A SINGLE-SAMPLE MULTIPLE DECISION PROCEDURE FOR RANKING MEANS OF NORMAL POPULATIONS WITH KNOWN VARIANCES[1]

By Robert E. Bechhofer

Cornell University[2]

Summary. This paper is concerned with a single-sample multiple decision procedure for ranking means of normal populations with known variances. Problems which conventionally are handled by the analysis of variance (Model I) which tests the hypothesis that k means are equal are reformulated as multiple decision procedures involving rankings. It is shown how to design experiments so that useful statements can be made concerning these rankings on the basis of a predetermined number of independent observations taken from each population. The number of observations required is determined by the desired probability of a correct ranking when certain differences between population means are specified.

1. Introduction. In many of the experimental situations to which tests of homogeneity conventionally are applied, such as the F-test that k population means are equal, or Bartlett's test that k population variances are equal, the tests (whether or not they yield statistically significant results) do not supply the information that the experimenter seeks. Thus in an agricultural problem the hypothesis that several essentially *different* varieties of grain have the *same* (population) mean yield is an unrealistic one since it is obvious that if the varieties actually are different, the (population) mean yields also will be different, and a sufficiently large sample will establish this fact at any preassigned level of significance. Moreover, should a significant result be obtained, the experimenter's problems usually have just begun. For having established that the varieties are different he may now desire to select the one which is "best." Here the best variety might be defined as the one having the *largest* (population) mean yield. Whenever the experimenter ultimately is faced by the prospect of having to choose a *best* variety, it seems reasonable that the experiment should have been designed with this outcome in mind. What is needed then is a decision procedure which will tell the experimenter which population or populations to choose, and an operating characteristic which will tell him the probability of his making a correct choice if he follows the given decision procedure. The experiment then should be so designed as to control (in some sense) this probability at some preassigned level.

Although the formulation of the problem as outlined above appears to be a reasonable one, little work along these lines has appeared in the literature. In

Received 2/6/52, revised 8/7/53.

[1] Based on research sponsored by the Office of Naval Research.

[2] The author initiated research on this project while he was at Columbia University.

this connection three papers by Paulson [12], [13], [14], all involving multiple decision procedures, deserve special mention. In the first he considers the $2^k - 1$ decision problem of dividing a set of k population means into a "superior" and an "inferior" group; in the second he considers the k decision problem of determining the "best" of k populations when comparing $k - 1$ experimental populations with a control population; in the third he finds an optimum solution to a $k + 1$ decision "slippage" problem of Mosteller [10]. Duncan [5], [6] has considered multiple decision procedures involving means. (It is not clear what kind of over-all confidence statement (with stated confidence coefficient) the experimenter can make if he uses Duncan's procedure.) Tukey [19] and Scheffé [16] have proposed very interesting alternate formulations of the analysis of variance problem; they are concerned with making multiple comparisons among the means.

The principal results of the present paper deal with a single-sample method of designing experiments to determine the ranking of k normal populations where the true ranking, concerning which information is sought, is based on the population means; in a later paper the writer intends to treat the similar situation where the true ranking, concerning which information is sought, is based on the population variances.

2. The test of homogeneity (analysis of variance) approach. The classical test procedure known as the analysis of variance was introduced by Fisher [8], [9] as a method of analyzing certain types of complex experiments and since has become one of the basic tools of the practicing statistician. At the time of its introduction the procedure represented a considerable contribution to the then available body of statistical techniques. Perhaps its greatest accomplishment lay in the fact that it stressed to the experimenter the principle of orthogonality—a principle which if carefully adhered to would permit him to extract from complex experiments considerable information concerning the effects of each of the factors that entered into the experiment; this same principle of orthogonality made it possible for him to test, without difficulty, hypotheses concerning the existence of these effects.

It has been recognized by many statisticians that the analysis of variance has certain deficiencies. However, these deficiencies do not lie in the design aspects of the procedure, but rather in the types of decisions which are made on the basis of the data. The substantial contribution of experimental design (in the Fisherian sense) to the planning of a meaningful experiment cannot be overemphasized. However, there seems to be considerable doubt as to the utility of the tests of hypotheses which usually are the end products of any analysis of variance. Cochran and Cox in their excellent book *Experimental Designs* point out that "On the whole . . . tests of significance are less frequently useful in experimental work than confidence limits. In many experiments it seems obvious that the different treatments must have produced some difference, however small in effect. Thus the hypothesis that there is no difference is unrealistic: The real problem is to obtain estimates of the sizes of the differences."

However, in many instances the purpose of the experiment is not to estimate the sizes of differences, but rather to find the "best" treatment or treatments. The method of estimating the sizes of differences often is used as a way of attempting, indirectly, to achieve this goal. The method described in the next sections is a direct approach to a solution of the ranking problem.

In these sections we shall assume the same underlying mathematical model as is assumed for the analysis of variance (Model I). However, we shall reformulate the purpose of our observation-taking. Instead of being interested in testing hypotheses that population means are equal, we shall be interested in making certain inferences concerning the ranking of these population means. It is important to emphasize, however, that in this ranking approach, experimental designs such as randomized blocks, Latin squares, etc., will play the same role as they do in the analysis of variance.

3. The ranking (multiple decision) approach: the one-way classification.

A. *Statement of the problem.* Let X_{ij} be normally and independently distributed chance variables $N(X_{ij} \mid \mu_i, \sigma_i^2)$, $(i = 1, 2, \cdots, k; j = 1, 2, \cdots, N_i)$. We assume that the μ_i are unknown; the σ_i^2 are known and may be equal or unequal. Let $\mu_{[1]} \leqq \mu_{[2]} \leqq \cdots \leqq \mu_{[k]}$ be the ranked μ_i ; we assume that it is not known which population is associated with $\mu_{[i]}$ $(i = 1, 2, \cdots,)k$.

We further assume that a population is characterized by its population mean, the "best" population being the one having the largest mean, the "second best" being the one having the second largest mean, etc. (Alternatively, we might have defined the "best" population as being the one having the smallest mean, etc.; however, the mathematical theory is the same for both cases.) Thus, the k populations might be k different varieties of grain, and μ_i might be the (population) mean yield per acre of the ith variety. We would like on the basis of a sample of $N = \sum_{i=1}^{k} N_i$ independent observations to make some inference about the "bestness" of the populations. (This statement will be made precise later.)

Our inferences will be based on the sample means. The sample mean from the ith population will be denoted by $\bar{X}_i$. (For the sake of simplicity, no attempt will be made in this paper to distinguish notationally between chance variables and their observed values.) The sample mean, population variance, and sample size associated with the population having population mean $\mu_{[i]}$ will be denoted by $\bar{X}_{(i)}$, $\sigma_{(i)}^2$, and $N_{(i)}$, respectively, $(i = 1, 2, \cdots, k)$; that is, the expected value of $\bar{X}_{(i)}$ is $\mu_{[i]}$ and the variance of $\bar{X}_{(i)}$ is $\sigma_{(i)}^2/N_{(i)}$. The ranked $\bar{X}_i$ will be denoted by

$$\bar{X}_{[1]} < \bar{X}_{[2]} < \cdots < \bar{X}_{[k]} . \tag{1}$$

The event $\bar{X}_i = \bar{X}_j (i \neq j)$ is an event of probability zero and can be ignored in probability calculations. However, in experimental situations this event can occur frequently because of the limitations of the measuring instrument. If it does occur, the tied means should be "ranked" using a randomized procedure which assigns equal probability to each ordering.

In order to apply our procedures, the experimenter must decide what his goal is *before* he takes his sample. For example, his goal may be to find any one of the following (or others unlisted):

(2) The "best" population.

(3) The "best two" populations *without* regard to order.

(4) The "best two" populations *with* regard to order.

(5) The "best three" populations *without* regard to order.

The choice of a goal will depend on economic and other considerations outside the control of the statistician.

Having chosen a goal the statistical procedure is elementary. We take N_i observations from the ith population ($i = 1, 2, \cdots, k$). We compute the k sample means $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_k$. We make the ranking (1). We then take action as follows. If our goal is to find (2), we make the statement, "The population associated with $\bar{X}_{[k]}$ is the 'best' population." If our goal is to find (4), we make the statement, "The populations associated with $\bar{X}_{[k]}$ and $\bar{X}_{[k-1]}$ are the 'best' and 'second best' populations, respectively." If our goal is to find (3), (5), etc., we make similar statements.

For fixed values of the μ_i and σ_i^2 ($i = 1, 2, \cdots, k$) the proportion of *correct* statements that we make will depend only on the N_i, but the proportion will differ, of course, for each kind of statement. We propose to design the experiment in such a way (that is, choose the N_i in such a way) that under specified conditions the proportion of correct statements associated with our decision procedure will be equal to or greater than some preassigned value.

B. *Expressions for the probabilities.* The required probabilities of a correct ranking can be expressed in two basically different forms, as volumes under multivariate normal surfaces, or as iterated integrals. We shall consider both forms. In order to do so we first must state our goal.

A general goal for the *one-way classification* of means can be expressed as follows. (See also (24)). To find

(6) The k_s "best" populations, the k_{s-1} "second best" populations, the k_{s-2} "third best" populations, etc., and finally the k_1 "worst" populations.

Here $k_1, k_2, \cdots, k_s (s \leqq k)$ are positive integers such that $\sum_{i=1}^{s} k_i = k$.

The probability of a correct ranking associated with (6) can be written as:

$$
\begin{aligned}
(7)\qquad & \Pr[\max\{\bar{X}_{(1)}, \cdots, \bar{X}_{(k_1)}\} < \min\{\bar{X}_{(k_1+1)}, \cdots, \bar{X}_{(k_1+k_2)}\}, \\
& \quad \max\{\bar{X}_{(k_1+1)}, \cdots, \bar{X}_{(k_1+k_2)}\} < \min\{\bar{X}_{(k_1+k_2+1)}, \cdots, \bar{X}_{(k_1+k_2+k_3)}\}, \\
& \quad \vdots \\
& \quad \max\{\bar{X}_{(k-k_s-k_{s-1}+1)}, \cdots, \bar{X}_{(k-k_s)}\} < \min\{\bar{X}_{(k-k_s+1)}, \cdots, \bar{X}_{(k)}\}].
\end{aligned}
$$

If we assign particular values to s and the k_i we obtain several special cases of interest, two of which we shall consider in some detail to illustrate the method. For example, for $s = 2$; $k_1 = k - t$; $k_2 = t$, we have

$$\text{(8)} \qquad \Pr[\max \{\bar{X}_{(1)}, \bar{X}_{(2)}, \cdots, \bar{X}_{(k-t)}\} < \min \{\bar{X}_{(k-t+1)}, \cdots, \bar{X}_{(k)}\}]$$

and for $s = k$; $k_1 = k_2 = \cdots = k_k = 1$, we have

$$\text{(9)} \qquad \Pr[\bar{X}_{(1)} < \bar{X}_{(2)} < \cdots < \bar{X}_{(k-1)} < \bar{X}_{(k)}].$$

Stated in words, (8) is the probability that the "best t" populations will yield the largest sample means; thus (8) for $t = 1, 2, 3$ is the probability of a correct ranking associated with (2), (3), (5), respectively.

If we consider (8) for $t = 1$ we have

$$\text{(10)} \quad \Pr[\max \{\bar{X}_{(1)}, \bar{X}_{(2)}, \cdots, \bar{X}_{(k-1)}\} < \bar{X}_{(k)}] = \Pr[0 < Y_1, 0 < Y_2, \cdots, 0 < Y_{k-1}]$$

where $Y_i = \bar{X}_{(k)} - \bar{X}_{(i)}$ $(i = 1, 2, \cdots, k-1)$. Then $E(Y_i) = \mu_{[k]} - \mu_{[i]} = \delta_{k,i}$ (say)

$$\sigma^2(Y_i) = \frac{\sigma^2_{(k)}}{N_{(k)}} + \frac{\sigma^2_{(i)}}{N_{(i)}} \qquad \text{for } (i = 1, 2, \cdots, k-1)$$

$$\sigma(Y_i Y_j) = \frac{\sigma^2_{(k)}}{N_{(k)}} \qquad \text{for } i \neq j (i, j, = 1, 2, \cdots, k-1),$$

and the Y_i have a $(k-1)$ − variate normal distribution.

If we denote the covariance matrix of the Y_i by $\mathfrak{Z}$, and denote the row vectors $(y_1, y_2, \cdots, y_{k-1})$ and $(\delta_{k,1}, \delta_{k,2}, \cdots, \delta_{k,k-1})$ by y' and δ', respectively, then (10) is given by

$$\text{(11)} \qquad \frac{|\mathfrak{Z}|^{-\frac{1}{2}}}{(2\pi)^{(k-1)/2}} \int_0^\infty \int_0^\infty \cdots \int_0^\infty e^{-\frac{1}{2}(y-\delta)'\mathfrak{Z}^{-1}(y-\delta)}\, dy_1\, dy_2 \cdots dy_{k-1}.$$

If all of the means have the same variance, that is, if

$$\text{(12)} \qquad \sigma^2_{(i)}/N_{(i)} = \sigma^2_{\bar{x}} \text{ (say)}, \qquad (i = 1, 2, \cdots, k),$$

then (10) is equal to

$$\text{(13)} \quad \frac{k^{-\frac{1}{2}}}{\pi^{(k-1)/2}} \int_{(-\delta_{k,k-1})/\sqrt{2}\sigma_{\bar{x}}}^{+\infty} \int_{(-\delta_{k,k-2})/\sqrt{2}\sigma_{\bar{x}}}^{+\infty} \cdots \int_{(-\delta_{k,1})\sqrt{2}\sigma_{\bar{x}}}^{+\infty} e^{-\frac{1}{2}x' P_1^{-1} x}\, dx_1 dx_2 \cdots dx_{k-1}$$

where $P_1 = \{\rho_{ij}\}$ is the $k-1$ by $k-1$ correlation matrix with

$$\text{(14)} \qquad \rho_{ij} = \begin{cases} 1 \text{ for } i = j \\ \frac{1}{2} \text{ for } i \neq j \end{cases} \qquad (i, j = 1, 2, \cdots, k-1)$$

and x' denotes the row vector $(x_1, x_2, \cdots, x_{k-1})$. Similarly, under condition (12), (9) is equal to

$$(15)\quad \frac{k^{-\frac{1}{2}}}{\pi^{(k-1)/2}} \int_{(-\delta_{k,k-1})/\sqrt{2}\sigma_{\bar{X}}}^{+\infty} \int_{(-\delta_{k-1,k-2})\sqrt{2}\sigma_{\bar{X}}}^{+\infty} \cdots \int_{(-\delta_{2,1})/\sqrt{2}\sigma_{\bar{X}}}^{+\infty} e^{-\frac{1}{2}z' P_2^{-1} z}\, dz_1\, dz_2 \cdots dz_{k-1}$$

where $P_2 = \{\rho_{ij}\}$ is the $k-1$ by $k-1$ correlation matrix with

$$(16)\quad \rho_{ij} = \begin{cases} 1 & \text{for} \quad i = j \\ -\frac{1}{2} & \text{for} \quad |i-j| = 1 \\ 0 & \text{for} \quad |i-j| > 1 \end{cases} \qquad (i, j = 1, 2, \cdots k-1)$$

and z' denotes the row vector $(z_1, z_2, \cdots, z_{k-1})$.

The probability (7) always is expressible as a sum of integrals of the form (11), and if (12) is true, each of these is reducible to integrals of the form (13) and (15). These integrals cannot be evaluated in finite terms, and the precise determination of probabilities would in general require special tables. However, for $k = 2$ the probabilities (13) and (15) are identical, simply being areas under univariate normal curves; Eisenhart [7] has tabulated unity minus these probabilities as a function of $\delta_{2,1}/\sigma$ and N for the special case $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $N_1 = N_2 = N$. For $k = 3$ the probabilities (13) and (15) are volumes under bivariate normal surfaces with correlation coefficients $\rho = +\frac{1}{2}$ and $\rho = -\frac{1}{2}$, respectively, and can be determined using [3] or [15]. For $4 \leqq k \leqq 10$ the probabilities (13) can be determined using [11]; for $k \geqq 4$ the probability (15) would require special tables which have not yet been prepared. (For related tables see [11] and Section 3D, and the tables at the end of this paper.)

In the first part of this section the probabilities (8) and (9) were expressed as volumes under multivariate normal surfaces. These probabilities also can be expressed as iterated integrals and for certain purposes the latter form is more convenient. We shall illustrate the method for the probability (8) but shall do so only for the special case

$$(17)\quad \sigma_i^2 = \sigma^2, N_i = N \qquad (i = 1, 2, \cdots, k)$$

which guarantees (12), and

$$(18)\quad \begin{aligned} \mu_{[k]} - \mu_{[k-t+1]} &= 0 \\ \mu_{[k-t+1]} - \mu_{[k-t]} &= \delta \text{ (say)} \\ \mu_{[k-t]} - \mu_{[1]} &= 0. \end{aligned}$$

(In Section 3C we shall refer to condition (18) as "the least favorable configuration of the population means.")

When (17) and (18) hold we can write (8) as

$$t \Pr[\max\{\bar{X}_{(1)}, \cdots, \bar{X}_{(k-t)}\} < \bar{X}_{(k-t+1)} < \min\{\bar{X}_{(k-t+2)}, \cdots, \bar{X}_{(k)}\}]$$

$$= t \int_{-\infty}^{+\infty} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\bar{X}_{(k-t+1)} - \mu_{[k-t]})/(\sigma/\sqrt{N})} e^{-z^2/2}\, dz \right]^{k-t}$$

$$(19)\quad \left[\frac{1}{\sqrt{2\pi}} \int_{(\bar{X}_{(k-t+1)} - \mu_{[k-t+1]})/(\sigma/\sqrt{N})}^{+\infty} e^{-z^2/2}\, dz \right]^{t-1}$$

$$\frac{\sqrt{N}}{\sigma\sqrt{2\pi}} e^{-N(\bar{X}_{(k-t+1)} - \mu_{[k-t+1]})^2/2\sigma^2}\, d\bar{X}_{(k-t+1)}$$

which, after making the transformation $y = \sqrt{N}(\bar{X}_{(k-t+1)} - \mu_{[k-t+1]})/\sigma$, yields[3]

$$t \int_{-\infty}^{+\infty} [F(y + d)]^{k-t}[1 - F(y)]^{t-1} f(y)\, dy \tag{20}$$

where

$$F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2}\, dx,\ F'(y) = f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{21}$$

and

$$d = \frac{\sqrt{N}\delta}{\sigma} = \sqrt{N}\lambda(\text{say}) \tag{22}$$

where λ is the standardized difference between the population means. If $d = 0$, the probability (20) is equal to

$$\frac{t!(k - t)!}{k!} = \frac{1}{C_t^k}. \tag{23}$$

The integral (20) is easier to evaluate numerically than the integral (13) even though the former is a more general case of (8) than the latter.

A more general goal than (6) also can be formulated. As an example of a special case of this we shall find

$$t_0 \text{ of the } t \text{ "best" populations} \qquad (1 \leqq t_0 \leqq t). \tag{24}$$

For (24) the probability of a correct ranking is given by

$$\frac{t!}{(t - t_0)!(t_0 - 1)!} \int_{-\infty}^{+\infty} [F(y + d)]^{k-t}[F(y)]^{t-t_0}[1 - F(y)]^{t_0-1} f(y)\, dy \tag{25}$$

under the assumptions (17) and (18). We note that (25) reduces to (20) for $t_0 = t$. In certain situations the experimenter may be willing to relax his requirements and specify (24) as his goal[4] rather than the corresponding case of (6).

C. *Determination of the sample sizes.* The "distances" of the k populations from each other can be expressed in terms of the $k - 1$ parameters

$$\delta_{i+1,i} = \mu_{[i+1]} - \mu_{[i]} \qquad (i = 1, 2, \cdots, k - 1). \tag{26}$$

To simplify notation let

$$\textstyle\sum_{j=1}^{i} k_j = \hat{k}_i \text{ (say)}. \tag{27}$$

Then we note that of the $k - 1$ parameters (26), $s - 1$ of them, namely,

$$\delta_{\hat{k}_i+1,\hat{k}_i} \qquad (i = 1, 2, \cdots, s - 1) \tag{28}$$

exercise a general over-all control on the probability (7), since if the parameters (28) are "small" the probability (7) is relatively low and if they are "large" the probability (7) is high. (For example, for (8) the parameter $\delta_{k-t+1,k-t}$ is the

[3] Paulson obtained this integral for the special case $t = 1$; see equations (2.2) in [12] and (2) in [13].

[4] The formulation of the problem as given in this paragraph was suggested to the writer by Dr. Milton Sobel who also derived the expression (25).

controlling one. If this parameter is made arbitrarily small, the probability (8) lies between (23) and one-half; by increasing this parameter, the probability (8) can be made arbitrarily close to unity.) It is obvious that for *fixed nonzero* values of the parameters (28), and for *fixed* population variances σ_i^2, the probability (7) can be made arbitrarily close to unity by making the N_i $(i = 1, 2, \cdots, k)$ sufficiently large. But if one or more of the $s - 1$ parameters (28) is very small, the $N = \sum_{i=1}^{k} N_i$ required to realize any probability close to unity will be extremely large.

Now in most experimental situations there seems to be little if any reason for attempting to differentiate between any pair of populations characterized by μ_{k_i} and $\mu_{k_{i+1}}$ if the corresponding parameter (28) is very small since the expense involved in guaranteeing a high probability of a correct ranking may be prohibitive and/or the economic loss involved in making an incorrect ranking may be negligible. In fact, in most situations it should be possible to specify $s - 1$ constants

$$\delta^*_{k_{i+1},k_i} \qquad (i = 1, 2, \cdots, s - 1) \tag{29}$$

which are the smallest values of the parameters (28) which are "worth detecting." We shall assume that these are given in what follows.

It is our purpose to find the smallest $N = \sum_{i=1}^{k} N_i$ which will guarantee a specified probability $\gamma < 1$ of a correct ranking whenever $\delta_{k_{i+1},k_i} \geqq \delta^*_{k_{i+1},k_i}$ $(i = 1, 2, \cdots, s - 1)$. As a device for doing this we consider the *least favorable configuration of the population means.* This configuration is defined as being the one which, for fixed N_i and σ_i^2 $(i = 1, 2, \cdots, k)$, yields the greatest lower bound of the probability of a correct ranking. Since the probability (7) is a strictly increasing function of each of the parameters $\delta_{i+1,i}$ $(i = 1, 2, \cdots, k - 1)$, it is easy to see that the greatest lower bound is achieved when each of the $s - 1$ parameters given by (28) has the corresponding parameter values given by (29), and each of the $k - s$ parameters given by (26) but not given by (28) has the value zero. The desired $N = \sum_{i=1}^{k} N_i$ is then the smallest one which will guarantee the probability γ for the least favorable configuration. Of course, the efficient choice of the N_i will depend on the σ_i^2 $(i = 1, 2, \cdots, k)$. For fixed N_i and σ_i^2 $(i = 1, 2, \cdots, k)$ the probability (7) considered as a function of the $\delta_{i+1,i}$ $(i = 1, 2, \cdots, k - 1)$ gives an analogue of power and might be termed the *operating characteristic curve* with respect to a correct ranking for the procedure.

(i) *Variances known and equal.* If $\sigma_i^2 = \sigma^2$ $(i = 1, 2, \cdots, k)$ where σ^2 is a known constant then it would appear to be most efficient to choose *equal* sample sizes from each population. (In this and the next two subsections, the most efficient allocation of the sample sizes will be defined as the one which, for fixed total sample size, maximizes the minimum probability of a correct ranking. The writer does not claim at this time that all of the procedures described in this paper are most efficient.) We choose the common sample size N' in such a way that $N = \sum_{i=1}^{k} N_i = kN'$ is the smallest integer which will guarantee γ for the least favorable configuration. The probability (7) then will prove to be a function of only the k_i $(i = 1, 2, \cdots, s)$ and the $s - 1$ constants

$$\sqrt{\frac{N'}{2}}\,\frac{\delta^*_{k_{i+1},k_i}}{\sigma} \qquad (i = 1, 2, \cdots, s-1). \tag{30}$$

For example, for (8) we have $s = 2$ and

$$\sqrt{\frac{N'}{2}}\,\frac{\delta^*_{k_1+t,k_1}}{\sigma} = \sqrt{\frac{N'}{2}}\,\frac{\delta^*_{k-t+1,k-t}}{\sigma}. \tag{31}$$

(ii) *Variances known and unequal.* If $\sigma_i^2 = a_i\sigma^2 (i = 1, 2, \cdots, k)$, where σ^2 is a known constant, and the a_i are known constants not all of which are equal to unity, then it may be desirable to choose the sample sizes so that the variances of the sample means are equal. This choice is *not* most efficient. (If we restrict our attention to procedures for which the sample sizes are taken to make the population variances of the sample means equal, then it can be shown the minimax procedure for (2) is: "select as the 'best' population the one having the largest sample mean.") However, it has a very important *practical* advantage; namely, that the tables which give the probability of a correct ranking for the special case $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ and $N_1 = N_2 = \cdots = N_k$ then become applicable. In order to apply these tables when the $N_i (i = 1, 2, \cdots, k)$ are subject to the restriction

$$\frac{\sigma_1^2}{N_1} = \frac{\sigma_2^2}{N_2} = \cdots = \frac{\sigma_k^2}{N_k} \tag{32}$$

we proceed as follows. We act as if the k populations had the common variance σ^2 which is the known constant referred to above. Using the method of the previous section, we find $N = kN'$ where N' is the number of observations taken from each population. We then set

$$\frac{\sigma_i^2}{N_i} = \frac{a_i}{N_i}\sigma^2 = \frac{\sigma^2}{N'} \qquad (i = 1, 2, \cdots, k) \tag{33}$$

from which it follows that we choose the individual N_i so that

$$N_i = a_i N' \qquad (i = 1, 2, \cdots, k). \tag{34}$$

If any N_i so chosen is not an integer, we replace it by the next largest integer. Because of (33), it is clear that these N_i guarantee γ.

As was indicated above, for fixed total sample size (34) does *not* define the most efficient choice of the N_i for arbitrary a_i. For example, for $k = 2$ it can be shown that the most efficient method of choosing N_1 and N_2 is to select that pair (N_1, N_2) which satisfies the equations $N_2\sigma_1 = N_1\sigma_2$ and $N_1 + N_2 = N$ where σ_1 and σ_2 are known and N is specified. For $k > 2$, the rule by which a most efficient choice of the $N_i (i = 1, 2, \cdots, k)$ is made appears to be too complicated for practical application; also, the number of tables needed would be prohibitively large.

(iii) *Variances unknown.* If the values of the $\sigma_i^2 (i = 1, 2, \cdots, k)$ are completely unknown (or even if it is known that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ where the common

value of the variances is unknown), it is not possible using a one-sample procedure to make any useful statement concerning the magnitude of the confidence coefficient. (Actually the experimenter faces the same type of dilemma here as he does when he desires to make a statement about the power of an analysis of variance test when the variance is unknown.) For this problem an analogue of Stein's two-sample procedure [17] has been developed to provide a solution of the ranking problem. This new procedure will be described in a forthcoming paper. (See [1].)

D. *Discussion of tables.* Tables have been prepared to assist the experimenter in designing and interpreting experiments for ranking means.

Table I is to be used for designing experiments involving k normal populations to decide which t have the largest (or smallest) population means. The table is based on probability (20), and gives the value of $d = \sqrt{N}\lambda$ associated with the probabilities 0.05 (0.05) 0.80 (0.02) 0.90 (0.01) 0.99, 0.995, 0.999, and 0.9995 for $k = 2(1)\ 10$ and $t = 1(1)\ [k/2]$ (as well as $k = 11(1)\ 15$ and selected values of t) where $[k/2]$ $k = 1, 2, \cdots, 15$ is the largest integer less than or equal to $k/2$. The table is based on the least favorable configuration of the population means which, for picking the t *largest*, is given by $\mu_{[k]} - \mu_{[k-t+1]} = 0$, $\mu_{[k-t+1]} - \mu_{[k-t]} = \lambda\sigma$, $\mu_{[k-t]} - \mu_{[1]} = 0$, and for picking the t *smallest* is given by the same expressions with t replaced by $k - t$. The values of d were obtained by inverse linear interpolation in [11].[5]

Table II is a special table to be used for designing experiments involving 3 normal populations to decide which one has the largest, which the second largest, and which the smallest population mean. The table is based on probability (15) for $k = 3$, and gives the value of $d = \sqrt{N}\lambda$ associated with the probabilities $\frac{1}{6}$, 0.20 (0.05) 0.80 (0.02) 0.90 (0.01) 0.99. The least favorable configuration $\mu_{[3]} - \mu_{[2]} = \mu_{[2]} - \mu_{[1]} = \lambda\sigma$ is assumed throughout. The values of d were obtained by inverse interpolation in [3] and [15]. For convenience of tabulation the standardized differences between the population means were taken as equal. A table for unequal differences could be prepared using [15].

Examples of the use of the tables are given in Section 8.

4. The ranking (multiple decision) approach: the two-way classification without interaction.

A. *Statement of the problem.* Let X_{ijm} be normally and independently distributed chance variables $N(X_{ijm} \mid \mu + \alpha_i + \beta_j, \sigma_{ij}^2)$, $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, c; m = 1, 2, \cdots, N_{ij})$, with $\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = 0$. We assume that μ, the α_i, and the β_j are unknown; the σ_{ij}^2 are known and may be equal or unequal. Let $\alpha_{[1]} \leqq \alpha_{[2]} \leqq \cdots \leqq \alpha_{[r]}$ and $\beta_{[1]} \leqq \beta_{[2]} \leqq \cdots \leqq \beta_{[c]}$ be the ranked α_i and β_j, respectively; we assume that it is not known which populations are associated with either $\alpha_{[i]}$ or $\beta_{[j]}$.

---

[5] These tables were computed for this project by the National Bureau of Standards, at the Institute for Numerical Analysis, Los Angeles; the computations were supported by the Office of Naval Research.

As in Section 3A we assume that a population is characterized by its population mean which for the two-way classification (no interaction) consists of two components of interest each one of which measures a classification "effect." The "best" *set* of populations with respect to the first classification is the one consisting of those populations having population means $\mu + \alpha_{[r]} + \beta_j$ $(j = 1, 2, \cdots, c)$; the "best" *set* of populations with respect to the second classification is the one consisting in those populations having population means $\mu + \alpha_i + \beta_{[c]}$ $(i = 1, 2, \cdots, r)$; the "second best," etc. sets of populations with respect to either the first or the second classifications are defined in the obvious way. Thus the rc populations might be the rc combinations of r different varieties of grain and c different types of fertilizer, and $\mu + \alpha_i + \beta_j$ might be the (population) mean yield per acre of the ith variety treated with the jth fertilizer. (Here we are assuming no variety-fertilizer interaction.) We would like on the basis of a sample of $N = \sum_{i=1}^{r} \sum_{j=1}^{c} N_{ij}$ independent observations to make inferences about the "bestness" of the populations for each of the two classifications.

Our inferences will be based on the sample means which will be denoted by

$$\bar{X}_{ij} = \frac{\sum_{m=1}^{N_{ij}} X_{ijm}}{N_{ij}} \qquad (i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c), \tag{35}$$

$$\bar{X}_{i.} = \frac{\sum_{j=1}^{c} \bar{X}_{ij}}{c} \qquad (i = 1, 2, \cdots, r), \tag{36}$$

and

$$\bar{X}_{.j} = \frac{\sum_{i=1}^{r} \bar{X}_{ij}}{r} \qquad (j = 1, 2, \cdots, c). \tag{37}$$

The sample mean, population variance, and sample size associated with the population having population mean $\mu + \alpha_{[i]} + \beta_{[j]}$ will be denoted by $\bar{X}_{(i)(j)}$, $\sigma^2_{(i)(j)}$, and $N_{(i)(j)}$, respectively, $(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, c)$; that is, the expected value of $\bar{X}_{(i)(j)}$ is $\mu + \alpha_{[i]} + \beta_{[j]}$ and the variance of $\bar{X}_{(i)(j)}$ is $\sigma^2_{(i)(j)}/N_{(i)(j)}$.

We also define

$$\bar{X}_{(i).} = \frac{\sum_{j=1}^{c} \bar{X}_{(i)(j)}}{c} \qquad (i = 1, 2, \cdots, r) \tag{38}$$

and

$$\bar{X}_{.(j)} = \frac{\sum_{i=1}^{r} \bar{X}_{(i)(j)}}{r} \qquad (j = 1, 2, \cdots, c). \tag{39}$$

The ranked $\bar{X}_{i.}$ and $\bar{X}_{.j}$ will be denoted by

$$\bar{X}_{[1].} < \bar{X}_{[2].} < \cdots < \bar{X}_{[r].} \tag{40}$$

and

$$\bar{X}_{.[1]} < \bar{X}_{.[2]} < \cdots < \bar{X}_{.[c]}\,, \tag{41}$$

respectively.

Goals for the two-way classification are of the same type as for the one-way classification except that they consist of two parts. For example, the experimenter's goal may be to find any one of the following (or others unlisted):

(42) The "best" set of populations according to the first classification *and* the "best" set of populations according to the second classification.

(43) The "best" set of populations according to the first classification *and* the "best two" sets of populations *without* regard to order according to the second classification.

Having chosen our goal, we take N_{ij} observations from the i, jth population and compute the $r + c$ sample means (36) and (37). We make the rankings (40) and (41). If our goal is to find (42), we make the statement, "The set of populations associated with $\bar{X}_{[r].}$ is the 'best' set of populations according to the first classification *and* the set of populations associated with $\bar{X}_{.[c]}$ is the 'best' set of populations according to the second classification." If our goal is to find (43), etc., we would make similar statements. For fixed values of the α_i, β_j, and σ_{ij}^2 $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, c)$ the proportion of *correct* statements that we make will depend only on the N_{ij}.

B. *Expression for the probabilities.* A general goal for the *two-way classification* is similar to (6) except that it consists of two parts. We shall not write it explicitly nor shall we write the associated probability of a correct ranking. However, to illustrate the method of evaluating such a probability we shall consider the following special case in some detail:

$$\begin{aligned}(44)\quad &\Pr[\bar{X}_{(1).} < \bar{X}_{(2).} < \cdots < \bar{X}_{(r).} \text{ and } \bar{X}_{.(1)} < \bar{X}_{.(2)} < \cdots < \bar{X}_{.(c)}] \\ &= \Pr[0 < Y_1, 0 < Y_2, \cdots, 0 < Y_{r-1} \text{ and } 0 < Z_1, 0 < Z_2, \cdots, 0 < Z_{c-1}]\end{aligned}$$

where

$$\begin{aligned}Y_i &= \bar{X}_{(i+1).} - \bar{X}_{(i).} \qquad (i = 1, 2, \cdots, r-1) \\ Z_j &= \bar{X}_{.(j+1)} - \bar{X}_{.(j)} \qquad (j = 1, 2, \cdots, c-1).\end{aligned} \tag{45}$$

We note that the Y_i and Z_j have a joint $(r + c - 2)$-variate normal distribution. If the means in each cell have the same variance, that is, if

$$\sigma_{ij}^2/N_{ij} = \sigma_{\bar{X}}^2 \text{ (say)} \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, c), \tag{46}$$

then the joint covariance matrix of the Y_i and Z_j simplifies considerably and we have

$$\sigma^2(Y_i) = \frac{2}{c}\sigma_{\bar{X}}^2 \qquad (i = 1, 2, \cdots, r-1)$$

$$\sigma^2(Z_j) = \frac{2}{r}\sigma_{\bar{X}}^2 \qquad (j = 1, 2, \cdots, c-1)$$

$$\sigma(Y_iY_j) = \begin{cases} -\frac{1}{c}\sigma_{\bar{X}}^2 & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases} \qquad (i, j = 1, 2, \cdots, r-1)$$

$$\sigma(Z_iZ_j) = \begin{cases} -\frac{1}{r}\sigma_{\bar{X}}^2 & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases} \qquad (i, j = 1, 2, \cdots, c-1)$$

$$\sigma(Y_iZ_j) = 0 \quad (i = 1, 2, \cdots, r-1;\ j = 1, 2, \cdots, c-1).$$

The fact that $\sigma(Y_iZ_j) = 0$ (all i, j) implies that the Y_i's are *independent* of the Z_j's (a sufficient condition for the Y_i's to be *independent* of the Z_j's is that within every row or every column the variances of the means are equal), and we have that (44) is equal to

$$\Pr[0 < Y_1, \cdots, 0 < Y_{r-1}]\Pr[0 < Z_1, \cdots, 0 < Z_{c-1}]. \tag{48}$$

Thus, if (46) holds, the probability for the two-way classification reduces to the product of the probabilities for two one-way classifications.

It is important to note that if it is desired to increase the first of the probabilities in the product (48), this is accomplished (for fixed c) by decreasing $\sigma_{\bar{X}}^2$ defined by (46), that is, by increasing the N_{ij}. But increasing the N_{ij} also has the effect of increasing the second of the probabilities in the product (48). Thus the *factorial* design of the experiment makes the data "work twice" and is in this sense more efficient than two separate experiments.

5. The ranking (multiple decision) approach: the r-way classification without interaction. For this problem the $X_{i_1i_2\cdots i_r}$ are normally and independently distributed chance variables

$$N(X_{i_1i_2\cdots i_rm} \mid \mu + \textstyle\sum_{j=1}^{r} \alpha_{i_j}, \sigma_{i_1i_2\cdots i_r}^2) \text{ with } \sum_{i_j=1}^{r_j} \alpha_{i_j} = 0,$$
$$(i_j = 1, 2, \cdots, r_j; j = 1, 2, \cdots, r; m = 1, 2, \cdots, N_{i_1i_2\cdots i_r}).$$

We would like on the basis of a sample of $\sum_{j=1}^{r}\sum_{i_j=1}^{r_j} N_{i_1i_2\cdots i_r}$ independent observations to make inferences about the "bestness" of the populations for each of the r classifications. This problem is a straightforward generalization of the case $r = 2$ treated in the previous section.

6. The ranking (multiple decision) approach: experimental designs. Designs such as randomized blocks, Latin squares, etc. are used in experimentation to

eliminate the effects of heterogeneity in one or two directions. Their use results in a reduction in the underlying variance of the experiment, and it therefore is possible to make more precise comparisons among the "treatment effects." These designs serve the same function in the ranking approach as they do in the analysis of variance. We shall illustrate this point with randomized blocks. and the carry-over to other more complex designs will be immediate.

We assume the same mathematical model for the randomized blocks design as for the two-way classification without interaction. However, in this case we are not concerned with the (block) effects β_j since the blocks are introduced only to reduce the σ_{ij}^2. We define our sample means as we did for the two-way classification. Since the expected value of the differences between the treatment (row) means involves only the α_i, our problem is thrown into the form of the one-way classification which we have considered already.

7. Large sample applications of the ranking theory. The results obtained in the previous section can be used to rank parameters other than the population means of normal distributions, provided that sufficiently large samples are available, and the statistics that are used to estimate these parameters are normally distributed in the limit. Since reasonably large sample sizes usually are required to achieve the desired probabilities in ranking problems, and since the central limit theorem applies under very general conditions, many ranking problems can be solved using the already-developed normal theory.

In many problems the approach of the statistic to normality will be accelerated, and the dependence of the mean on the variance will be minimized, if the statistic is appropriately transformed. Thus, for example, the population probabilities of "success" in binomial distributions, or the population means of Poisson distributions can be ranked using the transformations arc sin $\sqrt{\bar{x}}$ and $\sqrt{\bar{x}}$ respectively. Similarly, the population correlation coefficients of bivariate normal distributions can be ranked if the transformation $z = \frac{1}{2} \log_e [(1 + r)/(1 - r)]$ is used.

8. Examples. Several numerical examples will be given here to illustrate the use of the tables. It will be assumed that the mathematical models of Section 3 and Section 4 hold for Examples 1 and 2 and Example 3, respectively. No attempt will be made to relate these examples to any particular subject matter field.

Example 1. Given a one-way classification of three populations. Suppose that it is desired to find which population has the largest mean, and to guarantee that the probability of correctly choosing that population will be at least 0.75 when $\mu_{[3]} - \mu_{[2]} \geqq 4$. How many observations must be taken from each population?

Refer to Table I, column headed $k = 3$, $t = 1$.

a) Suppose that it is known that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2 = 100$. Then we follow the method of Section 3C(i). Entering the table we find that the value of $\sqrt{N}\lambda$ associated with a probability of 0.75 is 1.4338. We have $\lambda = 4/10$. Thus, $0.4\sqrt{N} = 1.4338$, and hence select 13 observations from each population.

TABLE I

Table of $\sqrt{N}\lambda$ corresponding to various probabilities, to be used for designing experiments involving k normal populations to decide which t have the largest (or smallest) population means

Prob. of Correct Ranking	$k = 2$ $t = 1$	$k = 3$ $t = 1$	$k = 4$ $t = 1$	$k = 4$ $t = 1$	$k = 5$ $t = 1$
0.9995	4.6535	4.9163	5.0639	5.1699	5.1661
0.9990	4.3703	4.6450	4.7987	4.9098	4.9049
0.9950	3.6428	3.9517	4.1224	4.2490	4.2394
0.99	3.2900	3.6173	3.7970	3.9323	3.9196
0.98	2.9045	3.2533	3.4432	3.5893	3.5722
0.97	2.6598	3.0232	3.2198	3.3734	3.3529
0.96	2.4759	2.8504	3.0522	3.2117	3.1885
0.95	2.3262	2.7101	2.9162	3.0808	3.0552
0.94	2.1988	2.5909	2.8007	2.9698	2.9419
0.93	2.0871	2.4865	2.6996	2.8728	2.8428
0.92	1.9871	2.3931	2.6092	2.7861	2.7542
0.91	1.8961	2.3082	2.5271	2.7075	2.6737
0.90	1.8124	2.2302	2.4516	2.6353	2.5997
0.88	1.6617	2.0899	2.3159	2.5057	2.4668
0.86	1.5278	1.9655	2.1956	2.3910	2.3489
0.84	1.4064	1.8527	2.0867	2.2873	2.2423
0.82	1.2945	1.7490	1.9865	2.1921	2.1441
0.80	1.1902	1.6524	1.8932	2.1035	2.0528
0.75	0.9539	1.4338	1.6822	1.9038	1.8463
0.70	0.7416	1.2380	1.4933	1.7253	1.6614
0.65	0.5449	1.0568	1.3186	1.5609	1.4905
0.60	0.3583	0.8852	1.1532	1.4055	1.3287
0.55	0.1777	0.7194	0.9936	1.2559	1.1726
0.50	0.0000	0.5565	0.8368	1.1093	1.0193
0.45		0.3939	0.6803	0.9633	0.8662
0.40		0.2289	0.5215	0.8156	0.7111
0.35		0.0585	0.3578	0.6635	0.5510
0.30			0.1855	0.5039	0.3827
0.25			0.0000	0.3325	0.2014
0.20				0.1424	0.0000
0.15					
0.10					
0.05					

TABLE I—*Continued*

Prob. of Correct Ranking	$k = 5$ $t = 2$	$k = 6$ $t = 1$	$k = 6$ $t = 2$	$k = 6$ $t = 3$	$k = 7$ $t = 1$
0.9995	5.3127	5.2439	5.4116	5.4529	5.3066
0.9990	5.0584	4.9856	5.1611	5.2043	5.0505
0.9950	4.4138	4.3280	4.5270	4.5756	4.3989
0.99	4.1058	4.0121	4.2244	4.2760	4.0861
0.98	3.7728	3.6692	3.8977	3.9530	3.7466
0.97	3.5635	3.4528	3.6925	3.7504	3.5324
0.96	3.4071	3.2906	3.5393	3.5992	3.3719
0.95	3.2805	3.1591	3.4154	3.4769	3.2417
0.94	3.1732	3.0474	3.3104	3.3735	3.1311
0.93	3.0795	2.9496	3.2187	3.2831	3.0344
0.92	2.9959	2.8623	3.1370	3.2026	2.9479
0.91	2.9201	2.7829	3.0628	3.1296	2.8694
0.90	2.8505	2.7100	2.9948	3.0627	2.7972
0.88	2.7257	2.5789	2.8729	2.9427	2.6676
0.86	2.6153	2.4627	2.7651	2.8368	2.5527
0.84	2.5156	2.3576	2.6677	2.7411	2.4486
0.82	2.4241	2.2609	2.5784	2.6535	2.3530
0.80	2.3391	2.1709	2.4955	2.5720	2.2639
0.75	2.1474	1.9674	2.3086	2.3887	2.0626
0.70	1.9765	1.7852	2.1421	2.2256	1.8824
0.65	1.8191	1.6168	1.9888	2.0756	1.7159
0.60	1.6706	1.4575	1.8443	1.9342	1.5583
0.55	1.5277	1.3037	1.7054	1.7985	1.4062
0.50	1.3879	1.1526	1.5694	1.6657	1.2568
0.45	1.2488	1.0019	1.4343	1.5339	1.1078
0.40	1.1081	0.8491	1.2977	1.4007	0.9567
0.35	0.9635	0.6915	1.1573	1.2640	0.8008
0.30	0.8119	0.5257	1.0103	1.1209	0.6369
0.25	0.6492	0.3472	0.8525	0.9675	0.4604
0.20	0.4691	0.1489	0.6780	0.7979	0.2643
0.15	0.2605		0.4760	0.6019	0.0364
0.10	0.0000		0.2239	0.3576	
0.05				0.0000	

TABLE I—*Continued*

Prob. of Correct Ranking	$k = 7$ $t = 2$	$k = 7$ $t = 3$	$k = 8$ $t = 1$	$k = 8$ $t = 2$	$k = 8$ $t = 3$
0.9995	5.4871	5.5501	5.3590	5.5480	5.6244
0.9990	5.2393	5.3052	5.1047	5.3023	5.3821
0.9950	4.6127	4.6867	4.4579	4.6815	4.7710
0.99	4.3140	4.3926	4.1475	4.3858	4.4807
0.98	3.9917	4.0758	3.8107	4.0669	4.1683
0.97	3.7895	3.8773	3.5982	3.8668	3.9728
0.96	3.6385	3.7293	3.4390	3.7175	3.8270
0.95	3.5164	3.6097	3.3099	3.5968	3.7093
0.94	3.4130	3.5086	3.2002	3.4946	3.6097
0.93	3.3228	3.4203	3.1043	3.4054	3.5229
0.92	3.2423	3.3417	3.0186	3.3258	3.4456
0.91	3.1693	3.2704	2.9407	3.2537	3.3755
0.90	3.1024	3.2051	2.8691	3.1876	3.3113
0.88	2.9824	3.0880	2.7406	3.0691	3.1963
0.86	2.8764	2.9847	2.6266	2.9644	3.0948
0.84	2.7806	2.8915	2.5235	2.8698	3.0032
0.82	2.6929	2.8061	2.4286	2.7832	2.9194
0.80	2.6113	2.7269	2.3403	2.7027	2.8416
0.75	2.4277	2.5485	2.1407	2.5215	2.6666
0.70	2.2641	2.3899	1.9621	2.3601	2.5111
0.65	2.1137	2.2442	1.7970	2.2116	2.3683
0.60	1.9719	2.1071	1.6407	2.0718	2.2340
0.55	1.8355	1.9754	1.4899	1.9374	2.1051
0.50	1.7022	1.8468	1.3418	1.8059	1.9792
0.45	1.5697	1.7191	1.1941	1.6753	1.8543
0.40	1.4358	1.5903	1.0443	1.5434	1.7284
0.35	1.2982	1.4581	0.8897	1.4079	1.5992
0.30	1.1542	1.3198	0.7272	1.2660	1.4641
0.25	0.9997	1.1717	0.5523	1.1139	1.3195
0.20	0.8288	1.0081	0.3579	0.9457	1.1599
0.15	0.6312	0.8192	0.1319	0.7511	0.9757
0.10	0.3846	0.5840		0.5085	0.7465
0.05	0.0232	0.2403		0.1530	0.4118

TABLE I—*Continued*

Prob. of Correct Ranking	$k = 8$ $t = 4$	$k = 9$ $t = 1$	$k = 9$ $t = 2$	$k = 9$ $t = 3$	$k = 9$ $t = 4$
0.9995	5.6463	5.4039	5.5988	5.6842	5.7196
0.9990	5.4049	5.1511	5.3550	5.4440	5.4809
0.9950	4.7966	4.5083	4.7388	4.8386	4.8798
0.99	4.5078	4.1999	4.4455	4.5513	4.5950
0.98	4.1972	3.8653	4.1292	4.2423	4.2888
0.97	4.0029	3.6543	3.9308	4.0489	4.0974
0.96	3.8581	3.4961	3.7829	3.9048	3.9548
0.95	3.7412	3.3679	3.6633	3.7885	3.8398
0.94	3.6424	3.2590	3.5620	3.6902	3.7426
0.93	3.5562	3.1637	3.4736	3.6045	3.6579
0.92	3.4794	3.0785	3.3948	3.5280	3.5825
0.91	3.4099	3.0012	3.3234	3.4589	3.5142
0.90	3.3462	2.9301	3.2579	3.3955	3.4516
0.88	3.2322	2.8024	3.1405	3.2820	3.3395
0.86	3.1316	2.6893	3.0368	3.1818	3.2408
0.84	3.0408	2.5868	2.9433	3.0915	3.1518
0.82	2.9577	2.4926	2.8575	3.0088	3.0703
0.80	2.8807	2.4049	2.7778	2.9321	2.9947
0.75	2.7074	2.2067	2.5984	2.7596	2.8249
0.70	2.5535	2.0293	2.4387	2.6064	2.6741
0.65	2.4122	1.8653	2.2919	2.4658	2.5359
0.60	2.2794	1.7102	2.1535	2.3335	2.4059
0.55	2.1520	1.5604	2.0206	2.2066	2.2814
0.50	2.0276	1.4133	1.8906	2.0828	2.1598
0.45	1.9042	1.2666	1.7615	1.9599	2.0393
0.40	1.7798	1.1178	1.6311	1.8360	1.9180
0.35	1.6523	0.9643	1.4971	1.7090	1.7936
0.30	1.5191	0.8030	1.3569	1.5763	1.6637
0.25	1.3765	0.6292	1.2065	1.4342	1.5248
0.20	1.2192	0.4361	1.0403	1.2775	1.3716
0.15	1.0377	0.2117	0.8481	1.0966	1.1950
0.10	0.8121		0.6085	0.8717	0.9757
0.05	0.4829		0.2575	0.5435	0.6560

TABLE I—*Continued*

Prob. of Correct Ranking	$k = 10$ $t = 1$	$k = 10$ $t = 2$	$k = 10$ $t = 3$	$k = 10$ $t = 4$	$k = 10$ $t = 5$
0.9995	5.4432	5.6425	5.7343	5.7788	5.7924
0.9990	5.1917	5.4000	5.4958	5.5422	5.5563
0.9950	4.5523	4.7878	4.8950	4.9468	4.9625
0.99	4.2456	4.4964	4.6100	4.6648	4.6814
0.98	3.9128	4.1823	4.3037	4.3619	4.3796
0.97	3.7030	3.9854	4.1120	4.1727	4.1911
0.96	3.5457	3.8385	3.9693	4.0319	4.0509
0.95	3.4182	3.7198	3.8541	3.9184	3.9378
0.94	3.3099	3.6193	3.7567	3.8224	3.8422
0.93	3.2152	3.5316	3.6718	3.7387	3.7589
0.92	3.1305	3.4534	3.5962	3.6643	3.6848
0.91	3.0536	3.3826	3.5277	3.5969	3.6177
0.90	2.9829	3.3176	3.4650	3.5351	3.5563
0.88	2.8560	3.2011	3.3526	3.4246	3.4463
0.86	2.7434	3.0983	3.2535	3.3272	3.3494
0.84	2.6416	3.0055	3.1642	3.2395	3.2621
0.82	2.5479	2.9203	3.0824	3.1591	3.1822
0.80	2.4608	2.8413	3.0065	3.0847	3.1082
0.75	2.2637	2.6635	2.8360	2.9174	2.9419
0.70	2.0873	2.5051	2.6845	2.7690	2.7944
0.65	1.9242	2.3595	2.5456	2.6330	2.6592
0.60	1.7700	2.2224	2.4149	2.5052	2.5322
0.55	1.6210	2.0907	2.2896	2.3827	2.4106
0.50	1.4748	1.9618	2.1673	2.2632	2.2920
0.45	1.3289	1.8339	2.0460	2.1448	2.1744
0.40	1.1810	1.7047	1.9237	2.0256	2.0561
0.35	1.0284	1.5720	1.7984	1.9035	1.9350
0.30	0.8679	1.4330	1.6674	1.7760	1.8085
0.25	0.6951	1.2841	1.5273	1.6398	1.6733
0.20	0.5032	1.1195	1.3727	1.4896	1.5244
0.15	0.2800	0.9292	1.1944	1.3166	1.3529
0.10	0.0000	0.6919	0.9728	1.1017	1.1401
0.05		0.3444	0.6495	0.7889	0.8303

TABLE I—*Continued*

Prob. of Correct Ranking	$k = 11$ $t = 2$	$k = 11$ $t = 3$	$k = 11$ $t = 4$	$k = 11$ $t = 5$	$k = 12$ $t = 3$
0.9995	5.6807	5.7773	5.8284	5.8511	5.8149
0.9990	5.4395	5.5402	5.5934	5.6170	5.5790
0.9950	4.8305	4.9432	5.0025	5.0288	4.9853
0.99	4.5408	4.6602	4.7229	4.7506	4.7039
0.98	4.2286	4.3560	4.4227	4.4522	4.4016
0.97	4.0329	4.1658	4.2353	4.2660	4.2126
0.96	3.8869	4.0242	4.0958	4.1274	4.0719
0.95	3.7689	3.9099	3.9834	4.0158	3.9584
0.94	3.6691	3.8133	3.8883	3.9214	3.8624
0.93	3.5819	3.7291	3.8055	3.8392	3.7788
0.92	3.5042	3.6541	3.7318	3.7661	3.7043
0.91	3.4338	3.5862	3.6652	3.6999	3.6369
0.90	3.3693	3.5239	3.6041	3.6393	3.5751
0.88	3.2536	3.4126	3.4948	3.5309	3.4645
0.86	3.1514	3.3143	3.3984	3.4354	3.3670
0.84	3.0592	3.2258	3.3117	3.3494	3.2791
0.82	2.9747	3.1447	3.2323	3.2707	3.1986
0.80	2.8963	3.0695	3.1587	3.1978	3.1240
0.75	2.7196	2.9006	2.9934	3.0341	2.9563
0.70	2.5624	2.7505	2.8468	2.8890	2.8075
0.65	2.4179	2.6129	2.7125	2.7560	2.6709
0.60	2.2818	2.4835	2.5863	2.6312	2.5426
0.55	2.1510	2.3594	2.4654	2.5117	2.4196
0.50	2.0231	2.2384	2.3476	2.3952	2.2995
0.45	1.8961	2.1183	2.2309	2.2799	2.1805
0.40	1.7679	1.9973	2.1133	2.1638	2.0606
0.35	1.6362	1.8733	1.9930	2.0450	1.9377
0.30	1.4984	1.7438	1.8673	1.9210	1.8093
0.25	1.3507	1.6052	1.7331	1.7886	1.6720
0.20	1.1874	1.4524	1.5852	1.6428	1.5206
0.15	0.9985	1.2761	1.4149	1.4749	1.3460
0.10	0.7632	1.0571	1.2035	1.2667	1.1291
0.05	0.4186	0.7376	0.8958	0.9640	0.8128

TABLE I—*Concluded*

Prob. of Correct Ranking	$k = 12$ $t = 4$	$k = 12$ $t = 5$	$k = 13$ $t = 4$	$k = 13$ $t = 5$	$k = 14$ $t = 5$
0.9995	5.8709	5.9002	5.9081	5.9424	5.9793
0.9990	5.6373	5.6678	5.6757	5.7113	5.7494
0.9950	5.0502	5.0841	5.0919	5.1314	5.1728
0.99	4.7725	4.8083	4.8158	4.8576	4.9005
0.98	4.4746	4.5126	4.5197	4.5641	4.6089
0.97	4.2886	4.3281	4.3350	4.3810	4.4271
0.96	4.1502	4.1909	4.1975	4.2449	4.2919
0.95	4.0387	4.0803	4.0867	4.1353	4.1831
0.94	3.9444	3.9870	3.9932	4.0427	4.0911
0.93	3.8623	3.9057	3.9117	3.9621	4.0111
0.92	3.7893	3.8333	3.8391	3.8904	3.9399
0.91	3.7232	3.7678	3.7735	3.8255	3.8756
0.90	3.6626	3.7079	3.7134	3.7661	3.8166
0.88	3.5543	3.6007	3.6059	3.6599	3.7113
0.86	3.4588	3.5063	3.5111	3.5664	3.6185
0.84	3.3729	3.4213	3.4259	3.4822	3.5350
0.82	3.2942	3.3435	3.3478	3.4052	3.4586
0.80	3.2213	3.2715	3.2755	3.3339	3.3879
0.75	3.0577	3.1098	3.1132	3.1739	3.2292
0.70	2.9125	2.9666	2.9693	3.0321	3.0887
0.65	2.7796	2.8354	2.8374	2.9023	2.9600
0.60	2.6547	2.7122	2.7137	2.7805	2.8394
0.55	2.5352	2.5944	2.5952	2.6640	2.7240
0.50	2.4186	2.4796	2.4797	2.5505	2.6116
0.45	2.3032	2.3659	2.3654	2.4382	2.5003
0.40	2.1870	2.2515	2.2503	2.3252	2.3885
0.35	2.0680	2.1345	2.1324	2.2096	2.2741
0.30	1.9439	2.0124	2.0095	2.0890	2.1548
0.25	1.8113	1.8821	1.8782	1.9604	2.0275
0.20	1.6652	1.7387	1.7337	1.8188	1.8875
0.15	1.4970	1.5736	1.5672	1.6560	1.7264
0.10	1.2883	1.3690	1.3608	1.4542	1.5270
0.05	0.9848	1.0716	1.0606	1.1611	1.2374

TABLE II

Table of $\sqrt{N}\lambda$ corresponding to various probabilities, to be used for designing experiments involving 3 normal populations to decide which one has the largest, which the second largest, and which the smallest population mean

Prob. of Correct Ranking	$\sqrt{N}\lambda$	Prob. of Correct Ranking	$\sqrt{N}\lambda$	Prob. of Correct Ranking	$\sqrt{N}\lambda$
0.99	3.6428	0.88	2.1981	0.50	0.9084
0.98	3.2900	0.86	2.0860	0.45	0.7836
0.97	3.0690	0.84	1.9855	0.40	0.6592
0.96	2.9044	0.82	1.8935	0.35	0.5328
0.95	2.7717	0.80	1.8094	0.30	0.4021
0.94	2.6598	0.75	1.6211	0.25	0.2635
0.93	2.5623	0.70	1.4560	0.20	0.1121
0.92	2.4756	0.65	1.3064		
0.91	2.3974	0.60	1.1674	1/6	0.0000
0.90	2.3258	0.55	1.0356		

b) Suppose that it is known that $\sigma_1^2 = 90$, $\sigma_2^2 = 130$, and $\sigma_3^2 = 191$. Following the method of Section 3C(ii) we see that we can let $\sigma^2 = 100$; $a_1 = 0.90$, $a_2 = 1.30$, and $a_3 = 1.91$. Then we have $\lambda = 4/10$. From a), above, we see that $N' = (1.4338/0.4)^2$. Using equation (34) we find that $N_1 = 11.6$, $N_2 = 16.7$, $N_3 = 24.6$; thus we select 12, 17, and 25 observations from populations 1, 2, and 3, respectively.

Example 2. Given a one-way classification of three populations. Suppose that we have selected 15 observations from each of the populations. What is the smallest difference $\mu_{[3]} - \mu_{[2]} = \mu_{[2]} - \mu_{[1]}$ that we can guarantee detecting with probability at least 0.80?

Refer to Table II.

Entering the table we find that the value of $\sqrt{N}\lambda$ associated with a probability of 0.80 is 1.8094. If $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$ is known, say equal to 6 units, we have $\sqrt{15}(\mu_{[i+1]} - \mu_{[i]})/6 = 1.8094$ for $i = 1, 2$; hence $\mu_{[3]} - \mu_{[2]} = \mu_{[2]} = \mu_{[1]} = 2.80$ units; if the variances are completely unknown, no useful statement can be made.

Example 3. Given a two-way classification (2 rows, 3 columns) of six populations, each having the same variance σ^2. Suppose that it is desired to find which *set* of populations has the largest row mean, and which set of populations has the smallest column mean, and to guarantee that the probability of correctly choosing these two sets will be at least 0.60 when $\alpha_{[2]} - \alpha_{[1]} \geqq 0.2\,\sigma$ and $\beta_{[2]} - \beta_{[1]} \geqq 0.4\sigma$. How many observations must be taken from each population?

Refer to Table I, columns headed $k = 2$, $t = 1$ and $k = 3$, $t = 1$.

For a sample of size 9 from each population we have (9)(2) = 18 and (9)(3) =

27 observations contributing to the column and row means, respectively. For $k = 2$, $t = 1$ we have $d_1 = \lambda_1 \sqrt{N_1} = 0.2 \sqrt{27} = 1.0392$, and hence the associated probability lies between 0.75 and 0.80; for $k = 3$, $t = 1$ we have $d_2 = \lambda_2 \sqrt{N_2} = 0.4 \sqrt{18} = 1.6971$, and hence the associated probability lies between 0.80 and 0.82. Interpolation will show that the associated probabilities are equal to 0.7688 and 0.8094, respectively, and hence their product is equal to 0.6223. If a sample of size 8 is taken from each population, the corresponding product is equal to 0.5960. Hence select 9 observations from each population.

Example 4. Given three bivariate normal populations with unknown population variances, covariances, and correlation coefficients. The population correlation coefficient associated with the ith population will be denoted by $\rho_i (i = 1, 2, 3)$; the ranked ρ_i will be denoted by $\rho_{[1]} \leqq \rho_{[2]} \leqq \rho_{[3]}$. It is not known which population is associated with $\rho_{[i]}$. Suppose that it is desired to find which population has the largest correlation coefficient, and to guarantee that the probability of correctly choosing that population will be at least 0.90 with $\rho_{[3]} = 0.7$ and $\rho_{[3]} - \rho_{[2]} \geqq 0.10$. How many observations must be taken from each population?

Refer to Table I, column headed $k = 3$, $t = 1$.

The quantity $d = \sqrt{N}\lambda = \sqrt{N}(\mu_{[3]} - \mu_{[2]})/\sigma$ now is replaced by $\sqrt{N - 3}(\frac{1}{2} \log_e (1 + \rho_{[3]})/(1 - \rho_{[3]}) - \frac{1}{2} \log_e (1 + \rho_{[2]})/(1 - \rho_{[2]}) = \sqrt{N - 3} \log_e 1.1902 = 0.174 \sqrt{N - 3} = 2.2302$ (from the table, for $P = 0.90$). Hence select 168 observations from each population.

9. Directions of future research. The results presented in this paper can be extended and generalized in several directions. The formulation of problems in terms of the *ranking (multiple decision) approach* rather than the *test of homogeneity approach* can be applied equally well to parameters other than population means of normal distributions. As an example of this, the writer has considered the problem of ranking the population variances of normal distributions. The results of this investigation, giving an *exact* rather than a large sample theory, will be presented in a later paper.

Ranking problems can be formulated as several sample, or completely sequential (rather than single sample) multiple decision procedures with resultant savings in the expected number of observations for a given probability of a correct ranking. Some promising results have been obtained thus far [2], [18], but many interesting unsolved problems remain, and additional research in this area should prove very fruitful.

Among the unsolved problems for the single sample procedure, two are of particular interest. It would be very desirable to know whether the multiple decision procedures described in this paper are optimum in any sense. Also, it would be useful to have a simple procedure for determining the most efficient allocation of the sample sizes when the population variances are known and unequal.

10. Acknowledgments. The writer is indebted to Mr. Cuthbert Daniel for having proposed to him a practical problem which led the writer to formulate the ranking approach.

Particular appreciation is recorded here for the efforts of Dr. Milton Sobel who read early versions of this manuscript with painstaking care, and who made many very helpful suggestions.

REFERENCES

[1] ROBERT E. BECHHOFER, CHARLES W. DUNNETT and MILTON SOBEL, "A two-sample multiple decision procedure for ranking means of normal populations with unknown variances (Preliminary Report)," abstract, *Ann. Math. Stat.*, Vol. 24 (1953), p. 136.

[2] ROBERT E. BECHHOFER and MILTON SOBEL, "A sequential multiple decision procedure for ranking means of normal populations with known variances (Preliminary Report)," abstract, *Ann. Math. Stat.*, Vol. 24 (1953), p. 136.

[3] GERTRUDE BLANCH, *NAML Report* 52-53, National Bureau of Standards, Los Angeles, January 5, 1952.

[4] WILLIAM G. COCHRAN and GERTRUDE M. COX, *Experimental Designs*, John Wiley and Sons, Inc., New York, 1950.

[5] DAVID B. DUNCAN, "A significance test for differences between ranked treatments in an analysis of variance," *Virginia J. Sci.*, Vol. 2, N. S. (1951), pp. 171-189.

[6] DAVID B. DUNCAN, "On the properties of the multiple comparisons test," *Virginia J. Sci.*, Vol. 3, N. S. (1952), pp. 49-67.

[7] CHURCHILL EISENHART, "Probability that sample means are in opposite order to population means," *Selected Techniques of Statistical Analysis*, McGraw-Hill Book Co., New York, 1947, pp. 377-382.

[8] R. A. FISHER, *The Design of Experiments*, 4th ed., Oliver and Boyd Ltd., Edinburgh and London, 1947.

[9] R. A. FISHER, *Statistical Methods for Research Workers*, 10th ed., Oliver and Boyd Ltd., Edinburgh and London, 1946.

[10] FREDERICK MOSTELLER, "A k-sample slippage test for an extreme population," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 58-65.

[11] National Bureau of Standards, Personal communication on unpublished tables.

[12] EDWARD PAULSON, "A multiple decision procedure for certain problems in the analysis of variance," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 95-98.

[13] EDWARD PAULSON, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23, (1952), pp. 239-246.

[14] EDWARD PAULSON, "An optimum solution to the k-sample slippage problem for the normal distribution," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 610-616.

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

[16] HENRY SCHEFFÉ, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40 (1953), pp. 87-104.

[17] CHARLES STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 253-258.

[18] CHARLES STEIN, "On sequences of experiments," abstract, *Ann. Math. Stat.*, Vol. 19 (1948), p. 117.

[19] JOHN TUKEY, "Allowances for various types of error rates," unpublished invited address presented before a joint meeting of the Institute of Mathematical Statistics and the Eastern North American Region of the Biometrics Society at Blacksburg, Va., March 19, 1952.