.
$G_3(x)$ = Edgeworth series expansion correct to order n^{-1}
$T_3(x)$ = cdf of the standardized Student's r.v.'s with 19 d.f.
$G_3'(x)$ = Edgeworth series expansion correct to order n^{-3}

Table 4. Values of the estimate $\hat{n}$ of the minimum sample size n for the single-stage procedure.

k	δ/σ	P^*			
		0.75	0.90	0.95	0.99
2	2.00	0.20	0.77	1.34	2.82
	1.00	0.81	3.22	5.40	10.94
	0.50	3.51	13.07	21.63	43.42
3	2.00	0.46	1.21	1.85	3.42
	1.00	1.95	4.93	7.36	13.25
	0.50	8.11	19.86	29.40	52.49
4	2.00	0.65	1.49	2.17	3.78
	1.00	2.74	5.99	8.55	14.61
	0.50	11.23	24.02	34.06	57.86
5	2.00	0.80	1.69	2.39	4.04
	1.00	3.34	6.76	9.40	15.58
	0.50	13.56	27.04	37.40	61.67
10	2.00	1.28	2.28	3.03	4.76
	1.00	5.12	8.96	11.80	18.29
	0.50	20.50	35.66	46.86	72.38
15	2.00	1.56	2.61	3.39	5.15
	1.00	6.14	10.17	13.12	19.77
	0.05	24.41	40.39	52.01	78.17

Table 5: Values of h_1 for the means rule R_1 for selecting the the subset containing the largest logistic population mean: $P^* = 0.75, 0.90, 0.95$

n	k	0.75	0.90	0.95
	2	0.5395	1.0351	1.3400
	3	0.8131	1.2794	1.5696
3	4	0.9572	1.4115	1.6953
	5	1.0538	1.5012	1.7813
	10	1.3068	1.7395	2.0115
	2	0.4213	0.8052	1.0388
	3	0.6342	0.9935	1.2143
5	4	0.7457	1.0945	1.3097
	5	0.8200	1.1628	1.3746
	10	1.0123	1.3423	1.5468
	2	0.2997	0.5712	0.7351
	3	0.4509	0.7039	0.8578
10	4	0.5296	0.7746	0.9242
	5	0.5818	0.8221	0.9691
	10	0.7159	0.9463	1.0875

Table 6: Performance characteristics of the median rule R_1 under the slippage configuration for $P^* = 0.90$ and $n = 3$.

		$\delta\sqrt{n}$				
k	P(Select $\pi_{(i)}$)	0.5	1.0	2.0	3.0	5.0
2	$i = 1$	0.824	0.717	0.440	0.192	0.013
	$i = 2$	0.948	0.975	0.995	0.999	1.000
	$E(S)$	1.772	1.692	1.436	1.192	1.013
	$E(S^*)$	0.824	0.717	0.440	0.192	0.013
	$E(SR)$	2.720	2.667	2.431	2.191	2.013
	$E(P)$	0.886	0.846	0.718	0.596	0.506
3	$i = 1, 2$	0.856	0.784	0.552	0.283	0.025
	$i = 3$	0.950	0.977	0.996	0.999	1.000
	$E(S)$	2.662	2.545	2.100	1.565	1.050
	$E(S^*)$	1.712	1.568	1.104	0.566	0.050
	$E(SR)$	5.418	5.283	4.644	3.847	3.076
	$E(P)$	0.887	0.848	0.700	0.522	0.350
5	$i = 1, \ldots, 4$	0.875	0.829	0.647	0.380	0.045
	$i = 5$	0.951	0.978	0.996	1.000	1.000
	$E(S)$	4.453	4.295	3.585	2.519	1.178
	$E(S^*)$	3.502	3.317	2.588	1.520	0.178
	$E(SR)$	13.510	13.183	11.453	8.797	5.446
	$E(P)$	0.891	0.859	0.717	0.504	0.236
10	$i = 1, \ldots, 9$	0.888	0.862	0.734	0.492	0.078
	$i = 10$	0.952	0.979	0.997	1.000	1.000
	$E(S)$	8.943	8.735	7.602	5.428	1.702
	$E(S^*)$	7.991	7.756	6.605	4.429	0.702
	$E(SR)$	49.476	48.572	42.992	32.140	13.509
	$E(P)$	0.894	0.874	0.760	0.543	0.170

Table 7: Performance characteristics of the median rule R_2 under the slippage configuration for $P^* = 0.90$ and $n = 3$.

		$\delta\sqrt{n}$				
k	P(Select $\pi_{(i)}$)	0.5	1.0	2.0	3.0	5.0
2	$i=1$	0.968	0.939	0.817	0.594	0.136
	$i=2$	0.992	0.997	0.999	1.000	1.000
	$E(S)$	1.960	1.935	1.816	1.594	1.136
	$E(S^*)$	0.968	0.939	0.817	0.594	0.136
	$E(SR)$	2.953	2.932	2.816	2.594	2.136
	$E(P)$	0.980	0.968	0.908	0.797	0.568
3	$i=1,2$	0.986	0.976	0.920	0.779	0.284
	$i=3$	0.996	0.998	1.000	1.000	1.000
	$E(S)$	2.968	2.949	2.839	2.557	1.567
	$E(S^*)$	1.972	1.951	1.840	1.557	0.567
	$E(SR)$	5.946	5.921	5.759	5.336	3.851
	$E(P)$	0.989	0.983	0.946	0.852	0.522
5	$i=1,\ldots,4$	0.993	0.990	0.967	0.893	0.468
	$i=5$	0.998	0.999	1.000	1.000	1.000
	$E(S)$	4.971	4.958	4.868	4.571	2.873
	$E(S^*)$	3.974	3.959	3.868	3.571	1.873
	$E(SR)$	14.923	14.893	14.669	13.927	9.682
	$E(P)$	0.994	0.992	0.974	0.914	0.575
10	$i=1,\ldots,9$	0.997	0.996	0.989	0.959	0.676
	$i=10$	0.999	1.000	1.000	1.000	1.000
	$E(S)$	9.971	9.963	9.897	9.627	7.088
	$E(S^*)$	8.972	8.963	8.897	8.627	6.088
	$E(SR)$	54.851	54.812	54.485	54.136	40.440
	$E(P)$	0.997	0.996	0.990	0.963	0.709

Table 8: Comparison of the rule R_1 and R_2: Slippage configuration.

$P^* = 0.90, n = 3, k = 4$						
	$\delta\sqrt{n} = 1.5$			$\delta\sqrt{n} = 3.0$		
Perf. Char.	R_1	R_2	R_1/R_2	R_1	R_2	R_1/R_2
$P(CS)$	0.991	1.000	0.991	1.000	1.000	1.000
$E(S)$	3.169	3.921	0.808	2.018	3.560	0.567
$E(S^*)$	2.179	2.922	0.746	1.018	2.560	0.398
$E(SR)$	8.320	9.841	0.845	6.034	9.120	0.662
$E(P)$	0.792	0.980	0.808	0.504	0.890	0.566

$P^* = 0.90, n = 5, k = 4$						
	$\delta\sqrt{n} = 1.5$			$\delta\sqrt{n} = 3.0$		
Perf. Char.	R_1	R_2	R_1/R_2	R_1	R_2	R_1/R_2
$P(CS)$	0.991	1.000	0.991	1.000	1.000	1.000
$E(S)$	3.167	3.999	0.792	2.025	3.917	0.517
$E(S^*)$	2.176	2.922	0.727	1.025	2.917	0.351
$E(SR)$	8.317	9.983	0.833	6.048	9.834	0.615
$E(P)$	0.792	0.998	0.794	0.506	0.979	0.517

$P^* = 0.95, n = 3, k = 4$						
	$\delta\sqrt{n} = 1.5$			$\delta\sqrt{n} = 3.0$		
Perf. Char.	R_1	R_2	R_1/R_2	R_1	R_2	R_1/R_2
$P(CS)$	0.996	1.000	0.996	1.000	1.000	1.000
$E(S)$	3.496	3.981	0.878	2.435	3.852	0.632
$E(S^*)$	2.500	2.981	0.839	1.435	2.852	0.501
$E(SR)$	8.985	9.962	0.902	6.869	9.703	0.689
$E(P)$	0.874	0.995	0.878	0.609	0.963	0.632

$P^* = 0.95, n = 5, k = 4$						
	$\delta\sqrt{n} = 1.5$			$\delta\sqrt{n} = 3.0$		
Perf. Char.	R_1	R_2	R_1/R_2	R_1	R_2	R_1/R_2
$P(CS)$	0.997	1.000	0.997	1.000	1.000	1.000
$E(S)$	3.489	3.999	0.827	2.429	3.987	0.609
$E(S^*)$	2.492	2.999	0.831	1.429	2.987	0.478
$E(SR)$	8.971	9.998	0.897	6.858	9.974	0.688
$E(P)$	0.872	1.000	0.872	0.607	0.997	0.609