ON A CLASS OF DECISION PROCEDURES FOR RANKING MEANS OF NORMAL POPULATIONS[1]

By K. C. Seal

University of North Carolina

Summary. An infinite class of decision rules having several desirable properties is suggested for choosing a group of populations from a given set of normal populations which should contain the population with the largest mean. The problem of selecting one member from this infinite class of rules has also been studied.

1. Introduction. In recent years it has been recognized ([1], [3], [4], [5], [9], [10], [11]) that the conventional test of homogeneity, such as the F-test in the analysis of variance for testing the equality of several population means, does not supply all the information that the experimenter seeks. In many practical situations it is unrealistic to assume that the population means of several essentially different populations will be equal. A sufficiently large sample will thus enable the experimenter to detect this difference at any preassigned level. In most cases what the experimenter actually wants is a decision procedure which would tell him which population or populations possess a desired characteristic. For example, the experimenter may be interested in determining the population with the largest mean, from a set of normal populations. Alternatively he may desire to select from a given number of populations a group containing the population having the largest mean.

Suppose there are $n + 1$ normal populations $N(\mu_i, \sigma_1^2)$, $i = 0, 1, 2, \cdots, n$, with unknown means and a common but unknown variance and that k random observations $x_{i\alpha}$ ($i = 0, 1, \cdots, n$; $\alpha = 1, 2, \cdots, k$) from each of the $n + 1$ normal populations are given, where $x_{i\alpha}$ is one of the k observations from the ith population. Under our assumptions the $n + 1$ sample means

$$x_i = \sum_{\alpha=1}^{k} x_{i\alpha} / k$$

will obey $N(\mu_i, \sigma_1^2 / k)$, $i = 0, 1, \cdots, n$, and an estimate

$$s_1^2 = \sum_{i=0}^{n} \sum_{\alpha=1}^{k} (x_{i\alpha} - x_i)^2 / (k - 1)(n + 1)$$

of σ_1^2 can be obtained which is independent of the sample means x_i, $i = 0, 1, \cdots, n$. We may, therefore, assume for mathematical convenience that just one random observation x_i from each of $n + 1$ normal populations $N(\mu_i, \sigma^2)$, $i =$

Received July 6, 1954.

[1] This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

0, 1, $\cdots$, n, is given, where $\sigma^2 = \sigma_1^2 / k$ is estimated by

$$s^2 = s_1^2 / k = \sum_{i=0}^{n} \sum_{\alpha=1}^{k} (x_{i\alpha} - x_i)^2 / [k(k-1)(n+1)]$$

and this is assumed to be known. Clearly this estimate s^2 of σ^2 is stochastically independent of the given observations x_i, $i = 0, 1, \cdots, n$. It is desired to choose a group of populations from the above $n + 1$ populations, with the help of some decision rule which ensures that the least upper bound of the probability of not including in the group the population with the largest mean is α $(0 < \alpha < 1)$, whatever may be the unknown μ_i's. Subject to this fundamental requirement we would like the rule to possess other desirable properties such as:

(a) The property of unbiasedness, i.e., the probability of rejecting any population not having the largest mean is not less than the probability of rejecting the population having the largest mean. (Analogy of this property to the property of unbiasedness in the theory of testing of hypothesis should be noticed.)

(b) The property of gradation, i.e., corresponding to any α $(0 < \alpha < 1)$, there exists a constant $\mu_{0\alpha}$ such that the chance of retaining the population with mean μ_0 in the group is greater or less than α, according as μ_0 is greater or less than $\mu_{0\alpha}$. The constant $\mu_{0\alpha}$ will in general depend on the decision rule as well as the unknown means of the remaining n populations, and the common variance σ^2.

An infinite class $\mathcal{C}$ of decision rules satisfying the fundamental requirement, together with the properties (a) and (b) is given in Section 2.1. Certain interesting properties of this class are studied in Section 3. The question of choosing one member from this infinite class having further desirable properties has been studied in Section 4.

2. Class $\mathcal{C}$ of decision procedures.

2.1. Let y_i, $i = 0, 1, \cdots, n$, be $n + 1$ random observations from $N(0, \sigma^2)$ and let $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$ be n ranked observations among $y_1, \cdots, y_n$. The $y_{(i)}$'s will then define another set of random variables $Y_{(i)}$, $i = 1, \cdots, n$. It is assumed $y_i \neq y_j$, $i \neq j$, since the set of points $(y_0, y_1, \cdots, y_n)$ in $(n + 1)$-dimensional Euclidean space where $y_i \neq y_j$, $i \neq j$ will be obtained with probability 1. Let $t_\alpha(c_1, \cdots, c_n)$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ denote the *upper* 100 α % point in the probability density function (pdf) of

$$t(c_1, \cdots, c_n) = \frac{\sum_1^n c_i Y_{(i)} - Y_0}{s} \tag{2.1.1}$$

The class $\mathcal{C}$ of decision rules $D(c_1, \cdots, c_n)$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ is defined as follows:

"Reject any observation x_0 from the given observations x_i, $i = 0, 1, \cdots, n$, if

$$\sum_{i=1}^{n} c_i x_{(i)} - x_0 > s t_\alpha(c_1, \cdots, c_n) \tag{2.1.2}$$

and accept otherwise, where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are n ordered observations among $(x_1, x_2, \cdots, x_n)$. (The $n + 1$ observations x_i, $i = 0, 1, \cdots, n$, taken from normal populations are again assumed to be distinct.) Proceed as above for each of $n + 1$ observations separately, so that each of $n + 1$ observations in due turn takes the place of x_0 and the remaining ordered observations play the part of $x_{(1)}, \cdots, x_{(n)}$." Thus in the above procedure we may start with the largest observation among x_i, $i = 0, 1, \cdots, n$, as x_0 and work downwards. If any particular observation is rejected, all other observations smaller than this observation are automatically rejected.

For the sake of convenience, we shall denote the decision rule $D(c_1, \cdots, c_n)$ when (i) $c_i = 1/n$, $i = 1, \cdots, n$, by $\bar{D}$ and when (ii) $c_r = 1$, and $c_j = 0$, $j \neq r$ by $D(r)$, $1 \leqq r \leqq n$. The corresponding auxiliary statistics $t(c_1, \cdots, c_n)$ will be denoted by $\bar{t}$ and $t(r)$.

It may also be noted here that $\sum_1^n c_i x_{(i)} - x_0$ $(c_i \geqq 0, i = 1, \cdots, n, \sum_1^n c_i = 1)$ can be written in the alternative form $\sum_1^n c_i(x_{(i)} - x_0)$.

3. Some properties of class $\mathcal{C}$.

3.1. *An inequality related to location parameters.*

THEOREM 3.1.1. *Suppose that* $F((x_1 - \mu_1) / \sigma_1, \cdots, (x_n - \mu_n) / \sigma_n)$ *is the cumulative distribution function* (cdf) *of n random variables X_i, $i = 1, \cdots, n$, and $T(u_1, \cdots, u_n)$ is a real-valued function of u_i, $i = 1, \cdots, n$, such that*

$$T(u_1 + \alpha_1, \cdots, u_n + \alpha_n) \geqq T(u_1, \cdots, u_n), \tag{3.1.1}$$

where $(\alpha_1, \cdots, \alpha_n)$ is a set of real numbers and $-\infty < u_i < \infty$, $i = 1, \cdots, n$ If for an arbitrary constant k,

$$P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix}\mu_1, \cdots, \mu_n\\ \sigma_1, \cdots, \sigma_n\end{matrix}\right]$$

denotes the probability of $T(X_1, \cdots, X_n) > k$ when $X_1, \cdots, X_n$ have the cdf *$F((x_1 - \mu_1) / \sigma_1, \cdots, (x_n - \mu_n) / \sigma_n)$, then*

$$P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix}\mu_1 + \alpha_1, \cdots, \mu_n + \alpha_n\\ \sigma_1, \cdots, \sigma_n\end{matrix}\right] \geqq P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix}\mu_1, \cdots, \mu_n\\ \sigma_1, \cdots, \sigma_n\end{matrix}\right].$$

PROOF.

$$\begin{aligned} P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix}\mu_1 + \alpha_1, \cdots, \mu_n + \alpha_n\\ \sigma_1 \quad, \cdots, \quad \sigma_n\end{matrix}\right] &= P\left[T(X_1 + \alpha_1, \cdots, X_n + \alpha_n) > k \,\middle|\, \begin{matrix}\mu_1, \cdots, \mu_n\\ \sigma_1, \cdots, \sigma_n\end{matrix}\right] \\ &\geqq P\left[T(X_1, \cdots, X_n) > k \,\middle|\, \begin{matrix}\mu_1, \cdots, \mu_n\\ \sigma_1, \cdots, \sigma_n\end{matrix}\right], \end{aligned}$$

since $T(X_1 + \alpha_1, \cdots, X_n + \alpha_n) \geqq T(X_1, \cdots, X_n)$ by hypothesis. Q.E.D.

From this theorem the following corollaries readily follow:

COROLLARY 3.1.1. *If* (3.1.1) *is satisfied for all*

$$\alpha_i \geqq 0, i = 1, \cdots, n, \quad \textit{then} \quad P[T(X_1, \cdots, X_n) > k]$$

is a nondecreasing function of each μ_i, $i = 1, \cdots, n$.

COROLLARY 3.1.2. *If*

$$T(u_1 + \alpha_1, \cdots, u_n + \alpha_n) > T(u_1, \cdots, u_n),$$

when $\alpha_i \geqq 0$ *and* $\alpha_i > 0$ *for at least one* i, $1 \leqq i \leqq n$, *and if the* cdf *of*

$$T(X_1, \cdots, X_n)$$

assigns a positive measure to every nondegenerate interval, then

$$P[T(X_1, \cdots, X_n) > k],$$

where k *is an arbitrary constant, is an increasing function of each* μ_i, $i = 1, \cdots, n$.

COROLLARY 3.1.3. *Any strictly monotonic functional of the* cdf *of*

$$T(X_1, \cdots, X_n),$$

which satisfies the conditions of Corollary 3.1.2, *is an increasing function of* μ_i, $i = 1, \cdots, n$.

EXAMPLE 1. Consider the pdf

$$f\left(\frac{x_1 - \mu_1}{\sigma_1}, \cdots, \frac{x_n - \mu_n}{\sigma_n}\right) = \prod_{i=1}^{n} (\sigma_i \sqrt{2\pi})^{-1} e^{-(x_i - \mu_i)^2/2\sigma_i}$$

and $T(X_1, \cdots, X_n) = (\sum_1^n c_i X_{(i)})^{2r+1}$, where $r = 0, 1, 2, \cdots$ and $c_i \geqq 0$, $i = 1, \cdots, n$, and $c_i > 0$ for at least one i.

Here the conditions of Corollary 3.1.2 are easily verified and it follows that

$$P\left[\left(\sum_{i=1}^{n} c_i X_{(i)}\right)^{2r+1} > k\right], \qquad r = 0, 1, 2, \cdots,$$

is an increasing function of each μ_i, $i = 1, \cdots, n$. This result for the particular case $r = 0$ will be used in Section 3 in proving the properties of unbiasedness and gradation for the class $\mathcal{C}$ of decision rules as defined in Section 2.1.

It is well known that if $F(x)$ is the cdf of a random variable X, then expectation $E(X)$, if it exists, is a strictly monotonic functional of F (cf. [6], p. 152–153; [12], p. 189). Hence we get from Corollary 3.1.3 that

$$E\left(\sum_{i=1}^{n} c_i X_{(i)}\right)^{2r+1} \qquad r = 0, 1, 2, \cdots,$$

is an increasing function of each of μ_i, $i = 1, \cdots, n$. Much more complicated functions can be constructed (cf. [13], pp. 25–26) having a similar property.

3.2. *Property of unbiasedness.* Let $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ denote the set of normal populations $N(\mu_i, \sigma^2)$, $i = 1, 2, \cdots, n$. Suppose that $X_{(1)} < \cdots < X_{(n)}$ are n order statistics from $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ when one random observation from each of these n normal populations with *common* variance equal to σ^2 is taken. Let X_0 be another independent variate obeying $N(\mu_0, \sigma^2)$. According to our decision rule $D(c_1, \cdots, c_n)$ as defined in Section 2.1, the probability of rejecting x_0 will then be given by

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha (c_1, \cdots, c_n)\right] \tag{3.2.1}$$
$$= \int_0^\infty ds \int_{-\infty}^{\infty} dx_0 \int \underset{A}{\cdots} \int p(s) e^{-(x_0-\mu_0)^2/2\sigma^2} \cdot g(x_{(1)}, \cdots, x_{(n)} \mid \mu_1, \cdots, \mu_n)\, dx_{(1)} \cdots dx_{(n)},$$

where

$$A = \begin{pmatrix} -\infty < x_{(1)} < \cdots < x_{(n)} < \infty \\ \sum_{i=1}^{n} c_i x_{(i)} > x_0 + st_\alpha(c_1, \cdots, c_n) \end{pmatrix}, \tag{3.2.2}$$

$g(x_{(1)}, \cdots, x_{(n)} \mid \mu_1, \cdots, \mu_n)$ represents the pdf of $X_{(1)}, \cdots, X_{(n)}$ from $\Omega(\mu_1, \cdots, \mu_n; \sigma)$ and

$$p(s) = \frac{\nu^{\nu/2}}{2^{(\nu-2)/2}\,\Gamma(\nu/2)\sigma^\nu} e^{-\nu s^2/2\sigma^2} s^{\nu-1}, \tag{3.2.3}$$

i.e., the pdf of sample standard deviation s based on $\nu = (k-1)(n+1)$ (cf. Section 1) degrees of freedom.

THEOREM 3.2.1. *$P[\sum_1^n c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)]$ is an increasing function of each $\mu_i' = \mu_i - \mu_0$, $i = 1, \cdots, n$.*

PROOF. Let $X_i' = X_i - \mu_0$, $i = 0, 1, 2, \cdots, n$. Since $\sum_1^n c_i = 1$,

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] = P\left[\sum_{i=1}^{n} c_i X'_{(i)} > X_0' + st_\alpha(c_1, \cdots, c_n)\right]. \tag{3.2.4}$$

For fixed values of X_0' and s, the conditional value of

$$P\left[\sum_{i=1}^{n} c_i X'_{(i)} > X_0' + st_\alpha(c_1, \cdots, c_n)\right]$$

is an increasing function of $\mu_i' = \mu_i - \mu_0$, by Corollary 3.1.2 and Example 1 of Section 3.1. Since the distribution of X_0' and s does not involve the μ_i', it is now obvious that the (unconditional) value of

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right]$$

is an increasing function of each μ_i'. Q.E.D.

From this theorem an interesting property of $D(c_1, \cdots, c_n)$ follows.

COROLLARY 3.2.1. *The probability of rejecting any undesirable population (i.e., any population which has not the largest mean) is never less than the probability of rejecting the desirable population (i.e., that population having the largest mean).*

PROOF. For the present proof let $\mu_{(0)} \geqq \cdots \geqq \mu_{(n)}$ denote the mean of the given $n + 1$ normal populations with common variance σ^2. By our decision rule $D(c_1, \cdots, c_n)$ the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$ will depend on

$$P_{c_1,\cdots,c_n}(\mu_{(1)} - \mu_{(0)}, \cdots, \mu_{(n)} - \mu_{(0)}), \tag{3.2.5}$$

which is defined as the conditional probability of

$$\sum_{i=1}^{n} c_i Y_{(i)} > y_0 + st_\alpha(c_1, \cdots, c_n),$$

when y_0 and s are assumed to be held constant. Here $Y_{(i)}$, $i = 1, \cdots, n$, and Y_0 are defined as in Section 2.1. The probability of rejecting any undesirable population $N(\mu_{(i)}, \sigma^2)$, $i = 1, 2, \cdots, n$, will, on the other hand, involve

$$P_{c_1,\cdots,c_n}(\mu_{(0)} - \mu_{(i)}, \mu_{(1)} - \mu_{(i)}, \cdots, \mu_{(i-1)} - \mu_{(i)}, \mu_{(i+1)} - \mu_{(i)}, \cdots, \mu_{(n)} - \mu_{(i)}). \tag{3.2.6}$$

Comparing the arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) and (3.2.6) we notice that

$$\mu_{(j)} - \mu_{(i)} \geqq \mu_{(j)} - \mu_{(0)}, \qquad \mu_{(0)} - \mu_{(i)} \geqq \mu_{(i)} - \mu_{(0)},$$

where $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, n$; $j \neq i$. Thus we can make a one to one correspondence between the n arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) and (3.2.6) in such a way that no argument of $P_{c_1,\cdots,c_n}$ in (3.2.6) is less than the corresponding argument of $P_{c_1,\cdots,c_n}$ in (3.2.5). Hence from the monotonic behavior of $P_{c_1,\cdots,c_n}(\delta_1, \cdots, \delta_n)$ with regard to δ_i, $i = 1, \cdots, n$, it follows that the probability of rejecting any undesirable population $N(\mu_{(i)}, \sigma^2)$, $i = 1, \cdots, n$, is never less than the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$. This property may be denoted by the *property of unbiasedness* which is therefore possessed by our decision rules $D(c_1, \cdots, c_n)$ $(c_i \geqq 0, \sum_1^n c_i = 1)$. It may also be noted that all the arguments of $P_{c_1,\cdots,c_n}$ in (3.2.5) are nonpositive and so (3.2.5) will not exceed $P_{c_1,\cdots,c_n}(0, \cdots, 0)$. This implies that the probability of rejecting the desirable population $N(\mu_{(0)}, \sigma^2)$ will not exceed the desired significance level α $(0 < \alpha < 1)$. Hence α will be the least upper bound of the probability of incorrect choice (i.e., not including the population with the largest mean in the selected group), whatever may be the population means. Thus any rule $D(c_1, \cdots, c_n)$ satisfies the fundamental requirement as stated in Section 1.

3.3. *Property of gradation.* From Theorem 3.2.1 it follows that

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] \tag{3.3.1}$$

is a decreasing function of μ_0; when $\mu_0 \to -\infty$, the value of (3.3.1) is equal to 1 and when $\mu_0 \to +\infty$, the same value is equal to 0. It is easily seen from (3.2.1) that (3.3.1) is a continuous function of μ_0. Hence corresponding to any assigned value γ $(0 < \gamma < 1)$ of (3.3.1) there exists a particular value $\mu_{0\gamma}$ of μ_0 for which (3.3.1) is exactly equal to γ. The value $\mu_{0\gamma}$ will clearly in general depend upon $\mu_1, \cdots, \mu_n, \sigma$ and $c_1, \cdots, c_n$ besides the assigned value γ, and if

$$\mu_1, \mu_2, \cdots, \mu_n$$

increase by a given constant Δ, then $\mu_{0\gamma}$ will also be increased by the same constant. In this situation we shall, therefore, find that

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right] \gtreqless \gamma, \tag{3.3.2}$$

according as $\mu_0 \lesseqgtr \mu_{0\gamma}$. This property will be designated as the *property of gradation*. We shall now study the nature of the unknown constant $\mu_{0\alpha}$ when γ is taken to be equal to α $(0 < \alpha < 1)$. It will be shown that $\mu_{0\alpha}$ for the decision rule $\bar{D}$ is very simple in form, but for other decision rules of class $\mathcal{C}$ no such simple explicit expression for $\mu_{0\alpha}$ can be given. Let $\bar{X}_n = \sum_1^n X_{(i)} / n$. Then $\bar{X}_n$ will obey $N(\sum_1^n \mu_i / n, \sigma^2 / n)$. Let $Y_{(1)} < \cdots < Y_{(n)}$ be n order statistics derived from a random sample of size n from $N(0, \sigma^2)$. Also let $Y_0 = X_0 - \mu_0$, so that Y_0 obeys $N(0, \sigma^2)$. Clearly the distribution of $\bar{X}_n - \sum_1^n \mu_i / n$ is identical with that of $\bar{Y}_n = \sum_1^n Y_{(i)} / n$.

Under our decision rule $\bar{D}$ the probability of rejecting x_0 in a single rejection is equal to

$$P[\bar{X}_n - X_0 > s\bar{t}_\alpha] = P\left[(\bar{Y}_n - Y_0) + \left(\sum_{i=1}^{n} \mu_i / n - \mu_0\right) > s\bar{t}_\alpha\right]$$

$$\gtreqless P[\bar{Y}_n - Y_0 > s\bar{t}_\alpha] = \alpha,$$

according as $\mu_0 \lesseqgtr \sum_1^n \mu_i / n$. Thus for $\bar{D}$ we have the special property that the probability of rejecting any population whose mean is greater than the average of the remaining n population means is less than α and the probability of rejecting any population whose mean is not greater than the average of the remaining n population means is at least equal to α.

It is now shown that $\mu_{0\alpha}$ for the general decision rule $D(c_1, \cdots, c_n)$ is not in general equal to $E(\sum_1^n c_i X_{(i)})$, although we have just shown that this is true for $\bar{D}$.

The existence of $\mu_{0\alpha}$ (which is a function of $\mu_1, \cdots, \mu_n, \sigma; c_1, \cdots, c_n$ besides α) for which the property of gradation holds for the general decision rule has already been shown. This implies that

$$\alpha = P\left[\sum_{i=1}^{n} c_i X_{(i)} - X_0 > st_\alpha(c_1, \cdots, c_n)\right], \tag{3.3.3}$$

when $E(X_0) = \mu_{0\alpha}(\mu_1, \cdots, \mu_n, \sigma; c_1, \cdots, c_n, \alpha)$. It is shown that the assumption $\mu_{0\alpha} = E(\sum_1^n c_i X_{(i)})$ (which implies that $\mu_{0\alpha}$ is independent of α)

leads to a contradiction for the general case. The right-hand side of (3.3.3) can be written as

$$P\left[\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha} - (X_0 - \mu_{0\alpha}) > st_\alpha(c_1, \cdots, c_n)\right] \tag{3.3.4}$$

But $Y_0 = X_0 - \mu_{0\alpha}$ obeys $N(0, \sigma^2)$; hence by (3.3.3), (3.3.4), and the definition of $t_\alpha(c_1, \cdots, c_n)$ we get

$$\begin{aligned} P\left[\sum_{i=1}^{n} c_i Y_{(i)} - Y_0 > st_\alpha(c_1, \cdots, c_n)\right] &= \alpha \\ &= P\left[\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha} - Y_0 > st_\alpha(c_1, \cdots, c_n)\right], \end{aligned} \tag{3.3.5}$$

where $Y_{(1)} < \cdots < Y_{(n)}$ are n order statistics obtained from a sample of size n from $N(0, \sigma^2)$. Hence it follows that, when $\mu_{0\alpha}$ is assumed to be independent of α, the distribution of $\sum_1^n c_i X_{(i)} - \mu_{0\alpha}$ and $\sum_1^n c_i Y_{(i)}$ must be identical. As a necessary condition for this we then have

$$\begin{aligned} E\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) &= E\left(\sum_{i=1}^{n} c_i X_{(i)} - \mu_{0\alpha}\right) \\ &= E\left(\sum_{i=1}^{n} c_i X_{(i)}\right) - \mu_{0\alpha}. \end{aligned} \tag{3.3.6}$$

Hence if we assume that the unknown constant $\mu_{0\alpha}$ is $E(\sum_1^n c_i X_{(i)})$, then it will follow that

$$E\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) = 0, \tag{3.3.7}$$

for an arbitrary set of c_i's such that $c_i \geqq 0$ and $\sum_1^n c_i = 1$. The equation (3.3.7) does not, however, hold in general and hence we arrive at the conclusion that $\mu_{0\alpha}$ is not in general equal to $E(\sum_1^n c_i X_{(i)})$. We can, however, easily derive the value of $\mu_{0\alpha}$ for $D(c_1, \cdots, c_n)$ when $\mu_1 = \mu_2 = \cdots = \mu_n$. It can be easily shown (cf. [13], p. 70) that in such a situation $\mu_{0\alpha}$ must also be equal to μ_1.

4. Selection of an optimum rule.

4.1. In this section we shall assume that the number of degrees of freedom $(k - 1)(n + 1)$ of s (cf. Section 1) is so large that σ may be considered to be known. Under this restriction the rule $D(c_1, \cdots, c_n)$ as described in Section 2.1 requires the obvious modification that s should be replaced throughout by the population standard deviation σ.

It has been shown that the class $\mathcal{C}$ of decision rules satisfies the fundamental requirement, i.e., the least upper bound of the probability of rejecting the population having the largest mean from the selected group is α $(0 < \alpha < 1)$, whatever may be the means of $n + 1$ given normal populations. If among the $n + 1$ population means all means except one are equal, then obviously it would be desirable to select that rule from the class $\mathcal{C}$ which

(i) maximizes the probability of retaining in the selected group the population with the unequal mean if this is larger than the common mean of the other n populations; and

(ii) maximizes the probability of not retaining the population with the unequal mean if this is smaller than the mean of the other n populations.

In case (i) the population with the largest mean will be designated as the "best" population, and in case (ii) the population with the smallest mean will be called the "worst" population. Thus if $X_{(1)} < \cdots < X_{(n)}$ are assumed to have come from $N(0, \sigma^2)$ and X_0 from $N(\delta, \sigma^2)$, then our desirable rule should ensure largest probability (i) for retaining x_0 in this selected group if $0 < \delta < \infty$, or, (ii) for rejecting x_0 from the group if $-\infty < \delta < 0$. From what we have observed in Section 3.2 it is clear that the above rule will be optimum when $X_{(1)} < \cdots < X_{(n)}$ are assumed to arise from $N(\mu, \sigma^2)$ and X_0 from $N(\mu + \delta, \sigma^2)$, $-\infty < \mu < \infty$.

We shall now show that among the class $\mathcal{C}$ of decision rules the rule $\bar{D}$ maximizes (approximately) the probability of retaining the "best" population in the selected group. In an exactly analogous way it can be shown that $\bar{D}$ maximizes also the probability of rejecting the "worst" population from the group. To derive this result we shall first prove the following:

LEMMA 4.1.1. *Let $Y_{(1)} < \cdots < Y_{(n)}$ be n order statistics from $N(0, 1)$. Then $\sum_1^n Y_{(i)}/n = \sum_1^n Y_i/n$ has minimum variance among all $\sum_1^n c_i Y_{(i)}$ such that $\sum_1^n c_i = 1$.*

PROOF. We have

$$\mathrm{Var}\left(\sum_{i=1}^{n} c_i Y_{(i)}\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j v_{ij}, \tag{4.1.1}$$

where v_{ij} denotes the covariance between $Y_{(i)}$ and $Y_{(j)}$. Let the variance-covariance matrix of $Y_{(i)}$ and $Y_{(j)}$ $(i = 1, \cdots, n; j = 1, \cdots, n)$ be denoted by $\Sigma(n \times n)$

To minimize (4.1.1) subject to the condition

$$\sum_{i=1}^{n} c_i = 1, \tag{4.1.2}$$

we get the following n equations

$$\sum_{j=1}^{n} c_j v_{ij} = \lambda, \qquad i = 1, \cdots, n, \tag{4.1.3}$$

where 2λ is used as Lagrangian multiplier. In matrix notation equations (4.1.3) can be written as

$$\Sigma \mathbf{c} = \lambda \mathbf{1}, \tag{4.1.3a}$$

where $\mathbf{c}'(1 \times n)$ and $\mathbf{1}'(1 \times n)$ denote the row vectors $(c_1, \cdots, c_n)$ and

$$(1, 1, \cdots, 1)$$

respectively. Since Σ is nonsingular, we get from (4.1.3a)

$$\mathbf{c} = \lambda\Sigma^{-1}\mathbf{1}. \tag{4.1.4}$$

But it is known ([7], [8]) that

$$\sum_{j=1}^{n} v_{ij} = 1. \tag{4.1.5}$$

Hence $\Sigma\mathbf{1} = \mathbf{1}$; this implies

$$\Sigma^{-1}\mathbf{1} = \mathbf{1}. \tag{4.1.6}$$

By (4.1.2), (4.1.4), and (4.1.6) it follows that $c_i = 1/n$, $i = 1, \cdots, n$.

This completes the proof of the lemma.

The probability of retaining x_0 arising from the "best" population when $D(c_1, \cdots, c_n)$ is followed will clearly be given by

$$\frac{n!}{(2\pi)^{(n+1)/2}\sigma^{n+1}} \int \cdots \int_B \exp\left[-(x_0 - \delta)^2/2\sigma^2 - \sum_{i=1}^{n} x_{(i)}^2/2\sigma^2\right] \prod_{i=0}^{n} dx_{(i)}, \tag{4.1.7}$$

where

$$B = \begin{bmatrix} -\infty < x_{(1)} < \cdots < x_{(n)} < \infty \\ \sum_{i=1}^{n} c_i x_{(i)} - x_0 < \sigma t_\alpha(c_1, \cdots, c_n) \\ -\infty < x_0 < \infty \end{bmatrix}$$

Our object is to show that the expression (4.1.7) is (approximately) maximum for $\bar{D}$. The arguments given in [13], pp. 71–84 and [14] suggest that

$$u(c_1, \cdots, c_n) = \sum_{i=1}^{n} c_i Y_{(i)} - Y_0,$$

where the c_i's, $Y_{(i)}$'s and Y_0 have the same meaning as in Section 1, may be assumed to be normally distributed for all practical purposes whatever may be the value of n. Let the (approximate) normal distribution of $u(c_1, \cdots, c_n)$ be denoted by $N(\xi_c, \sigma_c^2)$. Henceforth we shall consider this distribution to be exactly normal and hence the result derived below is correct only approximately. For the special case when all c_i's are equal, i.e., $c_i = 1/n$, $i = 1, \cdots, n$, we shall write $\bar{u}$ for $u(1/n, \cdots, 1/n)$ and $N(\bar{\xi}, \bar{\sigma}^2)$ for the (exact) distribution of $\bar{u}$. By Lemma 4.1.1 we know that $\bar{\sigma}$ is the minimum among all σ_c, where $\sum_1^n c_i = 1$.

In the given situation $\sum_1^n c_i X_{(i)} - X_0 + \delta$ will have the (approximate) normal distribution $N(\xi_c, \sigma_c^2)$, where mean ξ_c and variance σ_c^2 are independent of δ. Hence

$$v(c_1, \cdots, c_n) = \frac{\sum_1^n c_i X_{(i)} - X_0 + \delta - \xi_c}{\sigma_c} \tag{4.1.8}$$

will have standard normal distribution $N(0, 1)$.

Hence the expression in (4.1.7) can be written as

$$(2\pi)^{-1/2} \int_{-\infty}^{[\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c + \delta]/\sigma_c} e^{-v^2/2}\, dv. \tag{4.1.9}$$

Also from the definition of $t_\alpha(c_1, \cdots, c_n)$ (cf. Section 1) it is now evident that

$$(2\pi)^{-1/2} \int_{[\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c]/\sigma_c}^{\infty} e^{-v^2/2}\, dv = \alpha. \tag{4.1.10}$$

From (4.1.10) it follows that

$$\frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c}{\sigma_c} = \frac{\sigma \bar{t}_\alpha - \bar{\xi}}{\bar{\sigma}}. \tag{4.1.11}$$

From (4.1.9) it is easily seen that the probability of retaining x_0 in the selected group under the present situation is an increasing function of δ—a result which is a particular case of Corollary 3.1.2. Now for any arbitrary $\delta > 0$ the term in (4.1.9) will be maximum (when the c_i's are varied subject to the conditions $c_i \geqq 0$, $\sum_1^n c_i = 1$) when

$$\frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c + \delta}{\sigma_c} = \frac{\sigma t_\alpha(c_1, \cdots, c_n) - \xi_c}{\sigma_c} + \frac{\delta}{\sigma_c} \tag{4.1.12}$$

is maximum. But $\sigma_c \geqq \bar{\sigma}$ (for all c_i's subject to the above restrictions) implies that $\delta/\bar{\sigma} \geqq \delta/\sigma_c$ for any $\delta > 0$. Hence by (4.1.11) and (4.1.12) it follows that (4.1.12) is maximum for $\bar{D}$. Thus the rule $\bar{D}$ may be taken as the optimum rule.

It is interesting to note the close similarity of this optimum rule $\bar{D}$ to the usual (Student's) t-statistic for which a desirable property has been recently derived by Bahadur [2] while studying two normal populations with a common variance.

5. Acknowledgements. The author wishes to express his deep gratitude to Professor R. C. Bose for his inspiring guidance in the course of this investigation and to Professors S. N. Roy and Wassily Hoeffding for their helpful suggestions and criticism.

REFERENCES

[1] Raghu Raj Bahadur, "On a problem in the theory of k populations", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 362–375.

[2] R. R. Bahadur, "A property of the t-statistic", *Sankhyā*, Vol. 12 (1952), pp. 79–88.

[3] Robert E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 16–39.

[4] 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.

[5] 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), pp. 136–137.

[6] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge University Press, Cambridge, (1934).

[7] HOWARD L. JONES, "Exact lower moments of order statistics in small samples from a normal distribution", *Ann. Math. Stat.*, Vol. 19 (1948), pp. 270–273.

[8] E. H. LLOYD, "Least squares estimation of location and scale parameters using order statistics", *Biometrika*, Vol. 39 (1952), pp. 88–95.

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

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

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

[12] H. SCHEFFÉ AND J. W. TUKEY, "Non-parametric estimation. 1. Validation of order statistics", *Ann. Math. Stat.*, Vol. 16 (1945), pp. 187–192.

[13] K. C. SEAL, "On a class of decision procedures for ranking means", Unpublished Thesis (1954), University of North Carolina, Chapel Hill.

[14] K. C. SEAL, "Approximate distribution of certain linear function of order statistics", *Sankhyā*, submitted for publication.

[15] L. H. C. TIPPETT, "The extreme individuals and the range of samples taken from a normal population", *Biometrika*, Vol. 17 (1925), pp. 151–164.