ON THE COMPARISON OF SEVERAL EXPERIMENTAL CATEGORIES WITH A CONTROL[1]

By Edward Paulson

University of Washington

Summary. This paper investigates certain statistical problems arising in the determination of the "best" of k categories when comparing $k - 1$ experimental categories with a standard or control. The discussion is limited to the case of a single stage sampling procedure with an equal number of observations on each of the k categories. Results both of an exact and of an approximate nature are obtained when (a) the observations with each category are normally distributed, and (b) the observations with each category have a binomial distribution.

1. Introduction. In this paper we will be concerned with the problem of the selection of one of the k categories $\Pi_1, \Pi_2, \cdots, \Pi_k$ as best when category Π_1 plays a special role, since it represents the standard or control, while $\Pi_2, \Pi_3, \cdots, \Pi_k$ represent $k - 1$ experimental categories. For the type of application we have in mind, the k categories might represent k varieties of wheat, or k drugs, or k machines; the "goodness" of a category will depend on some parameter of the probability distribution associated with that category. The experimental categories can be classified into two groups: one group consisting of those categories which are superior to Π_1 and a second group consisting of those experimental categories which are inferior to or at most equal to Π_1. In such a situation it will usually be desirable to have special protection against the selection of an experimental category as best when it actually is inferior to Π_1. This will be accomplished by requiring that the statistical procedure used provide a special assurance that Π_1 will be selected as best if the second group happens to consist of all $k - 1$ experimental categories, that is, none of the experimental categories is superior to Π_1. Situations of this type are believed to be fairly common in experiments in medicine and agriculture.

We will therefore consider the following statistical problem: given a sample consisting of kn independent observations $\{x_{ij}\}$ $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, n)$, where x_{ij} is the jth observation with category Π_i, to devise a statistical procedure for selecting one out of the k categories as best so that if none of the experimental categories $\Pi_2, \Pi_3, \cdots, \Pi_k$ is actually "superior" to Π_1, then the probability that Π_1 is selected will be $\geqq 1 - \alpha$. We will also consider the related problem of deciding how large a sample will be required so that when one of the experimental categories is really superior to all the others including Π_1 by a specified amount the probability will also be $\geqq 1 - \beta$ that this experimental category will be selected as best. The constants α and β might be considered as

[1] Work on this paper done under the sponsorship of the office of Naval Research.

roughly analogous to the type I and type II errors in the Neyman-Pearson theory of testing a hypothesis. However the present problem is of a multiple decision type, and is not equivalent to one involving the testing of a hypothesis unless $k = 2$.

For the most part the discussion will be confined to the normal case, when the n observations with category Π_i ($i = 1, 2, \cdots, k$) are assumed to be normally and independently distributed with mean m_i and common variance σ^2; the best category is defined to be the one associated with the greatest value of m_i. A brief discussion will also be given of the binomial case, when each observation with category Π_i is classified as a "success" or "failure" with a probability P_i of being a "success"; the best category is defined to be the one associated with the greatest value of P_i.

2. The normal case with known variance. We will first treat the problem when σ is assumed to be known a priori. Let $\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$, $\bar{x}^* = \max(\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_k)$, $\alpha_1 = \alpha/(k-1)$, and for any a ($0 < a < 1$) let ν_a be defined by the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\nu_a}^{\infty} e^{-\frac{1}{2}t^2}\, dt = a.$$

Let Π^* be the experimental category whose mean is $\bar{x}^*$, and let λ be a constant whose value will be determined in a moment. The following statistical procedure is proposed for the selection of the best category:

$$\text{(1)} \qquad \begin{aligned} &\text{If } \bar{x}^* - \bar{x}_1 \geqq \lambda\sigma \sqrt{\frac{2}{n}}, \text{ select } \Pi^*; \\ &\text{If } \bar{x}^* - \bar{x}_1 < \lambda\sigma \sqrt{\frac{2}{n}}, \text{ select } \Pi_1. \end{aligned}$$

We now complete the specification of the statistical procedure by determining λ. It is obvious that when $m_1 \geqq \max(m_2, m_3, \cdots, m_k)$, the greatest lower bound of the probability that $\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n}$ will occur when $m_1 = m_2 = \cdots = m_k$. In order to evaluate $P\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n} \mid m_1 = m_2 = \cdots = m_k\}$ we use the fact that $\bar{x}^*$ and $\bar{x}_1$ are independent, and find, after some simplifications,

$$\begin{aligned} P&\left\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{\frac{2}{n}} \,\middle|\, m_1 = m_2 = \cdots = m_k\right\} \\ &= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}} \int_{-\infty}^{r+\lambda\sqrt{2}} (k-1)\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \left[\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt\right]^{k-2} dz\, dr \\ &= \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}r^2}}{\sqrt{2\pi}} \left[\int_{-\infty}^{r+\lambda\sqrt{2}} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}\, dt\right]^{k-1} dr. \end{aligned}$$

The constant λ will therefore be given as the root of the equation

$$\text{(2)} \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}r^2} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r+\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\, dt\right]^{k-1} dr = 1 - \alpha.$$

From equation (2) it is possible to tabulate the values of λ as a function of α and k. Pending the construction of adequate tables, we will give an approximate method for finding λ with a definite limit of error. For this purpose, let A_i stand for the event $\bar{x}_i - \bar{x}_1 \geqq \lambda\sigma\sqrt{2/n}$. We have $P\{\bar{x}^* - \bar{x}_1 < \lambda\sigma\sqrt{2/n}\} = 1 - P(A_2 + A_3 + \cdots + A_k)$, where the probabilities involved are to be calculated for the case when all the k means are equal. Making use of Bonferroni's Inequality (see [1], p. 75) we have

$$1 - \sum_{i=2}^{k} P(A_i) + \sum_{i}\sum_{\substack{j \\ j \geqq i}} P(A_i \cdot A_j) \geqq P\left\{\bar{x}^* - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} \geqq 1 - \sum_{i=2}^{k} P(A_i).$$

Due to the symmetry when the means are equal, this becomes

$$(3)\qquad \begin{aligned} 1 - (k-1)P(A_2) + \frac{(k-1)(k-2)}{2} P(A_2 \cdot A_3) \\ \geqq P\left\{\bar{x}^* - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} \geqq 1 - (k-1)P(A_2). \end{aligned}$$

Since $(\bar{x}_2 - \bar{x}_1)$ and $(\bar{x}_3 - \bar{x}_1)$ have a bivariate normal distribution with correlation $= \frac{1}{2}$, we obtain

$$P(A_2) = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-\frac{1}{2}t^2}\, dt,$$

$$P(A_2 A_3) = \frac{1}{\pi\sqrt{3}} \int_{\lambda}^{\infty}\int_{\lambda}^{\infty} e^{-\frac{2}{3}(x^2 - xy + y^2)}\, dx\, dy.$$

If we use as the approximate value for λ the solution $\tilde{\lambda}$ of the equation

$$(4)\qquad \frac{1}{\sqrt{2\pi}} \int_{\tilde{\lambda}}^{\infty} e^{-\frac{1}{2}t^2} dt = \alpha_1,$$

then from (3) it follows that $P(\tilde{\lambda}) = P\{\bar{x}^* - \bar{x}_1 < \tilde{\lambda}\sigma\sqrt{2/n}\}$ will exceed $1 - \alpha$ by an amount which is not greater than $\frac{1}{2}(k-1)(k-2)P(A_2 A_3)$. This quantity can be calculated from the tables of the volumes of the normal bivariate surface [2]. The calculations for several values of α and k are summarized in Table I. It appears that the approximation yields good results for values of α which ordinarily are of interest if k is not too large, say $\leqq 6$.

Any statistical procedure for selecting one of the k categories can, of course, lead to other types of error than that of selecting an experimental category as the best when it actually is inferior to the standard or control. In particular, the error in not selecting a particular experimental category as best when it actually is superior to all the other experimental categories and the standard or control by at least a specified amount is of considerable interest. For a fixed value of α, this new type of error can be reduced only by increasing n, the sample size. Suppose for convenience that Π_k is the particular experimental category that exceeds the others by an amount Δ; that is, $m_k \geqq \max(m_1, m_2, \cdots, m_{k-1}) + \Delta$. Using the statistical procedure of (1), it is easy to see that for a fixed λ, Δ, k, and n the greatest lower bound of the probability that Π_k will be selected as best will occur when $m_1 = m_2 = \cdots = m_{k-1} = m$ and $m_k = m + \Delta$.

TABLE I

Limits for $P(\tilde{\lambda})$

k \ α	.02	.05
3	$\tilde{\lambda} = 2.326$ $.981 \geqq P(\tilde{\lambda}) \geqq .98$	$\tilde{\lambda} = 1.960$ $.955 \geqq P(\tilde{\lambda}) \geqq .95$
6	$\tilde{\lambda} = 2.652$ $.984 \geqq P(\tilde{\lambda}) \geqq .98$	$\tilde{\lambda} = 2.326$ $.963 \geqq P(\tilde{\lambda}) \geqq .95$

If we denote this greatest lower bound by $P(n; \lambda, k, \Delta)$, we easily obtain

$$P(n; \lambda, k, \Delta) = P\left\{\bar{x}_k - \bar{x}_1 > \lambda\sigma\sqrt{\frac{2}{n}} \text{ and } \bar{x}_k > \max(\bar{x}_2, \bar{x}_3, \cdots, \bar{x}_{k-1}) \mid m_k = m_1 + \Delta\right\}$$

$$(5)\quad = \int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w-\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\,dt\right] \left\{\int_{-\infty}^{w}(k-2)\frac{e^{-\frac{1}{2}s^2}}{\sqrt{2\pi}}\left[\int_{-\infty}^{s}\frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}}\,dt\right]^{k-3} ds\right\}\frac{e^{-\frac{1}{2}(w-(\Delta/\sigma)\sqrt{n})^2}}{\sqrt{2\pi}}\,dw$$

$$= \int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w+(\Delta/\sigma)\sqrt{n}-\lambda\sqrt{2}} e^{-\frac{1}{2}t^2}\,dt\right]\cdot\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w+(\Delta/\sigma)\sqrt{n}} e^{-\frac{1}{2}t^2}\,dt\right]^{k-2}\frac{e^{-\frac{1}{2}w^2}}{\sqrt{2\pi}}\,dw.$$

In order to decide in advance of the experiment how large a sample should be taken, we might try to find n so that $P(n; \lambda, k, \Delta) = 1 - \beta$ where α, β, and Δ are determined by practical considerations depending on the particular experimental situation, and λ is found from (2) or (4). It will be very difficult to find n directly from (5) until tables are made available. However we can usually obtain a good approximation $\tilde{n}$ to the required value by solving the equation

$$(6)\quad P\left\{\bar{x}_k - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}} \,\middle|\, m_k = m_1 + \Delta\right\} = \frac{1}{\sqrt{2\pi}}\int_{(\Delta/\sigma)\sqrt{n/2}-\lambda}^{\infty} e^{-\frac{1}{2}t^2}\,dt = \beta.$$

The solution can be written $\tilde{n} = (2\sigma^2/\Delta^2)(\lambda + \nu_\beta)^2$ which reduces to $\tilde{n} = (2\sigma^2/\Delta^2)(\nu_{\alpha_1} + \nu_\beta)^2$ when $\tilde{\lambda}$ is used for λ. The adequacy of this approximation can be estimated with the help of the inequality[2]

$$(7)\quad P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - (k-2)P\{\bar{x}_k < \bar{x}_2\} \leqq P\{n; \lambda, k, \Delta\} \leqq P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\},$$

[2] The writer is indebted to the referee for this inequality, which is an improvement over the one originally used.

which holds when $m_1 = m_2 = \cdots = m_{k-1}$, $m_k = m_1 + \Delta$. To derive this inequality, it is obvious from the definition of $P\{n; \lambda, k, \Delta)$ that $P\{n; \lambda, k, \Delta\} \leqq P\{\bar{x}_k - \bar{x}_1 > \lambda\sigma\sqrt{2/n}\}$, while from Bonferroni's Inequality we have

$$P\{n; \lambda, k, \Delta\} \geqq 1 - P\left\{\bar{x}_k - \bar{x}_1 \leqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - \sum_{j=2}^{k-1} P\{\bar{x}_k \leqq \bar{x}_j\}$$

$$= P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda\sigma\sqrt{\frac{2}{n}}\right\} - (k-2)P\{\bar{x}_k \leqq \bar{x}_2\}.$$

Hence when $\tilde{n}$ is found from (6), it follows from (7) that $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ will exceed β by an amount which is less than

$$(k-2)P\{\bar{x}_k \leqq \bar{x}_2\} = \frac{(k-2)}{\sqrt{2\pi}} \int_{(\Delta/\sigma)\sqrt{\tilde{n}/2}}^{\infty} e^{-\frac{1}{2}t^2}\, dt.$$

We have attempted to indicate the adequacy of the approximation $\tilde{n}$ found from (6) by computing the upper bound for $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ for several

TABLE II

Upper bound for $[1 - P(\tilde{n}; \tilde{\lambda}, k, \Delta)]$

k	β \ α	.05	.02
3	.20	.2025	.2008
	.05	.0502	.0500
6	.20	.2031	.2010
	.05	.0501	.0500

values of α, β and k; in these calculations the value of $\tilde{\lambda}$ given in Table I was used for λ. The calculations are summarized in Table II.

It appears that for $\beta \leqq .20$, $\alpha \leqq .05$, the upper bound for $[1 - P(\tilde{n}; \lambda, k, \Delta)]$ exceeds the corresponding β by a small amount which can be neglected for most practical purposes.

3. The normal case with unknown variance. The case when σ is unknown will now be briefly considered. Let

$$s^2 = \sum_{i=1}^{k} \sum_{\alpha=1}^{n} \frac{(x_{i\alpha} - \bar{x}_i)^2}{k(n-1)}$$

be the pooled estimate of σ^2 based on $n' = k(n-1)$ degrees of freedom. The statistical procedure in (1) is modified as follows:

$$\text{select } \Pi^* \text{ if } \bar{x}^* - \bar{x}_1 \geqq \lambda_n s\sqrt{\frac{2}{n}};$$

(8)

$$\text{select } \Pi_1 \text{ if } \bar{x}^* - \bar{x}_1 < \lambda_n s\sqrt{\frac{2}{n}}.$$

The exact equation that λ_n must satisfy in order to have $P\{\bar{x}^* - \bar{x}_1 < \lambda_n s \sqrt{2/n} \mid m_1 = m_2 = \cdots = m_k\} = 1 - \alpha$ and an explicit expression for $P(n; \lambda_n, k, \Delta/\sigma)$, the probability that Π_k will be selected as best when $m_1 = m_2 = \cdots = m_{k-1}, m_k/\sigma = m_1/\sigma + \Delta/\sigma$, can easily be found by a procedure similar to that used for (2) and (5); however the results are complicated and instead we will proceed directly to discuss approximate procedures. Let C_j denote the event $\bar{x}_j - \bar{x}_1 > \lambda_n s \sqrt{2/n}$ $(j = 2, 3, \cdots, k)$, and let $t_{n'}$ denote a random variable having the t distribution with n' degrees of freedom. With the aid of tables of the t distribution an approximation $\tilde{\lambda}_n$ to λ_n can be found so that $P\{\bar{x}_2 - \bar{x}_1 > \tilde{\lambda}_n s \sqrt{2/n}\} = P\{t_{n'} > \tilde{\lambda}_n\} = \alpha/(k-1) = \alpha_1$. Then the probability that Π_1 will be selected as best when all the means are equal will exceed $1 - \alpha$ by an amount which is less than $\frac{1}{2}(k-1)(k-2)P(C_2 \cdot C_3)$. For bounds on the second type of error, we have

$$P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda_n s \sqrt{\frac{2}{n}} \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\} - (k-2)P\left\{\bar{x}_k \leqq \bar{x}_2 \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\}$$
$$\leqq P\left\{n; \lambda_n, k, \frac{\Delta}{\sigma}\right\} \leqq P\left\{\bar{x}_k - \bar{x}_1 \geqq \lambda_n s \sqrt{\frac{2}{n}} \,\middle|\, \frac{m_k}{\sigma} = \frac{m_1}{\sigma} + \frac{\Delta}{\sigma}\right\}.$$

All these inequalities are easily obtained as in Section 2. To evaluate the bound for the first type of error, a good approximation can usually be found by regarding s/σ to be normally distributed with mean 1 and variance $1/(2n')$; using this approximation it is easy to verify that

$$P(C_2 \cdot C_3) = P\left\{U \geqq \lambda_n \sqrt{\frac{2n'}{2n' + \lambda_n^2}} \quad \text{and} \quad V \geqq \lambda_n \sqrt{\frac{2n'}{2n' + \lambda_n^2}}\right\},$$

where (U, V) has a bivariate normal distribution with zero means, unit variances and correlation $\rho = (n' + \lambda_n^2)/(2n' + \lambda_n^2)$. This same device might also be used to approximate the upper and lower bounds for $P(n; \lambda_n, k, \Delta/\sigma)$ as an alternative to evaluating the bounds by using tables of the non-central t distribution. Finally, to obtain the value of n so that $P(n; \tilde{\lambda}_n, k, \Delta/\sigma) = 1 - \beta$ a good first approximation will usually be given by $n_0 = (2\sigma^2/\Delta^2)(\nu_{\alpha_1} + \nu_\beta)^2$; after computing $\tilde{\lambda}_{(n_0)}$ and the corresponding upper and lower bounds for $P(n_0; \tilde{\lambda}_{(n_0)}, k, \Delta/\sigma)$, the first approximation n_0 can be modified if necessary and the process iterated.

It should be noted that in order to find the sample size required to control the second type of error, either an approximate value of σ must be known from past experience, or else it must be sufficient for the practical problem under consideration to know the probability of selecting the best experimental category as a function of the ratio of Δ to σ. It is possible to eliminate the dependence of the result on σ by making use of a two-stage sampling scheme due to Stein [3]; this and other sequential procedures may be considered in another paper.

4. The binomial case. In this section a brief treatment of the binomial case will be given, based on the use of the inverse sine transformation. That is, we will use the fact that if $\bar{p}$ is the observed proportion of successes in n independent trials with a constant probability P of a success, then arc sin $\sqrt{\bar{p}}$ is for large n approximately normally distributed with mean (arc sin $\sqrt{P}$) and variance $1/(4n)$ (provided the angle is given in radian measure). This transformation was previously used by W. Allen Wallis and the present writer [4] to design experiments for comparing the percentages associated with one experimental and one standard category. The material in this section can be regarded as one possible extension of that work to the case where we are dealing with more than one experimental category.

Let r_i be the number of successes in the n observations with category Π_i. Let $\bar{p}_i = (r_i/n)$, let $u_i = \text{arc sin} \sqrt{\bar{p}_i}$ and let P_i = the true probability of a success with category Π_i. Let $\bar{p}^* = \max(\bar{p}_2, \bar{p}_3, \cdots, \bar{p}_k)$, $u^* = \max(u_2, u_3, \cdots, u_k)$, and let Π^* be the experimental category with observed percentage of successes $\bar{p}^*$. If there should happen to be more than one category with the observed percentage of successes $= \bar{p}^*$, select Π^* at random from the subset having $\bar{p}_i = \bar{p}^*$.

We now propose the following statistical procedure for selecting one of the k categories.

$$\text{(9)} \qquad \begin{aligned} &\text{select } \Pi^* \text{ if } u^* - u_1 \geqq \lambda \sqrt{\frac{1}{2n}}; \\ &\text{select } \Pi_1 \text{ if } u^* - u_1 < \lambda \sqrt{\frac{1}{2n}}, \end{aligned}$$

where λ is to be chosen so that if $P_1 \geqq \max(P_2, P_3, \cdots, P_k)$ the probability that Π_1 is selected as best will be $\geqq 1 - \alpha$. We assume that n is large enough so that the set $\{u_i\}$ can be regarded as normally distributed with common variance $1/(4n)$ and means arc sin $\sqrt{P_i}$. Therefore the problem is once again essentially equivalent to the normal case with known variance, which was treated in Section 2, and the value of λ is given by the solution of (2). To find the sample size n so that if $P_1 = P_2 = \cdots = P_{k-1} = P$ and $P_k = P + \delta$ $(\delta > 0)$, the probability that Π_k is selected as best will equal $1 - \beta$, set $\Delta = \text{arc sin} \sqrt{P + \delta} - \text{arc sin} \sqrt{P}$, and the required n will be given by (5) when $\sigma\sqrt{2/n}$ is replaced by $\sqrt{1/2n}$.

For values of α, β, and k so that $\alpha \leqq .05$, $\beta \leqq .20$, and $k \leqq 6$, it has been shown in Section 2 that if we use approximate values of λ and n given by (4) and (6), the change in the probabilities considered will be small, and will ordinarily be of little practical importance. Using the notation of the last section, (4) and (6) for the binomial case are equivalent to $\lambda' = \nu_{\alpha_1}$ and $n = (\nu_{\alpha_1} + \nu_\beta)^2/(2\Delta^2)$.

We conclude this section by discussing a specific problem. Consider a situation in which we are interested in investigating the effect of three experimental treatments on a certain disease, where it is known from previous experience that the probability of survival with the standard treatment is of the order of magnitude of .75. The problem we wish to consider is that of designing a statistical procedure (based on a single stage of sampling) for selecting one of the 4 treatments which will have the following properties: (a) the probability of selecting an experimental treatment as best when in fact it is inferior to the standard treatment is to be $\leqq .05$; (b) if one of the experimental treatments should happen to increase the probability of survival to .90, while the probabilities of survival for the three other treatments is $\leqq .75$, then the probability that the superior experimental treatment will be selected as best should be $\geqq .95$. Upon setting $\alpha = .05$, $\beta = .05$, and $k = 4$ we find $\lambda = 2.128$, $\Delta = .202$, $n = (2.128 + 1.645)^2/(2\Delta^2) = 174$.

The required statistical procedure having properties (a) and (b) is the following. A group of 696 animals are all innoculated with the specific disease under consideration, and then the animals are subdivided in some random manner into 4 groups each consisting of 174 animals. The first of these groups is given the standard treatment, and the remaining groups each receive one of the experimental treatments. After the experiment is completed, if arc sin $\sqrt{\bar{p}^*}$ − arc sin $\sqrt{\bar{p}_1} \geqq 2.128/\sqrt{2(174)} = .114$, we conclude that the experimental treatment with observed percentage of success $= \bar{p}^*$ is best, otherwise we conclude that the standard treatment is really better than any of the experimental treatments.

REFERENCES

[1] William Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, 1950.

[2] K. Pearson, *Tables for Statisticians and Biometricians*, Part II, 1st ed., Cambridge University Press, 1931.

[3] Charles Stein, "A two sample test for a linear hypothesis whose power is independent of the variance," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 253–258.

[4] E. Paulson and W. A. Wallis, "Planning and analyzing experiments involving two percentages," *Selected Techniques of Statistical Analysis*, McGraw-Hill Book Co., pp. 247–265.