ON SELECTING A SUBSET WHICH CONTAINS ALL POPULATIONS BETTER THAN A STANDARD

By Shanti S. Gupta[1] and Milton Sobel

Bell Telephone Laboratories

1. Summary. A procedure is given for selecting a subset such that the probability that all the populations better than the standard are included in the subset is equal to or greater than a predetermined number P^*. Section 3 deals with the problem of the location parameter for the normal distribution with known and unknown variance. Section 4 deals with the scale parameter problem for the normal distribution with known and unknown mean as well as the chi-square distribution. Section 5 deals with binomial distributions where the parameter of interest is the probability of failure on a single trial. In each of the above cases the case of known standard and unknown standard are treated separately. Tables are available for some problems; in other problems transformations are used such that the given tables are again appropriate.

2. Introduction. C. W. Dunnett [3] has considered a different but related problem of comparing several treatment means with a control mean for normal distributions with a common unknown variance. His goal is to separate those treatments which are better than the control from those that are worse (or not better). He controls the probability of selecting the standard as the best (i.e., classifying all other treatments as worse) *when the treatments are all equal to* (*or worse than*) *the standard*. Earlier, E. Paulson [8] considered the problem of selecting the best one of k categories when comparing $k - 1$ categories with a standard. He deals with means of normal distributions with a common unknown variance and also with binomial distributions. He controls the probability of selecting the standard as the best *when the categories are equal to* (*or worse than*) *the standard*.

The procedure described in this paper controls the probability that the selected subset *contains all* those populations better than the control *for any possible true configuration*. If we define a correct decision as a selected subset which *contains all* those populations better than the standard, then the procedure given below guarantees a probability of a correct decision to be at least P^*, not only when the $k - 1$ populations are equal to (or worse than) the standard, but for any possible true configuration. Although we are comparing the procedure with the work noted above, it should be stressed that the goals are different and the procedures are not interchangeable. It should be noted that the treatment of Secs. 3 and 4 could be applied to several other distributions in the Koopman-Darmois family.

The goal treated in this paper is more flexible in that it allows the experimenter to choose a subset and withhold judgment about which is the best one. Then, if

Received March 14, 1957; revised July 1, 1957.

[1] Now at the University of Alberta.

the best one is desired it can be chosen from the selected subset on the basis of economic or other considerations.

Although the title and discussion above use the phraseology "populations better than a standard" we shall actually be interested in selecting all populations as good as or better than the standard; for practical purposes the distinction is of minor importance since in most of the practical problems the parameters of interest can have any value in some interval and are very rarely equal.

To discuss confidence statements we consider first the problems below in which the better populations are the ones with the larger values of the main parameter of interest τ. After the experiment is performed, we can make with confidence P^* the joint statement that *for all populations which are eliminated the parameter value is less than that of the standard.* This joint confidence statement follows from the fact that in selecting a subset *containing all* populations as good as or better than a standard we are automatically eliminating a subset containing only populations worse than the standard. Hence this procedure can be used *to eliminate* those populations which are distinctly inferior to the standard.

For the case in which the better populations are defined to be the ones with the smaller values of τ, the statistical problem is identical and all the results and tables of this paper apply with the obvious modifications.

3. Location parameter—normal populations. We shall assume that populations $\Pi_1, \Pi_2, \cdots, \Pi_p$ with unknown means $\mu_1, \mu_2, \cdots, \mu_p$, respectively are given and that Π_0 is the standard or control, whose mean μ_0 may or may not be known. For clarity we shall discuss the various cases separately.

Case A. *Common known variance* (μ_0 *known*). From each of the p populations $\Pi_i (i = 1, 2, \cdots, p)$, n_i independent observations are taken. Let $\bar{x}_i$ denote the sample mean from Π_i and let σ^2 be the common known variance.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\bar{x}_i \geqq \mu_0 - d\sigma/\sqrt{n_i}\text{.''} \tag{3.1}$$

To determine the value of d let p_1, p_2 denote the true number of populations with $\mu \geqq \mu_0$ and $\mu < \mu_0$, respectively, so that $p_1 + p_2 = p$. Then the probability P of retaining all the p_1 populations with $\mu \geqq \mu_0$ is given by

$$\begin{aligned} P &= \prod_{i=1}^{p_1} P\{\bar{x}_i' \geqq \mu_0 - d\sigma/\sqrt{n_i'}\} \\ &= \prod_{i=1}^{p_1} P\{\sqrt{n_i'}(\bar{x}_i' - \mu_i')/\sigma \geqq -d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma\}, \end{aligned} \tag{3.2}$$

where primes refer to values associated with the p_1 populations for which $\mu \geqq \mu_0$. Hence

$$P = \prod_{i=1}^{p_1} \{1 - F(-d + \sqrt{n_i'}(\mu_0 - \mu_i')/\sigma)\} \tag{3.3}$$

where $F(x)$ refers to the standard normal cumulative distribution function. The μ_i' in (3.3) are restricted by the condition $\mu_i' \geqq \mu_0$ and the minimum of (3.3) is attained by setting $\mu_i' = \mu_0(i = 1, 2, \cdots, p_1)$. Since the result depends on the unknown integer p_1, we can obtain a lower bound by setting $p_1 = p$. Then using the symmetry of F we have

$$P \geqq F^p(d). \tag{3.4}$$

The equation determining d is obtained by setting the right-hand member of (3.4) equal to P^* and is given by

$$F(d) = (P^*)^{1/p}. \tag{3.5}$$

It should be noted that (3.4) is independent of μ_0, τ and n_i. Hence with a table of the standard normal c.d.f. one can easily find the appropriate d which satisfies (3.5) and is to be used in rule (3.1) for any μ_0, any σ and any vector n_i.

The case when the normal populations have different but known variances and the standard is known is treated similarly. The inequality defining the procedure for this problem, corresponding to (3.5), is

$$\bar{x}_i \geqq \mu_0 - d\sigma_i/\sqrt{n_i} \tag{3.6}$$

and the equation determining d is exactly the same as (3.5).

Case B. *Common known variance* (μ_0 *unknown*). In this case n_0 independent observations are taken on the standard Π_0. Let $\bar{x}_0$ denote the mean of these n_0 observations and let σ^2 be the known common variance for all the $(p + 1)$ populations. Then the procedure is to select all those populations for which the relation

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma/\sqrt{n_i} \tag{3.7}$$

is satisfied. The equation determining d is obtained by the same argument as in Case A and, letting $f(x)$ denote the standard normal density, we obtain

$$\int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u\sqrt{\frac{n_i}{n_0}} + d\right)\right] f(u)\, du = P^*. \tag{3.8}$$

For the special case $n_i = n(i = 0, 1, \cdots, p)$ this reduces to

$$\int_{-\infty}^{\infty} F^p(u + d)f(u)\, du = P^*. \tag{3.9}$$

Equation (3.9) is independent of σ. Hence a single two-way table of d-values for different values of P^* and p solves the problem for all values of σ when $n_i = n(i = 0, 1, \cdots, p)$. Tables of d-values satisfying (3.9) for several values of P^* are given in [2] for $p = 1\ (1)\ 10$ and in [5] for $p = 1\ (1)\ 50$. A short table, using only two decimals of the original four, is excerpted from [5] (see Table I). In the more general case when the populations have different but known variances the procedure is defined by

$$\bar{x}_i \geqq \bar{x}_0 - d\sigma_i/\sqrt{n_i} \tag{3.10}$$

TABLE I[a]

Table of d-values satisfying (3.9) *and used in the procedure defined by* (3.7)

p	P^*			
	.75	.90	.95	.99
1	0.95	1.81	2.33	3.29
2	1.43	2.23	2.71	3.62
3	1.68	2.45	2.92	3.80
4	1.85	2.60	3.06	3.92
5	1.97	2.71	3.16	4.01
6	2.06	2.80	3.24	4.09
7	2.14	2.87	3.31	4.15
8	2.21	2.93	3.37	4.20
9	2.26	2.98	3.42	4.25
10	2.31	3.03	3.46	4.29
15	2.50	3.20	3.63	4.44
20	2.62	3.32	3.74	4.54
30	2.79	3.48	3.89	4.68
40	2.90	3.58	4.00	4.78
50	2.99	3.67	4.08	4.85

[a] For a more complete table see [5].

and the equation determining d is

$$\int_{-\infty}^{\infty} \prod_{i=1}^{p} \left[F\left(u \frac{\sigma_0}{\sigma_i} \sqrt{\frac{n_i}{n_0}} + d \right) \right] f(u) \, du = P^*. \tag{3.11}$$

this reduces to (3.9) in the case when $\sigma_i/\sqrt{n_i}$ = constant ($i = 0, 1, \cdots, p$).

Case C. *Common unknown variance* (μ_0 *known*). As in Case A, n_i observations are taken only on the p populations $\Pi_i (i = 1, 2, \cdots, p)$. Let s_ν^2 denote the pooled estimate of σ^2 based on $\nu = \sum_{i=1}^{p} (n_i - 1)$ degrees of freedom ($n_i > 1$ for at least one i). Then the procedure is to select those and only those populations Π_i for which

$$\bar{x}_i \geqq \mu_0 - ds_\nu/\sqrt{n_i}\,. \tag{3.12}$$

The equation determining d is

$$\int_0^{\infty} F^p(yd) q_\nu(y) \, dy = P^*, \tag{3.13}$$

where $q_\nu(y)$ is the density of $y = s_\nu/\sigma = \chi_\nu/\sqrt{\nu}$. This result holds for any μ_0 and depends on n_i only through the value of ν.

Case D. *Common unknown variance* (μ_0 *unknown*). In this case n_i observations are taken on all the populations $\Pi_i (i = 0, 1, \cdots, p)$ and the pooled estimate s_ν^2 of σ^2 is based on $\nu = \sum_{i=0}^{p} (n_i - 1)$ d.f. ($n_i > 1$ for at least one i).

The inequality defining the procedure is

$$\bar{x}_i \geqq \bar{x}_0 - ds_\nu/\sqrt{n_i}\,. \tag{3.14}$$

The equation determining d is

$$\int_0^\infty \int_{-\infty}^\infty \left[\prod_{i=1}^p F\left(u\sqrt{\frac{n_i}{n_0}} + yd\right)\right] f(u)q_\nu(y)\, du\, dy = P^*. \tag{3.15}$$

For $n_i = n(i = 0, 1, \cdots, p)$ this reduces to

$$\int_0^\infty \int_{-\infty}^\infty F^p(u + yd)f(u)q_\nu(y)\, du\, dy = P^*. \tag{3.16}$$

Methods for evaluating this double integral and tables of d-values for selected values of P^*, p and ν are given in [6] and values of $d/\sqrt{2}$ for other values of p and ν are given in [3].

4. Scale parameter—gamma or chi-square populations. In this section it will be more natural to define the population Π_i as better than Π_0 if the scale parameter $\theta_i < \theta_0$.

Case A. θ_0 *known*. We assume that the population $\Pi_i(i = 1, 2, \cdots, p)$ has the density

$$\frac{1}{\Gamma\left(\frac{\alpha_i}{2}\right)} \frac{1}{\theta_i^{\alpha_i/2}} x^{\frac{\alpha_i}{2}-1} e^{-x/\theta_i}. \tag{4.1}$$

If $x_{ij}(j = 1, 2, \cdots, n_i)$ are the n_i observations on Π_i, then $t_i = \sum_{j=1}^{n_i} x_{ij}$ has the density (4.1) with α_i replaced by $\nu_i = n_i\alpha_i$ and the procedure is as follows.

Procedure: "Retain in the selected subset only those populations

$$\Pi_i(i = 1, 2, \cdots, p)$$

for which

$$\frac{t_i}{\nu_i} \leq (1 + d)\theta_0." \tag{4.2}$$

Let q_1 and q_2 denote the number of populations with $\theta \leq \theta_0$ and $\theta > \theta_0$, respectively, so that $q_1 + q_2 = p$. The probability P of a correct decision is given by

$$P = \prod_{i=1}^{q_1} P\left\{\frac{t_i'}{\theta_i'} \leq (1 + d)\frac{\theta_0\nu_i'}{\theta_i'}\right\}, \tag{4.3}$$

where primes refer to the q_1 populations with $\theta \leq \theta_0$. Hence,

$$P = \prod_{i=1}^{q_1} G_{\nu_i'}\left[(1 + d)\frac{\theta_0\nu_i'}{\theta_i'}\right], \tag{4.4}$$

where $G_{\nu_i}(x)$ is the c.d.f. of the gamma density in (4.1) with α_i replaced by ν_i and $\theta_i = 1$. A lower bound to this probability is obtained by setting $\theta_i = \theta_0$

and $q_1 = p$ so that the equation determining d can be written in the form

$$\prod_{i=1}^{p} \left\{ \frac{1}{\Gamma\left(\frac{\nu_i}{2}\right)} \int_0^{\nu_i(1+d)} u^{\frac{\nu_i}{2}-1} e^{-u}\, du \right\} = P^*. \tag{4.5}$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this is easily solved with the help of a table of the c.d.f. of $\gamma_\nu = \frac{1}{2}\chi_\nu^2$ with ν degrees of freedom.

Application to normal populations. If $\theta_i = 2\sigma_i^2 (i = 0, 1, \cdots, p)$ are the scale parameters for the $(p + 1)$ normal populations and $x_{ij} (j = 1, 2, \cdots, n_i)$ are the n_i observations on the population Π_i with the mean μ_i (known), then we retain the population Π_i in the selected subset if

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \leqq 2(1 + d)\sigma_0^2. \tag{4.6}$$

The equation determining the d in (4.6) is the same as (4.5) with ν_i replaced by n_i.

If the means μ_i are unknown and $n_i > 1 (i = 1, 2, \cdots, p)$, then in (4.6) we use the sample mean $\bar{x}_i$ in place of μ_i and $n_i - 1$ in place of n_i. The equation determining d is again (4.5) with $\nu_i = n_i - 1$.

Transformation: If we apply the transformation [1]

$$y_i = \ln\left(\frac{t_i}{\nu_i}\right) \qquad (i = 1, 2, \cdots, p), \tag{4.7}$$

then the procedure (4.2) of this section can be put in the form

$$y_i \leqq \ln\left(\frac{\theta_0}{2}\right) + d_1, \tag{4.8}$$

where

$$d_1 = \ln\,[2(1 + d)]. \tag{4.9}$$

Then using the normal approximation and the same argument as before, the approximate equation determining d_1 is

$$\prod_{i=1}^{p} \left\{ F\left(d_1 \sqrt{\frac{\nu_i}{2}}\right) \right\} = P^*. \tag{4.10}$$

For $\nu_i = \nu (i = 1, 2, \cdots, p)$ this gives an equation similar to (3.5). For the application to normal populations the equation corresponding to (4.8) is

$$\ln s_i^2 \leqq \ln \sigma_0^2 + d_1, \tag{4.11}$$

where d_1 is determined by (4.10) with $\nu_i = n_i$ or $n_i - 1$ according as the means μ_i are or are not known.

Case B. θ_0 *unknown.* The assumptions are the same as in Case A except that n_0 observations, viz., $x_{01}, x_{02}, \cdots, x_{0n_0}$ are taken on Π_0. The inequality defining the procedure and corresponding to (4.2) is

$$\frac{t_i}{\nu_i} \leqq (1 + d)\frac{t_0}{\nu_0}, \tag{4.12}$$

where $t_0 = \sum_{j=1}^{n_0} x_{0j}$ and $\nu_0 = n_0\alpha_0$. The equation determining d is obtained as before and is given by

$$\int_0^\infty \left[\prod_{i=1}^{p} \int_0^{\nu_i t(1+d)/\nu_0} \frac{u^{\frac{\nu_i}{2}-1} e^{-\frac{u}{2}}}{\Gamma(n_i/2)}\, du\right] \frac{t^{\frac{\nu_0}{2}-1} e^{-\frac{t}{2}}}{\Gamma(\nu_0/2)}\, dt. \tag{4.13}$$

Application to normal populations. For the case where the means are known, the rule takes the form

$$\frac{1}{n_i}\sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \leqq \frac{(1 + d)}{n_0}\sum_{j=1}^{n_0} (x_{0j} - \mu_0)^2; \tag{4.14}$$

where d is given by (4.13) with $\nu_i = n_i$. If μ_i's are not known and

$$n_i > 1(i = 0, 1, \cdots, p),$$

then the rule is the same as (4.14) with μ_i and n_i replaced by $\bar{x}_i$ and $n_i - 1$, respectively. The equation determining d is again (4.13) with $\nu_i = n_i - 1$.

Transformation: Using the transformation (4.7), we put the inequality defining the rule as

$$y_i \leqq y_0 + d_2. \tag{4.15}$$

The approximate equation determining d_2 is

$$\int_{-\infty}^{\infty} \left[\prod_{i=1}^{p} F\left(u\sqrt{\frac{n_i}{n_0}} + d_2\sqrt{\frac{n_i}{2}}\right)\right] f(u)\, du = P^*, \tag{4.16}$$

which for $n_i = n(i = 0, 1, \cdots, p)$ is of the same form as (3.9).

5. Binomial populations.

Case A. *Known standard.* It is assumed that $p + 1$ binomial populations Π_i with parameters $\theta_i(i = 0, 1, \cdots, p)$ are given where θ_0 is the known value of the probability of a unit being defective in the standard population Π_0. Again n_i independent observations are taken from each population

$$\Pi_i(i = 1, 2, \cdots, p).$$

Since θ_i is the probability of a unit being defective, we define Π_i to be better than Π_0 when $\theta_i < \theta_0$. Let x_i denote the number of defectives observed in the sample of n_i observations from $\Pi_i(i = 1, 2, \cdots, p)$.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i(i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i} x_i \leqq \theta_0 + d\sqrt{\frac{\theta_0(1 - \theta_0)}{n_i}}. \text{"} \tag{5.1}$$

Let q_1, q_2, be defined as in Sec. 4; let $[m_i(d)]$ denote the largest integer in

$$m_i(d) = n_i\theta_0 + d\sqrt{n_i\theta_0(1-\theta_0)} \qquad (i = 1, 2, \cdots, p). \tag{5.2}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ is given by

$$P = \prod_{i=1}^{q_1} \left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i} \theta_i^j (1-\theta_i)^{n_i-j} \right]. \tag{5.3}$$

A lower bound is obtained by setting $\theta_i = \theta_0 (i = 1, 2, \cdots, p)$ and $q_1 = p$. The fact that $\theta_1 = \theta_0$ gives a lower bound can be shown by writing the binomial sum as an incomplete Beta function. Hence the inequality determining d becomes

$$\prod_{i=1}^{p} \left[\sum_{j=0}^{[m_i(d)]} C_j^{n_i} \theta_0^j (1-\theta_0)^{n_i-j} \right] \geqq P^*, \tag{5.4}$$

and the solution is the smallest value of d satisfying (5.4). If $n_i = n$ then

$$[m_i(d)] = [m(d)]$$

and (5.4) reduces to

$$\sum_{j=0}^{[m(d)]} C_j^{n} \theta_0^j (1-\theta_0)^{n-j} \geqq (P^*)^{1/p}. \tag{5.5}$$

This is easily solved by consulting a table of cumulative binomial probabilities.

For large values of n_i (large enough for the normal approximation to give good results) the inequality determining d can be approximated by the simple equation

$$F(d) = (P^*)^{1/p}, \tag{5.6}$$

where F is the standard normal c.d.f. This equation is independent of n_i and is much easier to solve than (5.4).

Case B. *Unknown standard.* The assumptions are the same as in Case A except that n_0 observations are taken on the standard population Π_0. Let x_0 be the number of defectives among n_0.

Procedure: "Retain in the selected subset those and only those populations $\Pi_i (i = 1, 2, \cdots, p)$ for which

$$\frac{1}{n_i} x_i \leqq \frac{1}{n_0} x_0 + \frac{d}{2} \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}." \tag{5.7}$$

The probability P of retaining all the q_1 populations with $\theta \leqq \theta_0$ attains a minimum when $\theta_i = \theta (i = 0, 1, \cdots, p)$ and $q_1 = p$ and is given by

$$P(\theta, d) = \sum_{y=0}^{n_0} \prod_{i=1}^{p} \left[\sum_{j=0}^{[m_i(y,d)]} C_j^{n_i} \theta^j (1-\theta)^{n_i-j} \right] C_y^{n_0} \theta^y (1-\theta)^{n_0-y}, \tag{5.8}$$

where $[m_i(y, d)]$ is the largest integer contained in

$$m_i(y, d) = \frac{n_i}{n_0} y + \frac{dn_i}{2} \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}. \tag{5.9}$$

Then the desired value of d for (5.7) is the smallest number for which

$$\min_{0 \leqq \theta \leqq 1} P(\theta, d) \geqq P^*. \tag{5.10}$$

Since, except for very small n_i or very large p, the minimum occurs near $\theta = \frac{1}{2}$, we can obtain an approximate solution for d by finding the smallest number for which

$$P(\tfrac{1}{2}, d) \geqq P^*. \tag{5.11}$$

A simpler approximate solution, which gives good results when the n_i are not too small and p is not too large, is the normal approximation obtained under the assumption that $\theta_i = \frac{1}{2}(i = 0, 1, \cdots, p)$. Then from (5.7) we obtain for the approximate equation determining d

$$\int_{-\infty}^{\infty} \left[\prod_{i=1}^{p} F\left(u \sqrt{\frac{n_i}{n_0}} + d \sqrt{1 + \frac{n_i}{n_0}} \right) \right] f(u)\, du = P^*. \tag{5.12}$$

For $n_i = n(i = 0, 1, \cdots, p)$ the rule (5.7) can be written as

$$x_i \leqq x_0 + d', \tag{5.13}$$

where $d' = d\sqrt{n/2}$. In carrying out the rule we can assume that d' is an *integer*. The desired value of d' is the smallest integer for which

$$\min_{0 \leqq \theta \leqq 1} \left\{ \sum_{y=0}^{n} \left[\sum_{j=0}^{y+d'} C_j^n \theta^j (1 - \theta)^{n-j} \right]^p C_y^n \theta^y (1 - \theta)^{n-y} \right\} \geqq P^*. \tag{5.14}$$

Then (5.12) can be written in the form

$$\int_{-\infty}^{\infty} F^p(u + \hat{d}) f(u)\, du = P^* \tag{5.15}$$

and the relation between d' and $\hat{d}$, using a continuity correction, is

$$d' = d \sqrt{\frac{n}{2}} = \left\{ \frac{\hat{d}\sqrt{n} - 1}{2} \right\} \tag{5.16}$$

where $\{x\}$ is the smallest integer greater than or equal to x.

Transformation: It may be desirable to solve the binomial problem by using an arc sine transformation and converting it into one involving the location parameter of the normal distributions. For example, for the Case B above with $n_i = n(i = 0, 1, \cdots, p)$ if we use the arc sine transformation as given in [4], the inequality defining the procedure is

$$\text{arc sin} \sqrt{\frac{x_i}{n+1}} + \text{arc sin} \sqrt{\frac{x_i+1}{n+1}} \leqq \text{arc sin} \sqrt{\frac{x_0}{n+1}} + \text{arc sin} \sqrt{\frac{x_0+1}{n+1}} + \frac{d\sqrt{2}}{\sqrt{2n+1}}, \tag{5.17}$$

where the approximate equation determining d is the same as (3.9) so that Table I is applicable here also.

REFERENCES

[1] M. S. Bartlett and D. G. Kendall, "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *J. Roy. Stat. Soc. Suppl.*, Vol. 8 (1946), pp. 128–138.

[2] R. E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–29.

[3] C. W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control," *J. Amer. Stat. Assn.*, Vol. 50 (1955), pp. 1096–1121.

[4] M. F. Freeman and J. W. Tukey, "Transformations related to the angular and the square root," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 607–611.

[5] S. S. Gupta, "On a decision rule for a problem in ranking means," Mimeograph Series No. 150, May, 1956, Institute of Statistics, University of North Carolina, Chapel Hill, N. C.

[6] S. S. Gupta and M. Sobel, "On a statistic which arises in selection and ranking problems," submitted for publication in *Ann. Math. Stat.*

[7] S. S. Gupta and M. Sobel, "A procedure for selecting a subset containing the best of k binomial populations," to be published.

[8] E. Paulson, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 239–246.