A Comparison of Procedures for Multiple Comparisons of Means With Unequal Variances

AJIT C. TAMHANE*

Nine procedures for multiple comparisons of means with unequal variances are reviewed. Modifications in some procedures are proposed either for improvement in their performance or easier implementation. A Monte Carlo sampling study is carried out for pairwise differences as well as a few selected contrasts and the procedures are compared based on the results of this study. Recommendations for the choice of the procedures are given. Robustness of two procedures designed for homogeneous variances under violation of that assumption is also examined in the Monte Carlo study.

KEY WORDS: Multiple comparisons; Unequal variances; One-way fixed-effects ANOVA; Behrens-Fisher problem.

1. INTRODUCTION

Consider the usual one-way fixed-effects model of analysis of variance:

$$x_{ij} = \mu_i + e_{ij} \ ,$$

where all the e_{ij} are independent with $e_{ij} \sim N(0, \sigma_i^2)$ for $j = 1, 2, \ldots, n_i$; $i = 1, 2, \ldots, k$. The means μ_i and variances σ_i^2 are assumed to be unknown. Let $\bar{x}_i$ denote the sample mean and let s_i^2 denote an unbiased estimate of σ_i^2 based on ν_i degrees of freedom (df) that is independent of $\bar{x}_i$; for the most part we shall take s_i^2 to be the usual sample variance based on $\nu_i = n_i - 1$ df.

In recent years considerable attention has been focused on the problem of multiple comparisons among the μ_i when the σ_i^2 are unequal; for example, see Ury and Wiggins (1971), Spjøtvoll (1972), Brown and Forsythe (1974), Games and Howell (1976), Hochberg (1976), Tamhane (1977), and Dalal (1978). The primary purpose of the present article is to give a brief review of these procedures, point out any relationships between them, propose improvements in some procedures and, finally, make comparisons based on an extensive Monte Carlo (MC) sampling study. The criteria used for comparison are (a) the confidence level of the joint confidence intervals (CI's) of a suitable family of parametric functions (pairwise differences or contrasts) for the μ_i and (b) the widths of these CI's. Note that these two quantities respectively correspond to (a) the familywise Type I error rate and (b) the power of the simultaneous test procedures based on the corresponding joint confidence procedures.

The procedures included in this study fall into two groups. One group consists of procedures having resolutions (Gabriel 1969) for all linear combinations of the μ_i; the second group consists of procedures having resolutions for all pairwise differences that can be extended to all contrasts among the μ_i. (Brown and Forsythe's procedure also has resolution for all contrasts but it is somewhat different from the other procedures in the second group as pointed out in Section 2.1.2.) By carrying out the MC study for pairwise as well as general contrasts we have tried to provide "homegrounds" for both groups of procedures, thus making the comparison fair to the extent that is possible. It would have been preferable if a simple modification of the procedures in the first group were available that would reduce their resolution from all linear combinations to all contrasts. No such modification seems to exist, however.

The secondary purpose of this article is to study the robustness properties under variance heterogeneity of two generalized-Tukey procedures (generalized to cover the case of unequal n_i's) that are designed for the homogeneous variances case. These two procedures (Spjøtvoll and Stoline's 1973 Ext-T and Hochberg's 1974 GT2) were selected for comparison because they perform quite well relative to their competitors (see Ury 1976). The reason for including only the generalized Tukey and not, for example, the Scheffé procedure, was because of our predominant interest in pairwise comparisons for which the Tukey-type procedures are known to be more powerful. The second reason for not including more procedures was, of course, to keep the size of the study to manageable proportions. The inclusion of the generalized-Tukey procedures also serves a subsidiary purpose; namely, when the σ_i^2 are in fact equal, as is the case for some configurations studied in the sequel, they provide a standard against which the performance of the procedures designed for unequal σ_i^2 can be compared.

In the remainder of this article, Section 2 reviews the various procedures; Section 3 proposes modifications of

* Ajit C. Tamhane is Assistant Professor, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60201. The author is grateful to two referees, an associate editor, and the Editor for pointing out several references and suggesting many improvements in the earlier draft. This work is supported by NSF Grant ENG 77-06112.

© Journal of the American Statistical Association
June 1979, Volume 74, Number 366
Theory and Methods Section

some of the procedures either for their easier implementation in practice or for improvement in their performance. In part of Section 3 and in Section 4 some preliminary comparisons among the competing procedures are carried out, a few noncontenders being eliminated in the process. The details of the MC study and the results are presented in Section 5. Finally, the discussion of the results and recommendations for the choice of the procedures are given in Section 6.

2. A REVIEW OF THE PROCEDURES

2.1 Procedures for Unequal σ_i^2

The procedures in the first group guarantee the designated joint confidence level exactly. Because the contrasts problem is a generalization of the Behrens-Fisher problem for which no exact solution is known to exist, the procedures in the second group are inexact; that is, they are either conservative or approximate. Now we describe the procedures in the two groups.

2.1.1 Procedures for all linear combinations of the μ_i. Dalal (1978) proposed a family of procedures based on Hölder's inequality. Let $p \geq 0$, $q \geq 0$ satisfy $1/p + 1/q = 1$, and let $d_{q,\alpha} = d_{q,\alpha}(\nu_1, \ldots, \nu_k)$ denote the upper α point of the distribution of $d_q = (\sum_{i=1}^k |t_{\nu_i}|^q)^{1/q}$ where t_ν denotes a Student's t random variable (rv) with ν df and the t_{ν_i} are independent ($i = 1, \ldots, k$). Then the exact 100 $(1 - \alpha)\%$ joint CI's for all linear combinations $\sum_{i=1}^k a_i\mu_i$ are given by

$$\sum_{i=1}^k a_i\mu_i \in [\sum_{i=1}^k a_i\bar{x}_i \pm d_{q,\alpha}(\sum_{i=1}^k |a_i|^p s_i^p/n_i^{p/2})^{1/p}] .$$

Dalal remarks that these CI's are competitive for all q, although it is possible that for a specific subclass one might dominate the others. In general it is difficult to compute the distribution of d_q and, consequently, to compute $d_{q,\alpha}$. For $q = \infty$, $p = 1$ it is easy to see that $d_{\infty,\alpha} = d_{\infty,\alpha}(\nu_1, \ldots, \nu_k)$ is given by the solution in d to the equation

$$\prod_{i=1}^k \{2F_{\nu_i}(d) - 1\} = 1 - \alpha , \tag{2.1}$$

where $F_\nu(\cdot)$ denotes the distribution function of a t_ν rv. We shall refer to this procedure as the D procedure.

For the special case $q = 2$, the above CI's were proposed by Spjøtvoll (1972). He approximated the distribution of $d_2^2 = \sum_{i=1}^k t_{\nu_i}^2$ by a scaled F distribution and by using the method of moments gave the following approximation to $d_{2,\alpha} = d_{2,\alpha}(\nu_1, \ldots, \nu_k)$:

$$d_{2,\alpha}^2 \cong aF_\alpha(k, b) , \tag{2.2}$$

where $F_\alpha(m, n)$ denotes the upper α point of an F distribution with m and n df,

$$b = \frac{(k-2)[\sum_{i=1}^k \{\nu_i/(\nu_i-2)\}]^2 + 4k\sum_{i=1}^k \{\nu_i^2(\nu_i-1)/(\nu_i-2)^2(\nu_i-4)\}}{k\sum_{i=1}^k \{\nu_i^2(\nu_i-1)/(\nu_i-2)^2(\nu_i-4)\} - [\sum_{i=1}^k \{\nu_i/(\nu_i-2)\}]^2} ,$$

and

$$a = (1 - 2/b) \sum_{i=1}^k \{\nu_i/(\nu_i - 2)\} .$$

We shall refer to this procedure as the S procedure.

Hochberg (1976) proposed the following procedure based on a generalization of the Tukey method of multiple comparisons. Let $h_\alpha' = h_\alpha'(\nu_1, \ldots, \nu_k)$ denote the upper α point of the augmented range R' of $t_{\nu_1}, \ldots, t_{\nu_k}$; $R' = R'(\nu_1, \ldots, \nu_k) = \max\{\max_i |t_{\nu_i}|, \max_{i,j} |t_{\nu_i} - t_{\nu_j}|\}$. Then the exact $100(1 - \alpha)\%$ joint CI's for all linear combinations $\sum_{i=1}^k a_i\mu_i$ are given by

$$\sum_{i=1}^k a_i\mu_i \in [\sum_{i=1}^k a_i\bar{x}_i \pm h_\alpha' M(b_1, \ldots, b_k)] ,$$

where

$$M(b_1, \ldots, b_k) = \max(\sum_{i=1}^k b_i^+, \sum_{i=1}^k b_i^-) , \tag{2.3}$$

$b_i^+ = \max(b_i, 0)$, $b_i^- = \max(-b_i, 0)$, and $b_i = a_i s_i/\sqrt{n_i}$ ($i = 1, \ldots, k$). We shall refer to this procedure as the H1 procedure.

Because of difficulties in computing h_α', Hochberg suggested instead using h_α = the upper α point of the range R of $t_{\nu_1}, \ldots, t_{\nu_k}$; $R = R(\nu_1, \ldots, \nu_k) = \max_{i,j} |t_{\nu_i} - t_{\nu_j}|$; he noted that h_α' is well approximated by h_α provided $k \geq 3$ and $\alpha \leq .05$. In Section 5.2 we have developed an integral expression for the distribution of R' based on which h_α' can be easily computed for arbitrary combinations of the ν_i. In the MC studies we used H1 in its exact form.

2.1.2 Procedures for all contrasts among the μ_i. In this group we include the procedures proposed by Ury and Wiggins, Brown and Forsythe, Games and Howell, Hochberg, and Tamhane. All these procedures except that of Brown and Forsythe involve first constructing joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ and then extending them to all contrasts by using Lemma 3.1 of Hochberg (1975). (This extension is not in the original articles of Ury and Wiggins and Games and Howell.) Thus, if w_{ij} denotes the width of the $100(1 - \alpha)\%$ joint CI for $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$, then the corresponding joint CI's for all contrasts $\sum_{i=1}^k c_i\mu_i$, where $\sum_{i=1}^k c_i = 0$, are given by

$$\sum_{i=1}^k c_i\mu_i \in \left[\sum_{i=1}^k c_i\bar{x}_i \pm \frac{\sum_{i\in\mathcal{P}(\mathbf{c})} \sum_{j\in\mathfrak{N}(\mathbf{c})} c_i(-c_j)w_{ij}}{\sum_{i=1}^k |c_i|}\right], \tag{2.4}$$

where $\mathbf{c} = (c_1, \ldots, c_k)'$, $\mathcal{P}(\mathbf{c}) = \{i: c_i > 0\}$, and $\mathfrak{N}(\mathbf{c}) = \{j: c_j < 0\}$. On the other hand, Brown and Forsythe obtain joint CI's for all contrasts directly using Scheffé's projection method. Also the previous procedures typically involve bounding the probability of the joint statement related to the CI's for $\binom{k}{2}$ differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ by a function of the probabilities of individual statements by means of a Bonferroni-type inequality. All the procedures are based on some solution to the Behrens-Fisher problem, usually

Welch's (1938) approximate solution. Thus, most of the procedures described in the following paragraphs are approximate-conservative (approximate because of the Welch solution; conservative because of the Bonferroni-type inequality used). The validity of the Welch solution in the case of the two-sample problem has been demonstrated by Wang (1971). Wang recommends that it should be true that $n_i \geqq 6$ for each sample.

Ury and Wiggins (1971) proposed a procedure based on the Welch approximate solution and the Bonferroni inequality. According to this procedure the approximate-conservative $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ are given by

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm t_{\hat{\nu}_{ij},\beta}(s_i^2/n_i + s_j^2/n_j)^{\frac{1}{2}}] ,$$

where $t_{\nu,\beta}$ denotes the upper β point of a t_ν rv, $\beta = \alpha/2k'$, $k' = \binom{k}{2}$, and

$$\hat{\nu}_{ij} = \frac{(s_i^2/n_i + s_j^2/n_j)^2}{\{s_i^4/n_i^2(n_i - 1) + s_j^4/n_j^2(n_j - 1)\}} \cdot \quad (2.5)$$

We shall refer to this procedure as the UW procedure. Ury and Wiggins advocated a slight modification of (2.5), but we shall take up this modification in Section 3.2.

A closely related procedure was given by Hochberg (1976). According to this procedure the approximate-conservative $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ are given by

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm g_\alpha(s_i^2/n_i + s_j^2/n_j)^{\frac{1}{2}}] ,$$

where g_α solves the equation

$$\sum_{i=1}^{k} \sum_{j=i+1}^{k} P\{|t_{\hat{\nu}_{ij}}| > g\} = \alpha$$

in g, and $\hat{\nu}_{ij}$ is given by (2.5). We shall refer to this procedure as the H2 procedure.

Games and Howell (1976) proposed the following approximate $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$

$$\mu_i - \mu_j \in \left[\bar{x}_i - \bar{x}_j \pm q_{\hat{\nu}_{ij},k,\alpha} \cdot \frac{1}{\sqrt{2}} (s_i^2/n_i + s_j^2/n_j)^{\frac{1}{2}}\right],$$

where $q_{\nu,k,\alpha}$ denotes the upper α point of the studentized range distribution (see Miller 1966, Ch. 2, for a definition) with parameters k and ν and $\hat{\nu}_{ij}$ is given by (2.5). We shall refer to this procedure as the GH procedure. The use of the studentized range statistic in the GH procedure does not seem to have been adequately justified; in fact it turns out in the MC studies that in some instances GH procedure yields familywise Type I error rate greater than the specified level α, that is, it is radical.

In Tamhane (1977) we proposed two procedures, the first of which is based on Banerjee's (1961) conservative solution to the Behrens-Fisher problem and Šidák's (1967) multiplicative inequality. According to this procedure conservative $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ are given by

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm (t_{\nu_i,\gamma}^2 s_i^2/n_i + t_{\nu_j,\gamma}^2 s_j^2/n_j)^{\frac{1}{2}}] ,$$

where $\gamma = \frac{1}{2}\{1 - (1-\alpha)^{1/k'}\}$. We shall refer to this procedure as the T1 procedure. The second procedure was based on the Welch solution and the Šidák inequality. According to the second procedure approximate-conservative $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ are given by

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm t_{\hat{\nu}_{ij},\gamma}(s_i^2/n_i + s_j^2/n_j)^{\frac{1}{2}}] ,$$

where $\hat{\nu}_{ij}$ is given by (2.5). We shall refer to this latter procedure as the T2 procedure. In the MC studies carried out in Tamhane (1977) it was demonstrated that T1 is highly conservative relative to T2 and hence we shall drop T1 from further consideration. It is reviewed here for the sake of completeness.

Finally, we describe Brown and Forsythe's (1974) procedure. According to this procedure approximate $100(1-\alpha)\%$ joint CI's for all contrasts $\sum_{i=1}^k c_i\mu_i$, where $\sum_{i=1}^k c_i = 0$, are given by

$$\sum_{i=1}^{k} c_i\mu_i \in \left[\sum_{i=1}^{k} c_i\bar{x}_i \pm \{(k-1)F_\alpha(k-1, \hat{\nu}_c)\}^{\frac{1}{2}} \left\{\sum_{i=1}^{k} \frac{c_i^2 s_i^2}{n_i}\right\}^{\frac{1}{2}}\right],$$

where $\hat{\nu}_c$ is the Welch df given by

$$\hat{\nu}_c = \left(\sum_{i=1}^{k} \frac{c_i^2 s_i^2}{n_i}\right)^2 \left\{\sum_{i=1}^{k} \frac{c_i^4 s_i^4}{n_i^2(n_i - 1)}\right\}^{-1} \cdot \quad (2.6)$$

We shall refer to this procedure as the BF procedure. Note that BF gives the following approximate-conservative $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$:

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm \{(k-1)F_\alpha(k-1, \hat{\nu}_{ij})\}^{\frac{1}{2}}(s_i^2/n_i + s_j^2/n_j)^{\frac{1}{2}}] ,$$

where $\hat{\nu}_{ij}$ is given by (2.5). Closely related procedures have been described in Naik (1967).

2.2 Procedures for Equal σ_i^2

Spjøtvoll and Stoline (1973) extended Tukey's multiple comparison procedure to take into account the case of unequal n_i's. According to their procedure the exact (if the homogeneous variances assumption holds) $100(1-\alpha)\%$ joint CI's for all linear combinations $\sum_{i=1}^k a_i\mu_i$ are given by

$$\sum_{i=1}^{k} a_i\mu_i \in [\sum_{i=1}^{k} a_i\bar{x}_i \pm q_{k,\nu,\alpha}' sM(b_1, \ldots, b_k)] ,$$

where $q_{k,\nu,\alpha}'$ denotes the upper α point of the studentized augmented range distribution with parameters k and ν (see Miller 1966, Ch. 2, for a definition); s^2 is the usual "within" estimate of the common variance with $\nu = \sum_{i=1}^k n_i - k$ df; $b_i = a_i/\sqrt{n_i}$ and M is as defined in (2.3). If **a** is a contrast vector and the n_i are equal, then $q_{k,\nu,\alpha}'$ can be replaced by $q_{k,\nu,\alpha}$ (see Theorem 2.3 of

Hochberg 1975), thus yielding Tukey's original procedure. For this reason we shall refer to this procedure as the TSS procedure.

Hochberg (1975) proposed the following procedure for the case of unequal n_i's that he referred to as the GT2 procedure; we shall use the same nomenclature because of its widespread familiarity. According to GT2, the conservative (if the homogeneous variances assumption holds) $100(1-\alpha)\%$ joint CI's for all pairwise differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$ are given by

$$\mu_i - \mu_j \in [\bar{x}_i - \bar{x}_j \pm |m|_{k',\nu,\alpha} s(1/n_i + 1/n_j)^{\frac{1}{2}}] \ ,$$

where $|m|_{k',\nu,\alpha}$ denotes the upper α point of the studentized maximum modulus distribution with parameters k' and ν (see Miller 1966, Ch. 2, for a definition). One can extend the aforementioned pairwise intervals to all contrasts using (2.4).

3. MODIFICATIONS OF SOME PROCEDURES AND COMPARISONS

3.1 A Modification of the *D* Procedure

We note that for solving (2.1) one must use a tedious trial-and-error method with a hand calculator or use a digital computer, except possibly for the case $n_1 = \ldots = n_k = n$ (say). For this case we have $d_{\infty,\alpha} = t_{n-1,\delta}$ with $\delta = \frac{1}{2}\{1 - (1-\alpha)^{1/k}\}$, which can be obtained by interpolating in the t tables. This suggests the following modification of the D procedure that can be implemented with the help of the t tables alone even for the case of unequal n_i's. According to the modified procedure the exact $100(1-\alpha)\%$ joint CI's for all linear combinations $\sum_{i=1}^k a_i\mu_i$ are given by

$$\sum_{i=1}^{k} a_i\mu_i \in [\sum_{i=1}^{k} a_i\bar{x}_i \pm \sum_{i=1}^{k} t_{\nu_i,\delta}|a_i| s_i/\sqrt{n_i}] \ .$$

We shall refer to this modified procedure as the D' procedure. It has the following advantages over D:

1. D' is easier to implement in practice.
2. In D' the CI for a linear combination $\sum_{i\in E} a_i\mu_i$ where $E \subset \{1, 2, \ldots, k\}$ depends only on the ν_i for $i \in E$. In D, however, this CI depends on all the ν_i that is unappealing for obvious reasons.

The performance of D' should be on the average (over $\binom{k}{2}$ pairwise comparisons) equivalent to that of D; in fact if the ν_i are equal then D and D' are identical. In the unequal ν_i case, if all the $\nu_i \to \infty$ or, equivalently, if all the σ_i^2 are known, then $d_{\infty,\alpha}$ and all the $t_{\nu_i,\delta}$ tend to z_δ where z_δ denotes the upper δ point of the standard normal distribution. Thus D and D' are again equivalent. If the ν_i are unequal but small then it might be noted that if $w_{ij,D}$ denotes the width of the CI for $\mu_i - \mu_j$, using D and $w_{ij,D'}$, that using D' then (a.s.) we have

$$w_{ij,D} \le w_{ij,D'} \leftrightarrow d_{\infty,\alpha}(\nu_1, \ldots, \nu_k) \le \frac{(t_{\nu_i,\delta}s_i/\sqrt{n_i} + t_{\nu_j,\delta}s_j/\sqrt{n_j})}{(s_i/\sqrt{n_i} + s_j/\sqrt{n_j})} \ , \quad (3.1)$$

where $d_{\infty,\alpha}$ and $t_{\nu_i,\delta} (i = 1, \ldots, k)$ satisfy

$$1 - \alpha = P\{|t_{\nu_i}| \le d_{\infty,\alpha} \quad i = 1, \ldots, k\} = P\{|t_{\nu_i}| \le t_{\nu_i,\delta} \quad i = 1, \ldots, k\} \ .$$

Therefore, there exists a number $\nu^* \in (\min_i\nu_i, \max_i\nu_i)$ such that if $\nu_i, \nu_j > \nu^*$ then (3.1) is violated; if $\nu_i, \nu_j < \nu^*$ then (3.1) is satisfied; in all other cases (3.1) is satisfied (with high probability) if σ_i^2/n_i is highly different from σ_j^2/n_j. Thus we should expect D to do better than D' for those pairwise comparisons involving highly unbalanced (σ_i^2/n_i)-values. In spite of this one advantage of D over D', we drop D from further consideration because of the advantages in favor of D' cited earlier and because, on the average, performances of D and D' are similar. (The MC results for D are available from the author.)

H2 can be modified in a similar manner. The modified H2 (denoted by H2′) has the same advantages over H2 that D' has over D. For this reason we drop H2 from further consideration. Now note that the modified procedure H2′ is identical to UW and, therefore, H2′ will not be considered separately.

3.2 Modification of (2.5) and Procedures Affected by It

Ury and Wiggins (1971) pointed out that $\hat{\nu}_{ij}$ given by (2.5) ranges between $\min(n_i - 1, n_j - 1)$ and $n_i + n_j - 2$, but usually tends to be on the conservative side. Based on the work of Pratt (1964), they advocated that $\hat{\nu}_{ij}$ be taken to be equal to its upper limit $n_i + n_j - 2$ when at least one of the following four conditions holds:

1. $9/10 \le n_i/n_j \le 10/9$;
2. $9/10 \le (s_i^2/n_i)/(s_j^2/n_j) \le 10/9$;
3. $4/5 \le n_i/n_j \le 5/4$ and $1/2 \le (s_i^2/n_i)/(s_j^2/n_j) \le 2$;
4. $2/3 \le n_i/n_j \le 3/2$ and $3/4 \le (s_i^2/n_i)/(s_j^2/n_j) \le 4/3$.

This modification essentially means that if the sample sizes of the two groups are approximately balanced or the standard errors of the corresponding sample means are approximately balanced then one should use the "usual" df $n_i + n_j - 2$.

This modified value of $\hat{\nu}_{ij}$ can be used in UW, T2, GH, and BF (for pairwise comparisons) procedures; denote the modified procedures by UW′, T2′, GH′, and BF′, respectively. (For general contrasts we shall use (2.6) for the df $\hat{\nu}_c$ for BF without any modification although we shall still refer to that procedure as BF′.) Because UW, T2, and BF (for pairwise comparisons) are approximate-conservative, one can anticipate that this modification will make UW′, T2′, and BF′ sharper without making them radical; this is indeed so and we drop original procedures UW, T2, and BF from further consideration. The same statement cannot be made about GH′, however. In fact, in our MC studies GH′ was tried but turned out to be substantially radical. Therefore, we retained GH, which itself is somewhat radical. For the convenience of the reader we provide a glossary of all the procedures considered so far.

GLOSSARY OF THE PROCEDURES CONSIDERED

Mnemonic	*Procedure*
D, D′	Dalal's (1978) procedure and its modified version, respectively
S	Spjøtvoll's (1972) procedure
H1, H2	Hochberg's (1976) two procedures
UW, UW′	Ury and Wiggins's (1971) procedure and its modified version, respectively
GH, GH′	Games and Howell's (1976) procedure and its modified version, respectively
T1, T2, T2′	Tamhane's (1977) two procedures and the modified version of his latter procedure, respectively
BF, BF′	Brown and Forsythe's (1974) procedure and its modified version, respectively
TSS	Spjøtvoll and Stoline's (1973) procedure
GT2	Hochberg's (1974) procedure

4. SOME FURTHER COMPARISONS

First note that while UW′ is based on the Bonferroni inequality, T2′ is based on the multiplicative Šidák inequality; thus $\beta < \gamma$ and $t_{\nu,\beta} > t_{\nu,\gamma}$. Therefore, T2′ uniformly dominates UW′ in terms of the CI widths and hence we drop UW′ from further consideration.

It is easy to check that $q_{\nu,k,\alpha}/\sqrt{2} \leq t_{\nu,\gamma}$ with equality holding iff $k = 2$. Thus GH uniformly dominates T2. As mentioned earlier, however, GH tends to be radical and hence the comparison is not exactly fair. In any case, we use T2′ and not T2 and it is not clear whether GH uniformly dominates T2′ because the df of the critical points used in the two procedures can now be different.

A direct comparison between D′ and H1 seems difficult but one can compare D and H1 as follows. Comparing the respective associated widths $w_{ij,\mathrm{D}}$ and $w_{ij,\mathrm{H1}}$ of the CI's for $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$, one has (a.s.)

$$w_{ij,\mathrm{D}} \geq w_{ij,\mathrm{H1}} \leftrightarrow \frac{h_\alpha'(\nu_1, \ldots, \nu_k)}{d_{\infty,\alpha}(\nu_1, \ldots, \nu_k)} \geq \frac{(s_i/\sqrt{n_i} + s_j/\sqrt{n_j})}{\max(s_i/\sqrt{n_i}, s_j/\sqrt{n_j})} \cdot \quad (4.1)$$

It is easy to verify that the left side of (4.1) is greater than one and for the $(\nu_1, \ldots, \nu_k)$-combinations and α values studied in this article we verified that it is in the range 1.35 to 1.5 by computing h_α' and $d_{\infty,\alpha}$; the values of h_α' are given in Table 2. The right side of (4.1) is greater than one (a.s.), and it would be close to one with large probability if σ_i^2/n_i and σ_j^2/n_j are highly different. Thus except for comparisons involving highly unequal (σ_i^2/n_i)-values, it is likely that H1 would dominate D (and therefore D′); this is supported by the MC results.

Based on these considerations we finally keep eight procedures in our MC study: D′, S, H1, T2′, GH, BF′, TSS, and GT2.

5. MONTE CARLO STUDY

5.1 Choice of (σ^2, n)-Configurations and Other Parameters

The sampling experiments were conducted for all pairwise differences of the μ_i for $k = 4$ and 8 and for selected contrasts for $k = 8$. The values of α used were .05 and .10 although here we report the results only for $\alpha = .05$; the patterns in the results for $\alpha = .10$ are quite similar. For each combination of k and α, eight (σ^2, n)-configurations were studied. Our concern is mainly with the small-sample behavior of the procedures, and this guided our choice of the n_i's in the range 7 to 13. These configurations were ordered in terms of a measure of unbalance in the values of

$$\mathrm{var}(\bar{x}_i) = \tau_i = \sigma_i^2/n_i (i = 1, \ldots, k) \ .$$

This measure $\varphi(\boldsymbol{\tau})$ is given by

$$\varphi(\boldsymbol{\tau}) = \{\sum_{i=1}^{k} (\tau_i - \bar{\tau})^2/k\}^{\frac{1}{2}}/\bar{\tau} \ ,$$

where $\bar{\tau} = \sum_{i=1}^k \tau_i/k$. The same measure was used by Keselman and Rogan (1978) although they used it for measuring unbalance of σ_i^2's; we believe that τ is a more relevant parameter in the present problem than σ^2. The configurations in their "natural" order (primed serial numbers) are listed in Table 1. The φ values associated with the configurations and their order according to the φ values (unprimed serial numbers) are listed in the same table. Thus for both $k = 4$, 8, configuration 1′ with $\varphi(\boldsymbol{\tau}) = 0$ is the most balanced while configuration 8′ with $\varphi(\boldsymbol{\tau}) = 1.479$ is the most unbalanced. An alternative measure of unbalance that can be used is

$$\psi(\boldsymbol{\tau}) = \max_{1 \leq i \leq k} \tau_i / \min_{1 \leq i \leq k} \tau_i \ .$$

Note that ψ gives the same ordering as φ except that ranks of configurations 3′ and 6′ are interchanged, that is they are 6 and 7 according to φ while 7 and 6 according to ψ. The MC results are presented in terms of the unprimed serial numbers of the configurations, thus enabling the reader to see how the different procedures react to increasing unbalance in the τ_i's.

The class of general contrasts of practical interest typically involves comparing the average of one subset of the μ_i against the average of another subset of the μ_i. We selected three representative contrasts from this class for study in the MC experiments. Thus for $k = 8$, we selected one high-order contrast (4 :4 comparison) and two middle-order contrasts (2 :2 comparisons). They are

$$\begin{aligned} &\text{Contrast 1: } \tfrac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) - \tfrac{1}{4}(\mu_5 + \mu_6 + \mu_7 + \mu_8) \ ; \\ &\text{Contrast 2: } \tfrac{1}{2}(\mu_1 + \mu_2) - \tfrac{1}{2}(\mu_7 + \mu_8) \ ; \\ &\text{Contrast 3: } \tfrac{1}{2}(\mu_3 + \mu_4) - \tfrac{1}{2}(\mu_5 + \mu_6) \ . \end{aligned} \quad (5.1)$$

The pairwise comparisons are, of course, the lowest-order contrasts.

1. (σ^2, n)-Configurations

k	Configuration No.	$\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$	$n_1, n_2, \ldots, n_k$	$\varphi(\tau)$	Configuration No. in Terms of $\varphi(\tau)$
	1′	1,1,1,1	7,7,7,7	0	1
	2′	1,2,3,4	7,7,7,7	.447	4
	3′	1,4,7,10	7,7,7,7	.610	6
4	4′	1,1,1,1	7,9,11,13	.235	2
	5′	1,2,3,4	7,9,11,13	.262	3
	6′	1,2,3,4	13,11,9,7	.639	7
	7′	1,4,7,10	7,9,11,13	.500	5
	8′	1,4,7,10	13,11,9,7	1.479	8
	1′	1,1,1,1,1,1,1,1	7,7,7,7,7,7,7,7	0	1
	2′	1,1,2,2,3,3,4,4	7,7,7,7,7,7,7,7	.447	4
	3′	1,1,4,4,7,7,10,10	7,7,7,7,7,7,7,7	.610	6
8	4′	1,1,1,1,1,1,1,1	7,7,7,7,7,7,7,7	.235	2
	5′	1,1,2,2,3,3,4,4,	7,7,9,9,11,11,13,13	.262	3
	6′	1,1,2,2,3,3,4,4	13,13,11,11,9,9,7,7	.639	7
	7′	1,1,4,4,7,7,10,10	7,7,9,9,11,11,13,13	.550	5
	8′	1,1,4,4,7,7,10,10	13,13,11,11,9,9,7,7	1.479	8

5.2 Critical Points for the Various Procedures

In this section we describe how the critical points for the various procedures included in the MC study were obtained. The actual details of the MC study are given in the following section.

Unusual t values needed for D′ and T2′ were obtained by using the IMSL subroutine MDSTI. Similarly, the critical point $d_{2,\alpha}$ for S and the F values for BF′ were obtained by using the IMSL subroutine MDFI. Note that both MDSTI and MDFI give exact results even in the case of fractional df. The q and q' values needed for GH and TSS were obtained by interpolation in the tables given by Harter (1960) and Stoline (1978), respectively. The $|m|$ values needed for GT2 were obtained by interpolation in the tables given by Stoline and Ury (1978). Linear harmonic interpolation with respect to the df was used in all three cases.

Tables of values of h_α' needed for H1 are not available in the literature. To compute h_α' we develop an integral expression for the distribution function of R' that seems to be new and hence is given here. Represent $R' = \max_{0\le i<j\le k}|T_i - T_j|$ where $T_0 \equiv 0$, T_i is distributed as a t_{ν_i} rv (denoted by $T_i \sim t_{\nu_i}$), and $T_1, \ldots, T_k$ are independent. To obtain $h_\alpha' = h_\alpha'(\nu_1, \ldots, \nu_k)$ one solves the following equation in h':

$$\begin{aligned}1 - \alpha &= P\{\max_{0\le i\le k}(T_i) - T_j \le h' \quad \text{for all } j\} \\ &= \sum_{i=0}^{k} P\{T_i > T_j \ge T_i - h' \quad \text{for all } j \ne 1\} \\ &= \prod_{j=1}^{k}\{F_{\nu_j}(0) - F_{\nu_j}(-h')\} + \sum_{i=1}^{k}\int_0^{h'}\prod_{\substack{j=1\\ j\ne i}}^{k}\{F_{\nu_j}(t) \\ &\qquad - F_{\nu_j}(t-h')\}dF_{\nu_i}(t) \\ &= \prod_{j=1}^{k}\{F_{\nu_j}(h') - \tfrac{1}{2}\} + \int_0^{h'}[\sum_{i=1}^{k}\prod_{\substack{j=1\\ j\ne i}}^{k}\{F_{\nu_j}(t) \\ &\qquad - F_{\nu_j}(t-h')\}f_{\nu_i}(t)]dt\ , \qquad (5.2)\end{aligned}$$

where $f_\nu(\cdot)$ denotes the density function of a t_ν rv. The IMSL (1978) subroutine MDTD was used to evaluate $F_\nu(\cdot)$ and ZSYSTM was used to solve (5.2); Romberg quadrature method was used to evaluate the integral.

The values of $q_{k,\nu,\alpha}'$ ($q_{k,\nu,\alpha}$ if the n_i are equal), $|m|_{k',\nu,\alpha}$, $d_{2,\alpha}(\nu_1, \ldots, \nu_k)$, and $h_\alpha'(\nu_1, \ldots, \nu_k)$ used in the MC study are given in Table 2.

2. Critical Points for the Various Procedures ($1 - \alpha = .95$)

k	$\nu_1, \nu_2, \ldots, \nu_k$	ν	h_α'	$d_{2,\alpha}$	$q_{k,\nu,\alpha}$	$q_{k,\nu,\alpha}'$	$\lvert m\rvert_{k,\nu,\alpha}$
4	6,6,6,6	24	4.747	4.156	3.901	—	2.851
	6,8,10,12	36	4.374	3.812	—	3.810	2.775
8	6,6,6,6,6,6,6,6	48	5.903	5.396	4.481	—	3.286
	6,6,8,8,10,10,12,12	72	5.359	4.906	—	4.415	3.228

5.3 Details of the MC Study

For each combination of k and the (σ^2, n)-configuration 1,000 experiments were run. For studying the CI's we can take all the μ_i equal to zero without loss of generality. Therefore, in each experiment, k independent pairs of rv's $\bar{x}_i \sim N(0, \sigma_i^2/n_i)$ and $s_i^2 \sim \sigma_i^2\chi_{\nu_i}^2/\nu_i$ were generated. For generating normal rv's, the Box-Müller algorithm was used; the chi-squared rv's were generated by using the relation (for ν even) $\chi_\nu^2 \sim -\sum_{i=1}^{\nu/2}\log_e U_i$ where the U_i are independent uniform [0, 1] rv's. The FORTRAN library function RANF was used to generate the uniform rv's.

We carried out separate runs for pairwise comparisons and contrasts. For pairwise comparisons, for each procedure we obtained the estimates of (a) the achieved joint confidence level, (b) the expected half-widths of the CI's for all $\binom{k}{2}$ pairwise differences, and (c) the average of the $\binom{k}{2}$ expected half-widths. For a given procedure, the estimate of the joint confidence level was obtained by keeping a count of the number of runs in which zero

3. Estimated Confidence Levels for All Pairwise Comparisons ($1 - \alpha = 0.95$)[a]

k	Procedure	1	2	3	4	5	6	7	8
		(σ^2, n)-Configuration No.							
4	D′	.999	.998	.995	.998	.996	.993	.998	.988
	S	.994	.993	.990	.993	.991	.987	.992	.982
	H1	.985	.984	.975	.983	.989	.979	.986	.975
	T2′	.955	.956	.953	.954	.963	.947	.968	.948
	GH	.946	.944	.941*	.942	.950	.940	.957	.936*
	BF′	.957	.964	.964	.962	.973	.953	.975	.955
	TSS	.950	.960	.968	.938*	.970	.930*	.945	.909*
	GT2	.951	.955	.963	.950	.967	.940*	.928*	.891*
8	D′	.999	.996	.998	.992	.992	.993	.998	.993
	S	.997	.995	.997	.992	.993	.993	.996	.994
	H1	.992	.980	.985	.981	.984	.983	.986	.984
	T2′	.966	.963	.965	.942	.966	.963	.966	.973
	GH	.941*	.943	.935*	.916*	.948	.949	.946	.949
	BF′	.990	.985	.990	.975	.987	.977	.990	.988
	TSS	.960	.959	.962	.927*	.965	.914*	.922*	.897*
	GT2	.965	.960	.962	.942	.962	.938*	.916*	.882*

[a] Asterisk indicates that the achieved confidence level is less than the designated confidence level = .95 at 10 percent significance. The standard error of any entry P is given by $(P(1 - P)/1{,}000)^{1/2}$.

was included in all $\binom{k}{2}$ CI's for the differences $\mu_i - \mu_j (i, j = 1, \ldots, k; i < j)$. The results regarding estimated confidence levels are presented in Table 3. For lack of space we have not reported here the estimates of the expected half-widths for the individual $\binom{k}{2}$ comparisons, but in Table 4 we have given the average (over $\binom{k}{2}$ CI's) expected half-widths of the CI's for pairwise differences produced by each procedure. The half-widths of the CI's produced by each procedure for selected contrasts are given in Table 5. Note that the results for TSS and GT2 are not included, in an effort to keep the table small and also because sufficient information about the relative performance of these two procedures is obtained from the pairwise comparisons data. The standard errors of the estimates are given as parenthetical entries in the same tables. In arranging the tables we have put the procedures of similar type together so that their performances can be more readily compared; thus we have put D′, S, and H1 together because they possess the common property of having resolutions for all linear combinations. The discussion of the results is given in the following section.

6. DISCUSSION AND RECOMMENDATIONS

Let us study Table 3 first. To identify the liberal procedures we have marked with an asterisk those confidence levels that fall in the critical region for the one-sided hypothesis-testing problem H_0: $1 - \alpha = .95$ vs. H_1: $1 - \alpha < .95$ at 10 percent level of significance (i.e., confidence levels $< .942$). Thus we note that TSS and GT2 tend to be liberal for configurations 6, 7, and 8; for other configurations they seem to control the designated confidence level fairly well. One can thus conclude that for pairwise comparisons TSS and GT2 are fairly robust unless the unbalance in the τ_i values is extreme. Keep in mind that both TSS and GT2 are conservative for the problem of pairwise comparisons (if the homogeneous variances assumption holds) except when the n_i are equal when TSS is exact. Thus the apparent robust behavior of these procedures might be due to their inherent conservative nature.

Among the procedures designed for unequal variances we find that only GH tends to be liberal but there does not seem to be any specific pattern of configurations for which GH is liberal. This liberal nature of GH was noted in Games and Howell (1976), although in the MC study done by Keselman and Rogan (1978) GH is shown to control the confidence level more precisely. We also find that BF′, H1, S, and D′ are increasingly more conservative. This should not come as a surprise because BF′ is designed for all contrasts, while the other three procedures are designed for all linear combinations. T2′ controls the confidence level fairly well. One can compare the performance of T2′ with that of T2 (see Table 1 of Tamhane 1977; T2 is referred to as the W procedure there) and note the reduction in overprotection due to the use of the Ury and Wiggins df modification.

Next turn to Table 4. Among the procedures designed for unequal variances we note that GH produces the shortest CI's for all the eight configurations while T2′ and BF′ come next; H1, S, and D′ produce increasingly wider CI's. Although it seems that GH gives the best performance for pairwise comparisons, one must take into account the fact that GH is also somewhat liberal. If one does not want to run the risk of frequently liberal CI's produced by GH, then T2′ seems to offer the best choice for pairwise comparisons because it controls the confidence level more precisely and produces the CI's that are only slightly wider.

4. Average Estimated CI *Half-Widths for Pairwise Comparisons* *($1 - \alpha = 0.95$)*

k	*Procedure*	(σ^2, n)-Configuration No. 1	2	3	4	5	6	7	8
4	D′	2.541 (.017)[a]	1.987 (.011)	2.930 (.016)	3.924 (.027)	4.109 (.022)	5.532 (.039)	3.212 (.021)	4.633 (.032)
	S	2.173 (.014)	1.713 (.009)	2.576 (.014)	3.418 (.024)	3.720 (.020)	4.930 (.037)	2.809 (.019)	4.129 (.031)
	H1	2.000 (.014)	1.580 (.010)	2.370 (.015)	3.285 (.027)	3.594 (.022)	4.901 (.042)	2.774 (.023)	4.201 (.034)
	T2′	1.643 (.010)	1.376 (.008)	2.022 (.011)	2.583 (.018)	2.930 (.016)	3.727 (.028)	2.368 (.018)	3.560 (.030)
	GH	1.586 (.010)	1.309 (.007)	1.933 (.010)	2.535 (.018)	2.797 (.016)	3.722 (.030)	2.227 (.017)	3.330 (.027)
	BF′	1.692 (.011)	1.424 (.008)	2.103 (.011)	2.661 (.018)	3.045 (.017)	3.839 (.029)	2.426 (.018)	3.628 (.030)
	TSS	1.457 (.007)	1.323 (.005)	2.205 (.009)	2.335 (.012)	3.339 (.014)	3.371 (.018)	1.983 (.008)	2.840 (.013)
	GT2	1.506 (.007)	1.260 (.005)	2.100 (.009)	2.393 (.012)	3.181 (.014)	3.485 (.018)	1.889 (.008)	2.705 (.013)
8	D′	2.976 (.020)	2.284 (.014)	3.341 (.018)	4.534 (.036)	4.672 (.025)	6.524 (.046)	3.708 (.024)	5.359 (.037)
	S	2.830 (.018)	2.207 (.013)	3.305 (.018)	4.374 (.030)	4.750 (.026)	6.423 (.048)	3.612 (.024)	5.297 (.038)
	H1	2.498 (.018)	1.934 (.013)	2.910 (.017)	3.994 (.037)	4.325 (.027)	6.050 (.051)	3.362 (.027)	5.070 (.043)
	T2′	2.086 (.014)	1.721 (.010)	2.500 (.014)	3.225 (.022)	3.607 (.020)	4.736 (.035)	2.993 (.024)	4.526 (.039)
	GH	1.947 (.013)	1.590 (.009)	2.332 (.013)	3.057 (.022)	3.357 (.018)	4.572 (.034)	2.705 (.022)	4.044 (.033)
	BF′	2.367 (.015)	1.973 (.011)	2.894 (.016)	3.658 (.024)	4.169 (.023)	5.372 (.038)	3.365 (.027)	5.035 (.041)
	TSS	1.687 (.005)	1.522 (.004)	2.530 (.006)	2.641 (.009)	3.817 (.011)	3.932 (.015)	2.291 (.007)	3.281 (.010)
	GT2	1.750 (.006)	1.471 (.004)	2.445 (.007)	2.739 (.009)	3.689 (.011)	4.077 (.016)	2.215 (.007)	3.171 (.010)

[a] The entries in parentheses are the standard errors of the corresponding estimates.

It is not the purpose of this article to carry out an extensive comparison between TSS and GT2. But we note in passing that for all the configurations involving unequal n_i's (i.e., configuration numbers 2, 3, 5, 7, and 8) GT2 produces shorter CI's than TSS, while for the other cases, as one would expect, TSS produces shorter CI's.

By comparing the CI widths of GH and T2′ with that of TSS and GT2 for configuration numbers 1 and 2, respectively (i.e., the configurations for which the variances are homogeneous), one gets an idea of the loss in efficiency if good procedures for pairwise comparisons, in the case of unequal variances such as GH and T2′, are used when the underlying variances are equal. It would seem that the loss is not substantial, although it must be kept in mind that this conclusion only pertains when the sample sizes are not too different.

It should be pointed out that all the previous comparisons pertain to average CI widths. If the widths of the individual CI's for $\binom{k}{2}$ pairwise differences are compared, there are few instances in which even D′ beats T2′. Typically this occurs when the two τ_i values are highly different, namely, for the CI for $\mu_1 - \mu_k (k = 4, 8)$ for configuration 8 where $\tau_1 = 1/13$ and $\tau_k = 10/7$. It is not possible to give such detailed comparisons here because of lack of space.

Finally, we turn to Table 5. In this table we have listed the estimated half-widths of the CI's for selected contrasts for all the procedures designed for unequal variances (i.e., excluding TSS and GT2). Here we find that BF′ gives the best performance in all the cases. For the second-best performance, the contenders are S and GH. S seems to perform better than GH for the high-order contrast 1 and, in the case of middle-order contrasts 2 and 3, for configurations 7 and 8 corresponding to unbalanced τ_i values while GH performs better in the other cases. As in the case of pairwise comparisons, T2′ produces CI's that are only slightly wider than those produced by GH. H1 and D′ produce very wide CI's, H1 performing better than D′.

Perhaps a very surprising result (at least to us) of the MC study for contrasts was the performance of S. Although we anticipated that BF′ would perform very well for contrasts, we did not quite anticipate that S would be the second best. The poor performance of D′ and H1 even

5. *Estimated CI Half-Widths for Contrasts*
($1 - \alpha = 0.95$)

Contrast No.	Procedure	(σ^2, n)-Configuration 1	2	3	4	5	6	7	8
1	D′	2.975 (.009)[a]	2.285 (.007)	3.351 (.009)	4.586 (.015)	4.658 (.012)	6.527 (.022)	3.686 (.012)	5.361 (.019)
	S	1.436 (.004)	1.120 (.003)	1.681 (.004)	2.281 (.008)	2.431 (.006)	3.349 (.012)	1.868 (.007)	2.797 (.011)
	H1	2.316 (.008)	1.858 (.007)	2.828 (.008)	4.020 (.018)	4.377 (.014)	6.210 (.028)	3.469 (.016)	5.350 (.024)
	T2′	2.085 (.006)	1.740 (.006)	2.511 (.006)	3.292 (.012)	3.648 (.010)	4.820 (.017)	3.130 (.014)	4.841 (.022)
	GH	1.945 (.006)	1.599 (.005)	2.339 (.006)	3.149 (.012)	3.387 (.009)	4.725 (.019)	2.788 (.011)	4.252 (.018)
	BF′	1.022 (.003)	0.867 (.003)	1.286 (.003)	1.658 (.006)	1.863 (.005)	2.477 (.010)	1.499 (.006)	2.284 (.011)
2	D′	2.969 (.013)	2.383 (.010)	3.272 (.012)	4.489 (.022)	4.292 (.016)	6.136 (.031)	3.891 (.020)	5.578 (.031)
	S	2.016 (.009)	1.612 (.007)	2.310 (.008)	3.211 (.017)	3.270 (.013)	4.668 (.027)	2.771 (.016)	4.173 (.026)
	H1	2.383 (.011)	1.979 (.011)	2.918 (.012)	4.325 (.028)	4.581 (.021)	6.721 (.043)	3.928 (.026)	6.126 (.040)
	T2′	2.081 (.009)	1.842 (.010)	2.449 (.009)	3.300 (.018)	3.516 (.015)	4.787 (.027)	3.663 (.025)	5.852 (.041)
	GH	1.941 (.008)	1.671 (.008)	2.280 (.008)	3.230 (.019)	3.256 (.013)	4.901 (.031)	3.146 (.020)	4.907 (.033)
	BF′	1.571 (.007)	1.365 (.007)	1.847 (.007)	2.590 (.015)	2.662 (.011)	3.872 (.024)	2.501 (.015)	3.866 (.025)
3	D′	2.980 (.013)	2.187 (.008)	3.429 (.013)	4.684 (.022)	5.023 (.018)	6.917 (.031)	3.481 (.014)	5.143 (.021)
	S	2.025 (.009)	1.544 (.006)	2.428 (.009)	3.199 (.015)	3.574 (.013)	4.743 (.021)	2.473 (.010)	3.670 (.015)
	H1	2.401 (.011)	1.820 (.008)	2.855 (.012)	3.847 (.021)	4.270 (.018)	5.821 (.032)	3.050 (.017)	4.600 (.025)
	T2′	2.090 (.009)	1.642 (.007)	2.568 (.009)	3.299 (.016)	3.785 (.013)	4.892 (.022)	2.683 (.012)	4.008 (.018)
	GH	1.951 (.009)	1.532 (.006)	2.398 (.009)	3.094 (.015)	3.530 (.012)	4.609 (.022)	2.487 (.011)	3.705 (.016)
	BF′	1.578 (.007)	1.245 (.005)	1.942 (.007)	2.502 (.012)	2.862 (.010)	3.721 (.017)	2.031 (.009)	3.033 (.014)

[a] The entries in parentheses are the standard errors of the corresponding estimates.

for contrasts would seem to rule out their use in most practical applications. It should be mentioned that H1 is less conservative for pairwise comparisons than S, while D′ always seems to be the most conservative.

On the whole, we would recommend GH and T2′ for pairwise comparisons, GH giving slightly narrower CI's than T2′ at the risk of not attaining the designated confidence level by a small amount in some cases. For general contrast comparisons we recommend the BF′ procedure.

[*Received October 1977. Revised December 1978.*]

REFERENCES

Banerjee, Saibal K. (1961), "On Confidence Interval for Two Means Problem Based on Separate Estimates of Variances and Tabulated Values of *t*-Variable," *Sankhyā*, Ser. A, 23, 359–378.

Brown, Morton B., and Forsythe, Alan B. (1974), "The ANOVA and Multiple Comparisons for Data With Heterogeneous Variances," *Biometrics*, 30, 719–724.

Dalal, Siddharth R. (1978), "Simultaneous Confidence Procedures for Univariate and Multivariate Behrens-Fisher Type Problems," *Biometrika*, 65, 221–224.

Gabriel, K. Ruben (1969), "Simultaneous Test Procedures—Some Theory of Multiple Comparisons," *Annals of Mathematical Statistics*, 40, 224–250.

Games, Paul A., and Howell, John F. (1986), "Pairwise Multiple Comparison Procedures With Unequal *N*'s and/or Variances: A Monte Carlo Study," *Journal of Educational Statistics*, 1, 113–125.

Harter, H. Leon (1960), "Tables of Range and Studentized Range," *Annals of Mathematical Statistics*, 31, 1122–1147.

Hochberg, Yosef (1974), "Some Generalizations of the *T*-Method in Simultaneous Inference," *Journal of Multivariate Analysis*, 4, 224–234.

——— (1975), "An Extension of the *T*-Method to General Unbalanced Models of Fixed Effects," *Journal of the Royal Statistical Society*, Ser. B, 37, 426–433.

——— (1976), "A Modification of the *T*-Method of Multiple Comparisons for a One-Way Layout With Unequal Variances," *Journal of the American Statistical Association*, 71, 200–203.

International Mathematical and Statistical Libraries (1978), IMSL Manual (Vols. I and II), Houston, Tex.: IMSL.

Keselman, H.J., and Rogan, Joanne C. (1978), "A Comparison of the Modified Tukey and Scheffé Methods of Multiple Comparisons for Pairwise Contrasts," *Journal of the American Statistical Association*, 73, 47–52.

Miller, Rupert G. (1966), *Simultaneous Statistical Inference*, New York: McGraw-Hill Book Co.

Naik, Umesh D. (1967), "Simultaneous Confidence Bounds Concerning Means in the Case of Unequal Variances," *The Karnatak University Journal: Science*, XII, 93–104.

Pratt, John W. (1964), "Robustness of Some Procedures for the Two-Sample Location Problem," *Journal of the American Statistical Association*, 59, 665–680.

Šidák, Zbynek (1967), "Rectangular Confidence Regions for Means of Multivariate Normal Distributions," *Journal of the American Statistical Association*, 62, 626–633.

Spjøtvoll, Emil (1972), "Joint Confidence Intervals for All Linear Functions of Means in ANOVA With Unknown Variances," *Biometrika*, 59, 684–685.

———, and Stoline, Michael R. (1973), "An Extension of the *T*-Method of Multiple Comparisons to Include the Cases With Unequal Sample Sizes," *Journal of the American Statistical Association*, 68, 975–978.

Stoline, Michael R. (1978), "Tables of the Studentized Augmented Range and Applications to Problems of Multiple Comparison," *Journal of the American Statistical Association*, 73, 656–660.

———, and Ury, Hans K. (1979), "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons Among Means," *Technometrics*, to appear.

Tamhane, Ajit C. (1977), "Multiple Comparisons in Model I One-Way ANOVA With Unequal Variances," *Communications in Statistics*, A6(1), 15–32.

Ury, Hans K. (1976), "A Comparison of Four Procedures for Multiple Comparisons Among Means (Pairwise Contrasts) for Arbitrary Sample Sizes," *Technometrics*, 18, 89–97.

———, and Wiggins, Alvin, D. (1971), "Large Sample and Other Multiple Comparisons Among Means," *British Journal of Mathematical and Statistical Psychology*, 24, 174–194.

Wang, Y.Y. (1971), "Probabilities of Type I Error of Welch Tests for the Behrens-Fisher Problem," *Journal of the American Statistical Association*, 66, 605–608.