

## OPTIMUM SIGNIFICANCE LEVELS FOR MULTISTAGE COMPARISON PROCEDURES<sup>1</sup>

BY E. L. LEHMANN AND JULIET POPPER SHAFFER

*University of California, Berkeley*

The framework for multistage comparison procedures in the present paper is roughly that introduced by Duncan and treated more fully by Tukey. In the present paper we consider the problem of finding the optimum allocation of nominal significance levels for successive stages. The optimum procedure we obtain when the number  $s$  of treatments is odd, and the compromise procedure we propose for even  $s$ , essentially agree with a procedure suggested by Tukey. The agreement is exact when  $s$  is even and close when  $s$  is odd. The results of the present paper apply among others to the problem of distinguishing normal distributions with known variances, multinomial distributions, Poisson distributions, and distributions in certain nonparametric settings. However, they do not apply exactly to the comparison of normal distributions with a common unknown variance. When the variances are completely unknown, the method applies in principle but faces the difficulty that no exact test is then available for testing the homogeneity of a set of means.

**1. Introduction.** The framework for multistage comparison procedures in the present paper is roughly that introduced by Duncan [1, 2] and treated more fully by Tukey [9, Part F]. It is also used in the recent paper by Einot and Gabriel [3] which is concerned with comparing the power of several such procedures corresponding to different allocation of nominal levels. In the present paper we shall consider the problem of finding an optimum such allocation. Our formulation differs slightly from that of Einot and Gabriel in the standardization used to insure comparability of the different procedures. For this purpose, Einot and Gabriel fix the probability of at least one false significance statement when all the treatments are equal; instead, we prefer to use the maximum probability of at least one false significance statement. These coincide in many but not in all cases (for example, not in the Newman-Keuls allotment). Einot and Gabriel also compare the power of Studentized range and  $F$ -statistics. Such a comparison is outside the scope of the present paper, which is concerned only with the allocation of nominal significance levels among the different stages.

The optimum procedure we obtain when the number  $s$  of treatments is odd, and the compromise procedure we propose for even  $s$ , essentially agree with a procedure suggested by Tukey. The agreement is exact when  $s$  is even and close when  $s$  is odd. Tukey's purpose, to obtain "constant standard significance level," although

---

Received January 1977; revised August 1977.

<sup>1</sup>This paper was prepared with support of U.S. Office of Naval Research contract number N00014-75-C-0444/NR 042-036. Also this research was prepared with the support of National Science Foundation grant number MCS76-10238.

AMS 1970 subject classifications. 62F99 62G99 62J15.

*Key words and phrases.* Multistage comparison procedures, optimal significance levels, multiple comparisons, maximum error probability, power.

having a different motivation, thus turns out to be close to that of maximizing power.

The results of the present paper apply among others to the problem of distinguishing normal distributions with known variances, multinomial distributions, Poisson distributions, and distributions in certain nonparametric settings. However, they do not apply exactly to the comparison of normal distributions with a common unknown variance (or to other situations in which estimation of a common nuisance parameter destroys the independence of test statistics from nonoverlapping sets of samples). When the variances are completely unknown, the method applies in principle but faces the difficulty that no exact test is then available for testing the homogeneity of a set of means. These complications are discussed in part (i) of Section 7.

The general formulation of multistage comparison procedures adopted by Duncan [2], Tukey [9], Einot and Gabriel [3], and in the present paper runs into a difficulty when the test statistics used are the (Studentized) sample ranges. The resulting procedure may then not agree with the multiple range procedures as commonly defined and used, unless some additional conditions are satisfied. This difficulty is considered in part (iv) of Section 7.

**2. Multistage procedures.** Consider  $s$  independent random quantities  $X_i$ ,  $i = 1, \dots, s$  (they may, for example, be vector-valued) with distributions  $P_i$  ranging over sets  $\mathcal{P}_i$ . We shall be concerned with the problem of distinguishing among functions  $g_i$  defined over  $\mathcal{P}_i$  by means of a multistage procedure based on the  $X_i$ .

A multistage procedure is defined through a series of rejection regions. The first stage consists of a test of the hypothesis

$$(2.1) \quad H_s : g_1(P_1) = \dots = g_s(P_s)$$

by means of a rejection region  $R_s$  in the sample space of  $(X_1, \dots, X_s)$ . If  $(X_1, \dots, X_s) \in R_s$ , the hypothesis  $H_s$  is rejected or, as we shall say, the homogeneity of the set  $\{g_1, \dots, g_s\}$  is rejected and the procedure goes on to stage 2 in order to get more detail about the reasons for rejection. If  $(X_1, \dots, X_s)$  falls into the complement  $\bar{R}_s$  of  $R_s$ , we shall employ the traditional terminology of hypothesis testing and say that  $H_s$  or the homogeneity of the set  $\{g_1, \dots, g_s\}$  is accepted. This, of course, is not meant to imply that the data have convinced us of the validity of (2.1) but only that they have not enabled us to reject (2.1). In that case no further tests will be made.

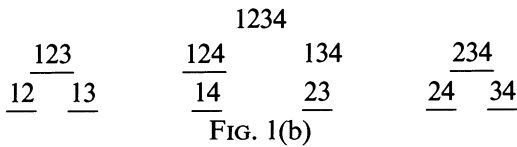
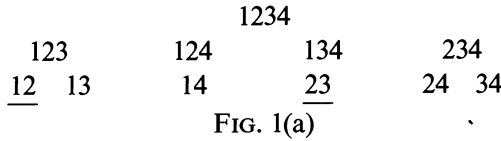
The second stage consists of tests of the  $s$  hypotheses

$$(2.2) \quad H_{s-1; i} : g_1 = \dots = g_{i-1} = g_{i+1} = \dots = g_s$$

by means of rejection regions  $R_{s-1; i}$ . If the  $(s-1)$ -tuple  $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_s)$  falls into the complement of  $R_{s-1; i}$ , homogeneity of the set  $\{g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_s\}$  is accepted and no subset will be tested. If on the other hand  $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_s)$  falls into  $R_{s-1; i}$ , homogeneity of the set  $\{g_1, \dots, g_{i-1}, g_{i+1}, \dots, g_s\}$  is rejected and one proceeds to the third stage.

This procedure is continued until nothing is left to be tested. As a shorthand notation we shall sometimes write  $H_k$  and  $R_k$  instead of the more complete  $H_{k; i_1, i_2, \dots}$  and  $R_{k; i_1, i_2, \dots}$  when the additional subscripts are clear from the context.

As an illustration we give in Figures 1(a) and 1(b) examples of possible outcomes of such a procedure for the case  $s = 4$  where homogeneity has been accepted for the underlined sets and rejected for those not underlined.



Typically, the rejection regions are defined in terms of statistics  $T$  such as ranges or  $\chi^2$ -statistics in the case of normal means, Kruskal-Wallis and Smirnov statistics in the nonparametric case, etc. These are computed for the variables in question and provide measures of inhomogeneity. Then  $R_s$  is the set for which  $T_s(X_1, \dots, X_s) \geq C_s$ ;  $R_{s-1; i}$  is the set for which  $T_{s-1; i}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_s) \geq C_{s-1; i}$ ; and so on. For the  $T$ 's and  $C$ 's we shall frequently delete the subscripts behind the semi-colon as we did for the  $H$ 's and  $R$ 's. In the most commonly used procedures, the statistics  $T$  are obtained from a single functional (for example, the range) evaluated at the respective empirical cumulative distribution functions, but this need not be the case and will not be assumed here.

In the present paper we shall suppose that the statistics  $T_s, T_{s-1}, \dots$  are given and we shall be concerned with the best choice of the critical values  $C_s, C_{s-1}, \dots$ . After some preparation in the next sections, a solution of this problem will be obtained in Section 4.

It follows from the definition of stagewise procedures given above that they have the following properties.

- (i) Rejection of homogeneity for any set of  $g$ 's implies rejection of homogeneity for all sets containing it.
- (ii) Acceptance of homogeneity for any set of  $g$ 's implies acceptance of homogeneity for any set contained in it.
- (iii) Homogeneity of the set  $\{g_i, \dots, g_k\}$  can be rejected only when  $X_{i_1}, \dots, X_{i_k}$  falls into  $R_k$ .

A procedure satisfying properties (i) and (ii) has been termed coherent by Gabriel.

**3. Separability.** We require another property of multistage procedures which is not implied by the definition. (For the normal location case this essentially reduces to condition (X) proposed by Tukey in Chapter 30 of [9]).

Suppose the  $g$ 's fall into  $t$  distinct groups of sizes  $v_1, v_2, \dots, v_t$  respectively ( $\sum v_i = s$ ), say

$$(3.1) \quad g_{i_1} = \dots = g_{i_{v_1}}; g_{i_{v_1+1}} = \dots = g_{i_{v_1+v_2}}; \dots$$

where  $(i_1, i_2, \dots, i_s)$  is a permutation of  $(1, \dots, s)$ . We shall say that the configuration (3.1) is *separable* by a given multistage procedure if there exists a sequence of distributions  $(P_1^{(m)}, \dots, P_s^{(m)})$  with  $P_i^{(m)} \in \mathcal{P}_i$  and satisfying (3.1) such that any hypothesis

$$(3.2) \quad g_{j_1} = \dots = g_{j_t}$$

with subscripts from at least two of the  $t$  distinct subsets of (3.1) is rejected with probability tending to 1 as  $m \rightarrow \infty$ . If every configuration (3.1) is separable by a given multistage procedure, we call the procedure *separating*.

**EXAMPLE 1.** Let  $X_i = (X_{i1}, \dots, X_{in})$  with the  $X_{ij}$  independently normally distributed with mean  $\theta_i$  and known common variance  $\sigma^2$ . Let  $g_i(P_i) = \theta_i$  and let  $X_i$  denote the mean of the  $i$ th sample. If each of the statistics  $T_s, T_{s-1}, \dots$  is for example the range of the  $X_i$ 's in question, or the sum of the squared deviations of the  $X_i$ 's from their mean, the procedure is clearly separating since such a  $T$  tends in probability to infinity if the difference between at least two of the involved  $\theta$ 's tends to infinity.

**EXAMPLE 2.** Let  $X_i = (X_{i1}, \dots, X_{in})$  with the  $X_{ij}$  independently distributed according to unknown continuous distributions  $F_i$ . Let  $g_i(P_i) = F_i$  and suppose that for testing the equality of, say,  $r$  of the  $F_i$ , the associated  $rn$  variables  $X_{ij}$  are ranked and the test is based on the Kruskal-Wallis statistic. For testing  $H: F_1 = \dots = F_r$ , this is given by

$$(3.3) \quad K = \frac{12n}{rn(rn+1)} \sum \left( R_i - \frac{rn+1}{2} \right)^2 = \frac{12}{rn^2(rn+1)} \sum R_i^2 - 3(rn+1),$$

where  $R_i$  is the rank-sum of the  $i$ th sample in the ranking of the  $rn$  variables, and  $R_i = R_i/n$ .

We shall now show that for any given critical values  $C_2, \dots, C_s$  this procedure is separating for sufficiently large  $n$ , provided there exist distributions  $F_i^0 \in \mathcal{F}_i$  (the class of  $F$ 's corresponding to  $\mathcal{P}_i$ ) such that

$$(3.4) \quad P[X_{i\alpha} < X_{j\beta}] = 1 \quad \text{whenever } i < j,$$

where the  $X_{i\alpha}$  and  $X_{j\beta}$  are distributed according to  $F_i^0$  and  $F_j^0$  respectively. It will be an obvious consequence that the conclusion continues to hold if there exist sequences  $F_i^{(m)} \in \mathcal{F}_i$  for which (3.4) holds in the limit. In particular it therefore holds for any location family given by  $F_i(x) = F(x - \theta_i)$  with unrestricted  $\theta$ 's.

To prove this result consider a set of  $g$ 's containing at least one element each from two of the subsets defined by (3.1). Suppose the lowest (according to the ordering (3.4)) and next lowest subsets from which the set contains elements are the  $i$ th and the  $j$ th, respectively, and that the set consists of  $\mathbf{a}$  elements from the  $i$ th,  $\mathbf{b}$  from the  $j$ th and  $\mathbf{c}$  from the other (larger) subsets. Then we shall have  $\mathbf{a}$  samples which are smaller than all other observations,  $\mathbf{c}$  which are larger than all other observations, and  $\mathbf{b}$  in the middle.

The ranks are: in the bottom group  $1, \dots, an$ ; in the middle group  $an + 1, \dots, (a + b)n$ ; in the top group  $(a + b)n + 1, \dots, (a + b + c)n$ . Since the sum of the ranks in each of the three groups is fixed,  $\sum R_i^2$  is minimized when the  $\mathbf{a}$  rank-sums in the bottom group are equal; the  $\mathbf{b}$  in the middle group are equal; and the  $\mathbf{c}$  in the top group are equal. An easy calculation shows that the resulting minimum value of  $K$  is

$$(3.5) \quad K_0(a, b, c) = 3n^2(a + b)(a + c)(b + c) / (a + b + c)[(a + b + c)n + 1],$$

so that

$$(3.6) \quad K \geq K_0(a, b, c)$$

for all samples that are possible under the given configuration. Since  $K$  has a limiting  $\chi^2$ -distribution with  $a + b + c - 1$  degrees of freedom under  $H$  as  $n \rightarrow \infty$ , it follows that  $K$  exceeds the given critical value with probability 1 for sufficiently large  $n$ .

To complete the proof, we must show that it is possible to choose  $n$  large enough so that  $K_0(a, b, c)$  exceeds the critical value for all choices of  $a, b$  and  $c$  that correspond to possible subsets of the given configuration which contain at least two distinct elements. Since the number of such choices is finite, the proof is complete.

EXAMPLE 3. In the situation of the preceding example, suppose that the test statistic for testing (3.2) instead of  $K$  is Kiefer's generalization [5] of the Smirnov statistic

$$(3.7) \quad K' = n \sup_x \sum [F_i^*(x) - \bar{F}^*(x)]^2 = n \sup_x [\sum F_i^{*2}(x) - r\bar{F}^{*2}(x)]$$

where  $F_i^*$  denotes the empirical cdf of the  $i$ th sample and where  $\bar{F}^* = \sum_{i=1}^r F_i^* / r$  is the empirical cdf of all  $rn$  observations.

Again the problem is to find a lower bound for  $K'$  in the case that  $\mathbf{a}$  samples fall below and  $\mathbf{c}$  samples above all other observations, with the remaining  $\mathbf{b}$  forming a middle group. Let  $z$  be a point such that

$$P(X_{i\alpha} \leq z) = P(X_{j\beta} > z) = 1 \quad \text{for all } i = 1, \dots, a$$

$$j = a + 1, \dots, a + b + c.$$

Then  $F_i^*(z) = 1$  and  $F_j^*(z) = 0$  for all  $i \leq a$  and  $a < j \leq a + b + c$  and  $\bar{F}^*(z) = a / (a + b + c)$ . It follows that

$$(3.8) \quad K' \geq K'_0(a, b, c) = na(b + c) / (a + b + c).$$

Since  $K'$  has a nondegenerate limit distribution, it follows as in Example 3 that  $K'$  is separating for sufficiently large  $n$ , provided there exist distributions  $F_i^0 \in \mathfrak{F}_i$  satisfying (3.4), or sequences  $F_i^{(m)} \in \mathfrak{F}_i$  for which (3.4) holds in the limit.

The separating property of the procedure based on  $K'$  holds also for (sequences of) distributions other than those defined by (3.4). An obvious example is provided by symmetric distributions, symmetric about a common point  $\mu$ , with different scales.

EXAMPLE 4. Let  $(X_{i1}, \dots, X_{id})$ ,  $i = 1, \dots, s$  be  $s$  independent multinomial vectors with distributions

$$P(X_{i1} = x_{i1}, \dots, X_{id} = x_{id}) = \frac{n_i!}{x_{i1}! \dots x_{id}!} p_{i1}^{x_{i1}} \dots p_{id}^{x_{id}}, \quad (\sum_{\alpha=1}^d p_{i\alpha} = 1).$$

The problem is that of distinguishing the vectors  $p_i = (p_{i1}, \dots, p_{id})$ .

Consider first the binomial case  $d = 2$ . It is clear that two binomial distributions can be distinguished with probability 1 if and only if one of the two binomial  $p$ 's is zero and the other is one. It follows that when  $d = 2$ ,  $s \geq 3$ , no procedure can be separating.

It might appear at first sight that the arc sin transformation provides an asymptotic way out, since it approximately transforms the binomial distributions into normal distributions with known variance and hence into the situation of Example 1. Notice, however, that Example 1 requires the normal means to tend to  $-\infty$  or  $+\infty$ ; this corresponds to letting the  $p$ 's tend to 0 and 1, which is just where the approximation breaks down.

An extension of the above argument shows that no procedure can be separating in the more general multinomial situation (3.9) when  $s > d$ . On the other hand, any reasonable procedure will be found to be separating when  $s \leq d$  by letting  $p_{ii} = 1$  for  $i = 1, \dots, d$  and  $p_{i\alpha} = 0$  otherwise.

EXAMPLE 5. Let  $X_i$  be  $s$  independent Poisson variables with parameters  $E(X_i) = \lambda_i$  and consider the problem of distinguishing among the  $\lambda$ 's. Suppose the procedure is based on exact tests of the hypotheses

$$\lambda_{i_1} = \dots = \lambda_{i_s}$$

which are carried out conditionally given

$$X_{i_1} + \dots + X_{i_s} = w.$$

To prove separability of a configuration of  $s$   $\lambda$ 's, let the distinct values  $\lambda_i^{(m)}$  tend to infinity in such a way that the ratios  $\lambda_i^{(m)}/\lambda_j^{(m)}$  all tend to limits  $\neq 1$ . Consider now testing the homogeneity of any subset of the given configuration which contains at least two distinct values. As  $m \rightarrow \infty$ , the sum  $w'$  of the associated Poisson variables tends to  $\infty$  in probability. The conditional distribution of these  $X$ 's given  $w'$  is a multinomial distribution corresponding to  $w'$  trials and with

probabilities of which at least two tend to different limits. Then any standard test of homogeneity will reject with probability tending to 1 as  $w' \rightarrow \infty$  and this proves that such a procedure is separating.

**4. Significance levels.** The choice of the critical values  $C_k$  is typically made in terms of the probability of rejecting homogeneity for the subset  $\{g_{i_1}, \dots, g_{i_k}\}$  when in fact

$$(4.1) \quad g_{i_1} = \dots = g_{i_k}.$$

The value of this probability depends also on the remaining  $g$ 's or  $P$ 's since an earlier hypothesis may be accepted thereby depriving the hypothesis (4.1) of the chance of being tested. A quantity of interest in this connection is

$$(4.2) \quad \alpha_k = P_{g_{i_1} = \dots = g_{i_k}}[(X_{i_1}, \dots, X_{i_k}) \in R_k],$$

where  $R_k$  may of course also depend on  $i_1, \dots, i_k$  but where we shall assume that the probability on the right hand side is the same for all  $(P_{i_1}, \dots, P_{i_k})$  satisfying (4.1).

It follows from property (iii) of Section 2 that

$$(4.3) \quad P[\text{rejecting homogeneity of } \{g_{i_1}, \dots, g_{i_k}\}] \leq \alpha_k$$

for all  $(g_1, \dots, g_s)$  satisfying (4.1). Consider now the configuration (3.1) with  $v_1 = k$  and  $v_2 = v_3 = \dots = 1$ , and suppose this configuration is separable by the given procedure. Then there exists a sequence of distributions  $(P_1^{(m)}, \dots, P_s^{(m)})$  satisfying (4.1) and for which the left hand side of (4.3) will tend to  $\alpha_k$ . It follows that

$$(4.4) \quad \sup P[\text{rejecting homogeneity of } \{g_{i_1}, \dots, g_{i_k}\}] = \alpha_k$$

where the sup is taken over all  $(P_1, \dots, P_s)$  satisfying (4.1). The relation (4.4) thus holds in particular whenever the given procedure is separating. For the normal case this result was already noted by Tukey in Chapter 30 of [6]. The levels  $\alpha_k$  are called apparent significance levels by Tukey [9],  $k$ -mean significance levels by Duncan [2], and nominal levels by Einot and Gabriel [3].

An aspect of a multiple comparison procedure which is of great interest is the maximum probability of rejecting the homogeneity of at least one set of  $g$ 's which are in fact equal, say

$$(4.5) \quad \alpha_0 = \sup \alpha(P_1, \dots, P_s)$$

where for a given  $s$ -tuple  $(P_1, \dots, P_s)$

$$(4.6) \quad \alpha(P_1, \dots, P_s) = P[\text{at least one false rejection}]$$

and where the sup in (4.5) is taken over all possible distributions  $P_i \in \mathcal{P}_i$ .

The relation of  $\alpha_0$  to the significance levels  $\alpha_2, \dots, \alpha_s$  can be seen from the following result which underlies all that follows, and which for the normal case was first stated informally by Tukey in Chapter 30 of [9].

**THEOREM 1.** *If the configuration of  $(g_1, \dots, g_s)$  is given by (3.1) and is separable by the given procedure,*

$$(4.7) \quad \sup \alpha(P_1, \dots, P_s) = 1 - \prod_{i=1}^t (1 - \alpha_{v_i})$$

where  $\alpha_1 = 0$  and where the supremum is taken over all  $s$ -tuples  $(g_1, \dots, g_s)$  which satisfy (3.1).

The proof of this theorem is essentially the same as that given for Theorem 1 in [7] and will therefore be omitted here. It follows from (4.7) that

$$(4.8) \quad \alpha_0 = \sup_{v_1, \dots, v_t} [1 - \prod_{i=1}^t (1 - \alpha_{v_i})]$$

holds for all configurations provided the given procedure is separating. It may be of interest to note that separability of the configuration is not only sufficient for Theorem 1 but also necessary.

The remarkable feature of these results is the fact that the suprema in question are independent of the statistics  $T$  defining the procedure, and depend only on the values  $\alpha_2, \dots, \alpha_s$ . Our primary concern in this paper is the choice of these  $\alpha$ 's.

In analogy with standard practice in hypothesis testing, we shall impose a bound on one of the rejection probabilities, namely  $\alpha_0$ , thereby insuring comparability of different procedures. We shall thus fix a value, say  $\alpha_0^*$  and restrict attention to procedures satisfying

$$(4.9) \quad \alpha_0 \leq \alpha_0^*.$$

Subject to this condition, we should like to maximize the "power" of the procedure, that is, the probability of detecting existing inhomogeneities. Since we are assuming the statistics  $T$  to be given and only the critical values  $C$  to be at our disposal, it is clear that power is maximized by maximizing the  $\alpha$ 's.

From (4.8) we see that the problem is that of maximizing  $\alpha_2, \dots, \alpha_s$  subject to

$$(4.10) \quad \prod_{i=1}^t (1 - \alpha_{v_i}) \geq 1 - \alpha_0^* \text{ for all } v_1, \dots, v_t \text{ satisfying } \sum_{i=1}^t v_i = s.$$

Let us begin with  $\alpha_s$ . If the left hand side of (4.10) is to involve  $\alpha_s$ , we must have  $t = 1$ ,  $v_1 = s$  and the only restriction on  $\alpha_s$  is  $\alpha_s \leq \alpha_0^*$ , so that we shall maximize  $\alpha_s$  by putting

$$(4.11) \quad \alpha_s = \alpha_0^*.$$

Analogously, we see that we can also put

$$(4.12) \quad \alpha_{s-1} = \alpha_0^*$$

without imposing any additional restrictions on  $\alpha_2, \dots, \alpha_{s-2}$ . This shows that for  $s = 3$ , there exists a uniformly best choice of the  $\alpha$ 's, namely  $\alpha_2 = \alpha_3 = \alpha_0^*$ .

Consider next the case  $s = 4$ . The possible configurations are  $t = 1$ ,  $v_1 = 4$ ;  $t = 2$ ,  $\{v_1, v_2\} = \{1, 3\}$  or  $\{2, 2\}$ ;  $t = 3$ ,  $\{v_1, v_2, v_3\} = \{1, 1, 2\}$  and  $t = 4$ ,  $v_1 = v_2 = v_3 = v_4 = 1$  and it is seen that a uniformly best choice of the  $\alpha$ 's is given by  $\alpha_3 = \alpha_4 = \alpha_0^*$ ,  $\alpha_2 = 1 - (1 - \alpha_0^*)^{\frac{1}{2}}$ .



This happy state of affairs however does not extend to higher values of  $s$ . For  $s = 5$ , for example, it is no longer possible simultaneously to maximize  $\alpha_2$  and  $\alpha_3$ . The maximum value of  $\alpha_3$  is  $\alpha_0^*$  but this requires  $\alpha_2 = 0$ ; alternatively the maximum value of  $\alpha_2$  is  $1 - (1 - \alpha_0^*)^{\frac{1}{2}}$  in which case  $\alpha_3$  cannot exceed  $1 - (1 - \alpha_0^*)^{\frac{1}{2}}$ .

When a uniformly best choice of the  $\alpha$ 's does not exist, it is convenient to introduce the following definition.

DEFINITION. A set  $(\alpha_2, \dots, \alpha_s)$  for which  $\alpha_0$  satisfies (4.9) will be called inadmissible if there exists another set  $(\alpha'_2, \dots, \alpha'_s)$  for which  $\alpha'_0$  also satisfies (4.9) and such that

$$(4.13) \quad \alpha_i \leq \alpha'_i \quad \text{for all } i, \text{ with strict inequality} \\ \text{for at least some } i.$$

This use of the term is analogous to that which is customary in the theory of hypothesis testing.

THEOREM 2. Any admissible procedure which is separating satisfies

$$(4.14) \quad \alpha_2 \leq \alpha_3 \leq \dots \leq \alpha_s.$$

PROOF. If (4.14) does not hold, there exists  $k$  such that  $\alpha_{k+1} < \alpha_k$ . For any such  $k$ , consider the procedure in which  $\alpha'_i = \alpha_i$  for  $i \neq k+1$  and  $\alpha'_{k+1} = \alpha_k$ . Then clearly  $\alpha'_0 \geq \alpha_0$ . To show that  $\alpha'_0 \leq \alpha_0^*$ , we need only show that  $\prod_{i=1}^t (1 - \alpha'_{v_i}) \geq 1 - \alpha_0^*$  for all  $(v_1, \dots, v_t)$ .

If none of the  $v$ 's is equal to  $k+1$ ,  $\alpha'_{v_i} = \alpha_{v_i}$  for all  $i = 1, \dots, t$  and the result holds. Otherwise, replace each  $v$  that is equal to  $k+1$  by two  $v$ 's: one equal to  $k$  and one equal to  $1$ , and denote the resulting set by  $v'_1, \dots, v'_t$ . Then  $\prod_{i=1}^t (1 - \alpha'_{v_i}) = \prod_{i=1}^t (1 - \alpha_{v'_i}) \geq 1 - \alpha_0^*$ , and this completes the proof.

Having fixed  $\alpha_0 = \alpha_s = \alpha_{s-1}$  at  $\alpha_0^*$ , how should we choose the remaining  $\alpha$ 's? In order to have a reasonable chance of detecting existing inhomogeneities irrespective of their pattern we should like to have none of the  $\alpha$ 's too small. In fact, at the end of the section we shall propose a set of  $\alpha$ -values which for  $s \leq 21$  will keep  $\alpha_0$  less than .1 without letting any of  $\alpha_2, \dots, \alpha_s$  fall below .01. These considerations suggest a minimax approach: namely subject to  $\alpha_0 \leq \alpha_0^*$  to maximize  $\min(\alpha_2, \dots, \alpha_s)$ . In view of Theorem 2, this is achieved by maximizing  $\alpha_2$ .

THEOREM 3. The maximum value of  $\alpha_2$  for a multistage procedure which is separating and satisfies (4.9) is given by

$$(4.15) \quad \alpha_2 = 1 - (1 - \alpha_0^*)^{\lfloor s/2 \rfloor - 1}$$

where  $\lfloor A \rfloor$  as usual denotes the largest integer  $\leq A$ .

PROOF. Instead of fixing  $\alpha_0$  at  $\alpha_0^*$  and maximizing  $\alpha_2$ , it is more convenient to invert the problem and fix  $\alpha_2$  at, say,  $\alpha$  and minimize  $\alpha_0$ . The theorem will be proved by showing that the resulting minimum value of  $\alpha_0$  is

$$(4.16) \quad \alpha_0^* = 1 - (1 - \alpha)^{\lfloor s/2 \rfloor}.$$

Suppose first that  $s$  is even. Since  $\alpha_2$  is fixed at  $\alpha$ , it then follows from the fact that the procedure has been assumed to be separating that the right hand side of (4.7) can be made arbitrarily close to  $\alpha_0^*$ . This is seen by using the separability of the configuration  $v_1 = \cdots = v_{[s/2]} = 2$ .

When  $s$  is odd, we put an additional  $v$  equal to 1 and use separability of the resulting configuration.

In either case it follows that

$$(4.17) \quad \sup_{v_1, \dots, v_s} [1 - \prod_{i=1}^s (1 - \alpha_{v_i})] = \sup \alpha(P_1, \dots, P_s) \geq \alpha_0^*.$$

To complete the proof, we must exhibit values of  $\alpha_2, \dots, \alpha_s$  for which the left hand side of (4.17) equals  $\alpha_0^*$ . To this end consider the procedure defined by

$$(4.18) \quad \alpha_3 = \cdots = \alpha_s = \alpha.$$

The right hand side of (4.7) then becomes  $1 - (1 - \alpha)^{t'}$  where

$$(4.19) \quad t' = \text{number of } v\text{'s} > 1.$$

This quantity takes on its maximum value for  $t' = [s/2]$  and this completes the proof.

Theorems 2 and 3 provide some guidance for the choice of the levels  $\alpha_2, \dots, \alpha_s$  and  $\alpha_0$ . If  $\alpha_2$  is fixed at a conventional level  $\alpha$  independent of  $s$ , then as  $s$  increases, so does  $\alpha_0$  (and hence  $\alpha_{s-1}$  and  $\alpha_s$ ) to values which soon become intolerably high. Conversely, if  $\alpha_0$  is fixed at a conventional level independent of  $s$ , the level  $\alpha_2$  soon becomes so small that there is virtually no chance of detecting moderate differences between pairs when they exist. It turns out however that for  $s \leq 21$ , one can let  $\alpha_2$  vary from .01 to .05 while at the same time keeping  $\alpha_0$  between .05 and .10. The remaining  $\alpha$ 's, by Theorem 2, will then also lie between these limits.

While 10% is larger than is conventionally recommended for a single significance test, it must be remembered that  $\alpha_0$  is the maximum probability of at least one false significance statement when many comparisons are being made. We believe that in these circumstances such a higher level often can be tolerated, and therefore propose the sets of levels of  $\alpha_2$  and  $\alpha_0$  shown in Table 1.

Having decided on  $\alpha_0$  and  $\alpha_2$ , how should one choose the remaining  $\alpha$ 's? We shall show in the next section that when  $s$  is odd there is a uniformly best choice

TABLE 1

$\alpha_2 = .05$		$\alpha_2 = .03$		$\alpha_2 = .02$						
$s$	2	3	4	5	6	7	8	9	10	11
$\alpha_0$	.05	.05	.059	.059	.059	.059	.078	.078	.096	.096

$\alpha_2 = .01$										
$s$	12	13	14	15	16	17	18	19	20	21
$\alpha_0$	.059	.059	.068	.068	.077	.077	.086	.086	.096	.096

for  $\alpha_3, \dots, \alpha_s$ . The result is not quite as clear-cut when  $s$  is even. However, we shall see that even in that case, admissibility determines rather narrow limits within which  $\alpha_2, \dots, \alpha_s$  must lie.

**5. Most powerful test when  $s$  is odd.** We shall now prove the result indicated at the end of the preceding section: that if  $s$  is odd and  $\alpha_2$  and  $\alpha_0$  have been fixed at  $\alpha$  and  $\alpha_0^*$  (given by (4.16)) respectively it is possible to maximize all the remaining  $\alpha$ 's simultaneously. Subject to the two restrictions, the resulting choice thus provides a uniformly most powerful set of  $\alpha$ 's.

To simplify the notation, instead of the probability  $\alpha(P_1, \dots, P_s)$  of at least one error, let us work with its complement

$$(5.1) \quad \gamma(P_1, \dots, P_s) = P(\text{no error}).$$

If we then put

$$(5.2) \quad \gamma_k = 1 - \alpha_k$$

so that in particular

$$(5.3) \quad \gamma_2 = \gamma = 1 - \alpha,$$

it follows from (4.10) and (4.16) that

$$(5.4) \quad \prod_{i=1}^t \gamma_{v_i} \geq \gamma^{\lfloor s/2 \rfloor} = \gamma^{(s-1)/2}.$$

**THEOREM 4.** *Subject to (5.3) and (5.4), the minimum value of  $\gamma_i$  ( $i = 3, \dots, s$ ) is, for any odd  $s$ ,*

$$(5.5) \quad \gamma_i^* = \gamma^{\lfloor i/2 \rfloor},$$

*and this value can be attained simultaneously for all  $i$ .*

**PROOF.** (i) Consider any procedure satisfying (5.3) and (5.4). First let  $i$  be odd and consider the configuration in which  $v_1 = i$  and all the remaining  $v$ 's are equal to 2. Then by (5.4)

$$\gamma_i \gamma^{(s-i)/2} \geq \gamma^{(s-1)/2}$$

and hence

$$(5.6) \quad \gamma_i \geq \gamma_i^*.$$

A completely analogous argument shows that (5.6) also holds when  $i$  is even.

(ii) Consider now the procedure defined by

$$(5.7) \quad \gamma_i = \gamma_i^*.$$

This clearly satisfies (5.3) and it only remains to check that it also satisfies (5.4). Since

$$\prod_{i=1}^t \gamma_{v_i} = \gamma^{\sum_{i=1}^t \lfloor v_i/2 \rfloor},$$

we only need to show that

$$\sum_{i=1}^t \lfloor v_i/2 \rfloor \leq (s-1)/2.$$

Now

$$\begin{aligned} [v_i/2] &= \frac{1}{2}v_i && \text{when } v_i \text{ is even} \\ &= \frac{1}{2}(v_i - 1) && \text{when } v_i \text{ is odd} \end{aligned}$$

so that  $\Sigma[v_i/2] = (s - b)/2$ , where  $b$  is the number of odd  $v$ 's (including ones). Since  $s$  is odd,  $b \geq 1$  and this completes the proof.

The levels defined by (5.7) are close to the levels

$$(5.8) \quad \begin{aligned} \gamma_i &= \gamma^{i/2}, & 1 < i < s - 1 \\ &= \gamma^{s/2}, & i = s - 1, s \end{aligned}$$

proposed by Tukey in Chapter 31 of [9]; see also Ryan [8]. These levels are admissible but, of course, do not satisfy (4.15). Other widely discussed sets of levels have been proposed:  $\gamma_i = \gamma^{i-1}$  by Duncan, and  $\gamma_i = \gamma$  by Newman and Keuls. (See Einot and Gabriel [3] for references.) Neither of these sets is admissible according to our definition. (Actually, the procedures of these authors are for unknown  $\sigma$  while our discussion assumes  $\sigma$  to be known. See Section 6.)

**6. Choice of  $\alpha_3, \dots, \alpha_{s-1}$  when  $s$  is even.** When  $s$  is odd, there exists a uniformly best choice of  $\gamma_3, \dots, \gamma_{s-1}$  satisfying

$$(6.1) \quad \prod_{i=1}^t \gamma_{v_i} \geq \gamma^{[s/2]}$$

where  $\gamma_2 = \gamma$ . The following theorem shows that this is not the case when  $s$  is even.

**THEOREM 5.** *Let  $s$  be even and let the  $\gamma$ 's satisfy (5.3) and (6.1).*

(i) *The values of  $\gamma_3, \dots, \gamma_{s-1}$  are bounded below by*

$$(6.2) \quad \begin{aligned} \gamma_i &\geq \gamma^{i/2} && \text{when } i \text{ is even and when } i \text{ is odd and } 3 \leq i \leq s/2 \\ &\geq \gamma^{(i+1)/2} && \text{when } i \text{ is odd and } s/2 < i \leq s - 1. \end{aligned}$$

(ii) *For any given  $i$ , there exists a set of  $\gamma$ 's for which  $\gamma_i$  attains the lower bound (6.2).*

(iii) *If  $s > 6$ , the lower bounds (6.2) cannot be attained simultaneously for all  $i$ .*

**PROOF.** (i) follows directly from (6.1) with the following values of the  $v$ 's:

$$\begin{aligned} i = \text{even:} & \quad v_1 = i, \quad \text{the other } v\text{'s} = 2; \\ i = \text{odd,} & \quad 1 < i \leq s/2; \quad v_1 = v_2 = i, \quad \text{the other } v\text{'s} = 2; \\ i = \text{odd,} & \quad s/2 < i \leq s - 1: \quad v_1 = 1, \quad v_2 = i, \quad \text{the other } v\text{'s} = 2. \end{aligned}$$

(ii) For the cases with  $\gamma_i = \gamma^{i/2}$ , this is seen by putting  $\gamma_j = \gamma^{j/2}$  for all  $j$ , since then

$$\prod_{i=1}^t \gamma_{v_i} = \gamma^{\frac{1}{2}\Sigma_{i=1}^t v_i} = \gamma^{s/2}$$

which verifies (6.1).

For the cases with  $\gamma_i = \gamma^{(i+1)/2}$ , it is seen by putting

$$\begin{aligned} \gamma_i &= \gamma^{(i+1)/2} \\ \gamma_j &= \gamma^{(j-1)/2} && \text{for all odd } j \neq i \\ \gamma_j &= \gamma^{j/2} && \text{for all even } j. \end{aligned}$$

To verify (6.1) one need only note that at most one  $\gamma$  can be equal to  $\gamma^{(i+1)/2}$  and that the subscript of at least one other  $\gamma$  must be odd.

(iii) If  $s > 6$ ,  $t = 2$ ,  $v_1 = 3$ ,  $v_2 = s - 3$ , then  $v_2$  is odd and  $> s/2$ , and we have

$$\prod_{i=1}^t \gamma_{v_i} < \gamma^{s/2}$$

if the  $\gamma_{v_i}$  are set equal to their lower limits for all  $i$ , so that (6.1) is not satisfied.

It is seen from the proof of (iii) that one cannot lower any of the  $\gamma_i$  with  $i = \text{odd}$  and  $s/2 < i < s - 1$  below  $\gamma^{i/2}$  without increasing at least one of the  $\gamma$ 's with  $i = \text{odd}$  and  $< s/2$  by the same (or a larger) factor above its minimum level. Under these circumstances, a reasonable compromise seems to be to put

$$(6.3) \quad \begin{aligned} \gamma_i &= \gamma_i^* = \gamma^{i/2} && \text{for all } 1 < i < s - 1 \\ &= \gamma^{s/2} && \text{for } i = s - 1, s. \end{aligned}$$

This rule, which was proposed by Tukey [9], is clearly admissible. As in the case of odd  $s$ , the sets of levels proposed by Duncan and by Newman and Keuls are not admissible.

The values (6.3) do not constitute the only admissible set of  $\gamma$ 's, and the question arises as to how far an admissible set can deviate from the compromise solution (6.3). A partial answer is obtained by supplementing the lower bounds (6.2) by a corresponding set of upper bounds for all admissible sets of  $\gamma$ 's. Together, the two sets of bounds can be summarized in the following theorem.

**THEOREM 6.** *Let  $s$  be even and let the  $\gamma$ 's satisfy (5.3) and (6.1). Then any admissible set of  $\gamma$ 's satisfies, for  $i = 3, \dots, s$*

$$(6.4) \quad \begin{aligned} \gamma_i &= \gamma_i^* \text{ when } i = \text{even and } > \frac{s}{2} - 4 \\ &\text{or } i = \text{odd and } = \frac{s}{2} \text{ or } = s - 1 \end{aligned}$$

$$(6.5) \quad \begin{aligned} \gamma_i^* \leq \gamma_i \leq \gamma_i^* \gamma^{-\frac{1}{2}} = \gamma_{i-1}^* \text{ when } i = \text{even and } \leq \frac{s}{2} - 4 \\ \text{or } i = \text{odd and } < \frac{s}{2} \end{aligned}$$

and

$$(6.6) \quad \gamma_{i+1}^* = \gamma_i^* \gamma^{\frac{1}{2}} \leq \gamma_i \leq \gamma_i^* \quad \text{when } i = \text{odd and } \frac{s}{2} < i < s - 1.$$

Before turning to the proof of this result, let us consider its implications. The bounds show that an admissible  $\gamma_i$  can differ from  $\gamma_i^*$  by at most  $\gamma^{i/2} - \gamma^{(i+1)/2}$  or  $\gamma^{(i-1)/2} - \gamma^{i/2}$ . For  $i \geq 3$ , the largest possible value of  $|\gamma_i - \gamma_i^*| = |\alpha_i - \alpha_i^*|$  is thus

$\gamma(1 - \gamma^{\frac{1}{2}})$  which for all  $0 < \gamma < 1$  is  $\leq \frac{1}{2}(1 - \gamma) = \frac{1}{2}\alpha$ . For the values of  $\alpha$  proposed in Section 3 an admissible  $\gamma_i$  can therefore differ from  $\gamma_i^*$  by at most .01 for  $s = 8$  and 10, and by at most .005 for any even  $s$  between 12 and 20.

The other point worth noting is that Theorem 2 is a consequence of Theorem 4 (for  $s = \text{odd}$ ) and Theorem 6 (for  $s = \text{even}$ ). In the former case this is obvious; in the latter it follows from the fact that if the lower and upper bounds for  $\gamma_i$  are denoted by  $\underline{\gamma}_i$  and  $\bar{\gamma}_i$ , then

$$\bar{\gamma}_{i+1} \leq \underline{\gamma}_i \leq \bar{\gamma}_i \leq \gamma_{i-1}.$$

**SKETCH OF PROOF OF THEOREM 6.** To establish the upper bounds  $\bar{\gamma}_i$  stated in Theorem 6, a single method of proof suffices for all cases. Consider any set  $(\gamma_3, \dots, \gamma_i, \dots, \gamma_s)$  with  $\gamma_i > \bar{\gamma}_i$  and satisfying (6.1); then the set in which  $\gamma_i$  is replaced by  $\bar{\gamma}_i$  also satisfies (6.1) and hence proves the first set inadmissible. We shall not carry out this check but should perhaps explain the somewhat surprising quantity  $(s/2) - 4$  which appears in (6.4) and (6.5) when  $i$  is even. This stems from the fact that if  $\gamma_i > \gamma_i^*$  for an even  $i$ , it follows from Theorem 5 that admissibility requires a compensating odd  $j$  with  $\gamma_j < \gamma_j^*$ . Since the  $v$ 's must add up to  $s$  which is even, at least one other  $v$  must be odd. If the only other odd  $v$  is equal to 1, the left hand side of (6.1) is  $\geq \gamma^{s/2}$  even when  $\gamma_i = \gamma_i^*$ . It thus requires an odd factor with  $v \geq 3$  to justify letting  $\gamma_i > \gamma_i^*$ . Hence  $i + j + 3 \leq s$  where  $j > s/2$ . Thus  $\gamma_i > \gamma_i^*$  is possible only if  $i < s - \frac{s}{2} - 3 = \frac{s}{2} - 3$ , or  $i \leq \frac{s}{2} - 4$  as was to be proved.

## 7. Applications and further extensions.

(i) *The normal case.* The results of the preceding sections apply to the normal location problem with known variances  $\sigma_i^2$ . Let us now consider the more important situation in which the variances are unknown.

(a) Suppose first that the unknown variances are assumed to be equal (or have known ratios). If the usual procedure for estimating the unknown  $\sigma^2$  is followed, the relations between the  $\alpha$ 's obtained in the preceding sections will no longer hold exactly. In particular, if we define  $\alpha_3, \dots, \alpha_s$  as before in terms of  $\alpha = \alpha_2$ , the quantity  $\alpha_0^*$  given by (4.16) will continue to be a sharp upper bound for  $\alpha_0$ . However, given the same value of  $\alpha_s = \alpha_{s-1}$  we can now raise  $\alpha_2, \dots, \alpha_{s-2}$  without raising  $\alpha_0^* = \alpha_s$ . This follows from the fact that the  $s$  Studentized sample means will now be positively dependent instead of independent as were the means without Studentization (see [7]). The dependence, and hence the permissible increase in  $\alpha_2, \dots, \alpha_{s-1}$  is, however, likely to be so small when  $\sigma^2$  is estimated with at least moderate degrees of freedom, as to be of no practical importance. The conclusions may therefore be expected to hold to a close approximation unless the number of degrees of freedom is quite small.

(b) If a common variance is not assumed (as it often should not be), the situation is somewhat different. No exact test of homogeneity is then available; however a test proposed by Welch [10] seems to provide a highly satisfactory

approximate test. (See Kohr and Games [6] for a comparison with several other tests.) The values of  $\alpha_2, \dots, \alpha_s$  defined in Sections 4, 5 and 6 are then no longer exact and in fact depend slightly on the true values of the variance ratios. However if none of the sample sizes is too small, the approximation can be expected to be close and to that degree the results of the paper continue to apply.

(c) Another problem of interest in the normal case with unknown variances is the comparison of the variances instead of the means. The results of the paper apply to this problem with any of the standard tests of homogeneity of variances.

(ii) *Lowering  $\alpha_s$ .* We saw earlier that subject to (4.10) the maximum value of  $\alpha_s$  is given by (4.11) and that setting  $\alpha_s$  at that value puts no new restrictions on the remaining  $\alpha$ 's. However, some practitioners attach special importance to  $\alpha_s$ , the probability of falsely rejecting the overall null hypothesis  $g_1 = \dots = g_s$ , and would prefer a smaller value for this probability. What are the consequences of lowering  $\alpha_s$  from the value  $\alpha_0$  of Table 1 to a lower value such as .05 or even .01? Theorem 1 shows that such a lowering, with values of  $\alpha_2, \dots, \alpha_{s-1}$  remaining unchanged, will not affect any of the maximum probabilities (4.7), except when  $g_1 = \dots = g_s$ , although it will lower the probabilities for all actual configurations (3.1). The resulting procedure will of course no longer be admissible according to our definition of this concept. In spite of this, one may wish to impose the additional requirement that the probability of falsely rejecting the overall null hypothesis not exceed the new lower value assigned to  $\alpha_s$ . In this case, if the original procedure is admissible under the requirements (5.3) and (6.1), the procedure obtained by lowering  $\alpha_s$  to this new value will clearly be admissible among the procedures satisfying (5.3), (6.1) and the additional bound on  $\alpha_s$ .

(iii) *Deleting stages; Fisher's LSD test.* The following method, not covered by the multistage procedures described in Section 2, although it is concerned with the comparison of a set of normal means, is often referred to as Fisher's least significant difference test. It consists of first carrying out an  $F$ -test say, at level  $\alpha$ , of the hypothesis  $H_s: g_1 = \dots = g_s$ ; if  $H_s$  is accepted, the procedure stops and homogeneity of the set  $\{g_1, \dots, g_s\}$  is accepted; in case of rejection, all  $\binom{s}{2}$  pairs  $\{g_i, g_j\}$  are compared at level  $\alpha$  by means of  $t$ -tests. (If  $\sigma^2$  is known, the  $F$ - and  $t$ -tests are replaced by the corresponding  $\chi^2$  and normal tests.)

Let us now consider the more general set-up of Section 2 but restrict ourselves to the following two stages. First, the overall test of  $H_s$  is performed at level, say,  $\alpha'$ . If  $H_s$  is accepted, the procedure stops and homogeneity of the full set  $\{g_1, \dots, g_s\}$  is accepted. In case of rejection, all pairs are tested by means of the rejection regions  $R_2$ , each at level  $\alpha$ . The probability of at least one false rejection is then given by the formula

$$\begin{aligned}
 (7.1) \quad & \sup P[\text{at least one false rejection}] \\
 &= \alpha' \quad \text{if } g_1 = \dots = g_s \\
 &= 1 - \prod_{i=1}^s (1 - \alpha_{v_i}) \quad \text{if (3.1) holds with } t > 1,
 \end{aligned}$$

where the sup extends over all  $(g_1, \dots, g_s)$  satisfying the condition indicated on the right hand side of (7.1) and where  $\alpha_k$  is given by

$$(7.2) \quad \alpha_k = P[(X_i, X_j) \in R_2 \text{ for at least one pair } 1 \leq i < j \leq k].$$

Consider a fixed value of

$$(7.3) \quad \alpha_0 = \max[\alpha'; 1 - \prod_{i=1}^t (1 - \alpha_{v_i}), \text{ all configurations }].$$

The most powerful choices of  $\alpha$  and  $\alpha'$  are then clearly given by

$$(7.4) \quad \alpha' = \alpha_0$$

and

$$(7.5) \quad \max[1 - \prod_{i=1}^t (1 - \alpha_{v_i})] = \alpha_0$$

where the maximum is taken with respect to all configurations  $(v_1, \dots, v_t)$  with  $\sum_{i=1}^t v_i = s$  and  $t > 1$ .

(iv) *A subset option.* The generalization (iii), required to cover the least significant difference and related procedures, consisted in allowing the deletion of some of the stages in the multistage procedure of Section 2.

The most commonly used multistage comparison procedures, the multiple range tests as customarily described (discussed for example in [7]), also do not fall within the framework of our analysis but for a different reason. To cover procedures of this kind, let us consider a multistage procedure with the following added *subset option*. When homogeneity of a set  $\{P_{i_1}, \dots, P_{i_k}\}$  is rejected we permit the additional conclusion that a subset of these  $P$ 's, selected on the basis of the data from a specified family of subsets, is also rejected. If such a subset is rejected, all parameter sets containing this subset are also rejected and will not be tested any more. Actually, in what follows we shall apply this option only to the case that the subsets are pairs. The procedures we shall discuss are somewhat less powerful than those of Sections 5 and 6; however, they have the advantage that whenever homogeneity of a set is rejected, at least one pair in this set will also be declared to be inhomogeneous.

Let us assume that  $X_i = (X_{i1}, \dots, X_{in})$  and that the rejection regions  $R_k$  of Section 2 are defined in terms of statistics  $T_s, T_{s-1}, \dots$  and critical values  $C_s, C_{s-1}, \dots$  as described there but that now the  $T$ 's and  $C$ 's depend only on the size of the subset being tested, so that they do not require any additional subscripts.

If the procedure has not already stopped at an earlier stage, homogeneity of a pair  $(X_i, X_j)$  is then rejected whenever

$$(7.6) \quad T_2(X_i, X_j) \geq C_2.$$

We shall restrict attention to procedures in which the remaining  $T$ 's are defined by

$$(7.7) \quad T_k(X_1, \dots, X_k) = \max[T_2(X_i, X_j); 1 \leq i < j \leq k],$$

and then adopt the following rule: whenever homogeneity of a set  $\{g_{i_1}, \dots, g_{i_k}\}$  is rejected because  $T_k(X_{i_1}, \dots, X_{i_k})$  is too large, then homogeneity of the pair of  $g$ 's



corresponding to the two  $X$ 's for which  $T_2(X_i, X_j)$  is maximum, is also rejected.

A well-known example of a procedure of this kind in the normal case is that based on the range: that is, in which

$$(7.8) \quad T_2(X_i, X_j) = |X_j - X_i|$$

where  $X_i$  and  $X_j$  denote the means of the  $i$ th and  $j$ th sample.

Another illustration is provided if in the notation of the Smirnov example (Example 3) of Section 4 we take

$$(7.9) \quad T_2(X_i, X_j) = \sup_x |F_i^*(x) - F_j^*(x)|.$$

As a third example, suppose that for each  $i$  and  $j$  the  $2n$  observations constituting the  $i$ th and  $j$ th samples are ranked from 1 to  $2n$  and that  $R_i$  and  $R_j$  denote the average rank of the  $i$ th and  $j$ th sample in this ranking. Then we could take

$$(7.10) \quad T_2(X_i, X_j) = |R_j - R_i|.$$

Consider finally the multinomial example of Section 3 with  $n_i = n$  for all  $i$ . Here we can take, for instance for  $T_2(X_i, X_j)$  the usual  $\chi^2$ -statistic for comparing two multinomial distributions. If  $n$  is sufficiently large,  $C_2$  can be obtained approximately from the  $\chi^2$ -distribution with  $d - 1$  degrees of freedom and the resulting procedure satisfies the assumptions made above.

With this type of procedure, properties (i) and (ii) of Section 2 continue to hold; however, property (iii) need no longer be valid, since it may now happen that homogeneity of a set  $\{g_i, \dots, g_k\}$  is rejected at an early stage in spite of the fact that  $(X_i, \dots, X_k)$  falls into  $\bar{R}_k$ . As shown in the case of multiple range tests in [7], Theorem 1 need then no longer be true, and the analysis of the preceding sections is no longer applicable.

Whether or not (iii), and hence Theorem 1, holds in the present situation typically will depend on the critical values  $C$ . However, a sufficient condition for them to hold is

$$(7.11) \quad C_2 \leq \dots \leq C_s.$$

Under mild additional assumptions which are satisfied in all our examples, (7.11) is also necessary. These statements follow from the argument given for the case of range-based procedures in [7].

The values of the  $C$ 's of course depend on the  $\alpha$ 's, and unfortunately it turns out that for the  $\alpha$ 's given by (5.3), (5.7) and (6.3) the  $C$ 's will not always satisfy (7.11). There is then no guarantee that  $\alpha_0$  will have the prescribed value  $\alpha_0^*$ . To avoid this difficulty we can proceed as follows. We can replace  $C_j$  by

$$(7.12) \quad C_j' = \max(C_2, \dots, C_j)$$

so that the resulting  $C_j'$  will satisfy (7.11). We shall now consider the properties of this approach. Note that for this discussion, we need not concern ourselves with the subset option. We are only concerned with the relationships of the  $\alpha_k$  and  $\alpha_k'$

defined by (7.12) and

(7.13)

$$\alpha_k = P[T_k(X_1, \dots, X_k) \geq C_k], \quad \alpha'_k = P[T_k(X_1, \dots, X_k) \geq C'_k]$$

where the probabilities are computed assuming  $g_1 = \dots = g_k$ .

We shall begin by proving the following lemma.

LEMMA. *If  $\alpha_2, \dots, \alpha_s$  are any levels satisfying  $\alpha_2 = \alpha$ ,  $\alpha_k \leq \alpha_{k+1}$  for all  $k$  and if  $\alpha_0$  is equal to  $\alpha_0^*$  given by (4.16), then the  $\alpha_k$  and  $\alpha'_k$  satisfy the following relationships:*

$$(7.14) \quad \alpha'_2 = \alpha_2, \quad \alpha'_k \leq \alpha_k \quad \text{for } k = 3, \dots, s$$

and

$$(7.15) \quad \alpha'_2 \leq \alpha'_3 \leq \dots \leq \alpha'_s \leq \alpha'_0 = \alpha_0.$$

PROOF. The relations (7.14) follow from the facts that  $C'_2 = C_2$  and  $C'_k \geq C_k$  for  $k > 2$ . To prove (7.15) note that  $C'_{k+1} = \max(C'_k, C_{k+1})$ .

If  $C'_k \leq C_{k+1}$  we have  $\alpha'_{k+1} = \alpha_{k+1} \geq \alpha_k \geq \alpha'_k$  as desired. If  $C_{k+1} < C'_k$ , we have

$$\alpha'_{k+1} = P[T_{k+1}(X_1, \dots, X_{k+1}) \geq C'_{k+1}] > P[T_k(X_1, \dots, X_k) \geq C'_k] = \alpha'_k$$

and this shows that  $\alpha'_k \leq \alpha'_{k+1}$  for all  $k = 2, \dots, s$ . That  $\alpha'_s \leq \alpha'_0$  is obvious. Finally, it follows from (7.14) that  $\alpha'_0 \leq \alpha_0$  and hence from the proof of Theorem 3 that  $\alpha'_0 = \alpha_0$ .

Since the levels  $\alpha'_k$  satisfy the requirements

$$(7.16) \quad \alpha = \alpha'_2 \leq \alpha'_k \leq \alpha'_0 = \alpha_0^*$$

for all  $k$ , the procedure defined by the  $C'_k$  is a possible procedure satisfying the conditions imposed earlier. Is it admissible within the class satisfying the additional restrictions (7.11)?

We shall now show that this is the case when  $s$  is odd. The proof breaks down for even  $s$ , and we shall deal with this case in a more pragmatic manner.

THEOREM 7. *Let  $s$  be odd. Then the procedure  $\Pi'$  obtained by applying (7.12) to the procedure defined by (5.7) is admissible, and is in fact the unique admissible procedure among all those satisfying (7.16) and (7.11).*

PROOF. Suppose that  $\Pi'$  is not the unique admissible procedure. Then we shall show that there exists a procedure  $\Pi''$  with levels  $\alpha''$  satisfying (7.16) and (7.11) and such that

$$(7.17) \quad \alpha''_k > \alpha_k \quad \text{for some } k.$$

This contradicts the consequence of Theorem 4 that there exists a unique admissible procedure satisfying (7.16), and hence proves that the  $\alpha'$ -procedure is admissible.

To prove the existence of the required  $\alpha''$ , note that if  $\Pi'$  is not the unique admissible procedure satisfying (7.11) and (7.16), there exists a procedure  $\Pi''$  satisfying these conditions and such that  $\alpha'_k < \alpha''_k$  for some  $k$ . If  $\alpha'_k = \alpha_k$ , then

$\alpha_k < \alpha_k''$  as was to be proved. If  $\alpha_k' < \alpha_k$ , then by (7.12) there exists  $j < k$  such that  $C_j = C_k'$ . Since  $C_k'' < C_k'$  we have by (7.11)

$$C_j'' < C_k'' < C_k' = C_j$$

and hence  $\alpha_j'' > \alpha_j$ , as was to be proved.

The above argument makes essential use of the uniqueness of the admissible solution in the case that  $s$  is odd and no longer applies when  $s$  is even and  $\geq 8$ . As an example of what may occur when  $s$  is even, we will consider the normal location problem with known variance  $\sigma^2$ . Easy calculations (using Harter's tables [4]) show that for the even values  $s = 8, 10, \dots, 20$ , reversals of the  $C$ 's occur only in the cases  $s = 8$ ,  $C_7^* < C_6^*$  and  $s = 12$ ,  $C_{11}^* < C_{10}^*$  and possibly in the cases  $s = 10$ ,  $(C_9^*, C_8^*)$ ;  $s = 14$ ,  $(C_{13}^*, C_{12}^*)$  and  $s = 16$ ,  $(C_{15}^*, C_{14}^*)$ . In these latter three cases, the  $C^*$ 's are equal to the two decimal places recorded in the tables. In each of these cases (7.12) would require increasing  $C_{s-1}$  to the value of  $C_{s-2}^*$ , and no other  $C$ -value could be decreased to compensate for this increase without violating (6.1). Therefore, in this normal location problem, the  $C'$ -procedures are admissible, at least for all  $s \leq 21$ .

**Acknowledgment.** The authors wish to thank Emil Spjøtvoll for pointing out a possible generalization of their original results which greatly improved the paper. They would also like to thank the referee for a most careful reading of the manuscript and a number of very helpful comments.

#### REFERENCES

- [1] DUNCAN, C. B. (1952). Testing the homogeneity of treatment means in an analysis of variance of engineering data. Paper presented at annual meetings, Amer. Statist. Assoc., Chicago, December 1952.
- [2] DUNCAN, D. B. (1955). Multiple range and multiple  $F$ -tests. *Biometrics* **11** 1-42.
- [3] EINOT, ISRAEL and GABRIEL, K. R. (1975). A study of the powers of several methods of multiple comparisons. *J. Amer. Statist. Assoc.* **70** 574-583.
- [4] HARTER, H. L. (1969). *Order Statistics and their use in Testing and Estimation*. Vol. I. Aerospace Research Laboratories, Office of Aerospace Research, U.S. Air Force.
- [5] KIEFER, J. (1959).  $K$ -sample analogues of the Kolmogorov-Smirnov and Cramer-von Mises tests. *Ann. Math. Statist.* **30** 420-447.
- [6] KOHR, RICHARD L. and GAMES, PAUL A. (1974). Robustness of the analysis of variance, the Welch procedure and a Box procedure to heterogeneous variances. *J. Experimental Education* **43** 61-69.
- [7] LEHMANN, E. L. and SHAFFER, JULIET POPPER (1977). On a fundamental theorem in multiple comparisons. *J. Amer. Statist. Assoc.* **72** 576-578.
- [8] RYAN, THOMAS A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. *Psychological Bulletin* **57** 318-328.
- [9] TUKEY, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton Univ.
- [10] WELCH, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika* **38** 330-336.

DEPARTMENT OF STATISTICS  
UNIVERSITY OF CALIFORNIA  
BERKELEY, CALIFORNIA 94720