DOCUMENT RESUME

ED 357 072 TM 019 851

AUTHOR	Hsiung, Tung-Hsing; Olejnik, Stephen
TITLE	Contrast Analyses for the Additive Nonorthogonal Two-Factor Design in the Unequal Variance Case.
PUB DATE	Apr 93
NOTE	38p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Analysis of Variance; *Comparative Analysis; Computer Simulation; Equations (Mathematics); *Factor Analysis; *Mathematical Models; *Sample Size
IDENTIFIERS	*Contrast Analysis; Dunnett Test; Multiple Comparisons; *Nonorthogonal Variables; Paired Comparisons; Power (Statistics); Type I Errors; Variance (Statistical)

ABSTRACT

This study considers the problem of performing all pairwise comparisons of column means for a two-by-four additive nonorthogonal factorial analysis of variance (ANOVA) model where cell variances are heterogeneous. Extensions of the following procedures are considered: (1) Games-Howell (1976) procedure; (2) the C. W. Dunnett (1980) T3 and C procedures; (3) the B. S. Holland and M. D. Copenhaver (1987) technique; (4) the A. J. Hayter (1986) procedure; and (5) the James (1951) second-order test. Using computer simulated data, Type I error rates and statistical power for these multiple comparison procedures are estimated. Sixty-six combinations of sample size and variance patterns are examined. The results suggest that these procedures maintain the familywise Type I error rate under the nominal 0.05 level. In terms of statistical power, the Games-Howell procedure generally provides the greater any-pair power, but the extension of the Hayter technique provides greater average power per contrast as well as in identifying all significant pairwise differences. Fourteen tables present details from the analyses. (Author/SLD)

CONTRAST ANALYSES FOR THE ADDITIVE NONORTHOGONAL TWO-FACTOR DESIGN IN THE UNEQUAL VARIANCE CASE

Tung-Hsing Hsiung

Stephen Olejnik

University of Georgia

Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA. April, 1993.

Abstract

This study considers the problem of performing all pairwise comparisons of column means for a two by four additive nonorthogonal factorial ANOVA model where cell variances are heterogeneous. Extensions of the Games-Howell (1976) procedure, the Dunnett (1980) T3 and C procedures, the Holland and Copenhaver (1987) technique, the Hayter (1986) procedure, and the James (1951) second-order test are considered. Using computer simulated data, Type I error rates and statistical power for these multiple comparison procedures are estimated. Sixty-six different combinations sample size and variance patterns are examined. The results suggest that these procedures maintain the familywise Type I error rate under the nominal .05 level. In terms of statistical power, the Games-Howell procedure generally provides the greater any-pair power but the extension of the Hayter technique provides greater average power per contrast as well as in identifying all significant pairwise differences.

Over the last twenty years, considerable attention has focused on the performance of the multiple comparison procedures (MCP) among group means when variances are homogeneous or heterogeneous (e.g., Dunnett, 1980; Hsiung & Olejnik, 1991; Keselman & Rogan, 1978; Klockars & Hancock, 1992; Seaman, Levin, & Serlin, 1991). Each of these studies has dealt only with the performance of MCP in the single-factor design. As yet, only Wilcox (1987) has considered a two-factor model but limited his investigation to pairwise comparisons of cell means in each row or column. We have not found any studies in the literature which have investigated the properties of MCP to contrasts among marginal means in a factorial design when population variances are heterogeneous.

In a nonorthogonal two factor design a pairwise contrast between column means using an unweighted means solution can be formed by summing the cell means within a column and subtracting the sum of cell means from the contrasting column. That is, $\hat{\psi}_{kk'} = \Sigma_j\bar{X}_{jk} - \Sigma_j\bar{X}_{jk'}$ ($k \neq k'$, and k, $k' = 1, \ldots, K$; j represents the j_{th} row and $j = 1, \ldots, J$). To test the null hypothesis: H_0: $\hat{\psi}_{kk'} = 0$, a Welch (1938) type test statistic can be formed by dividing the point estimate for the contrast by the standard error of the contrast. Where the standard error is the square root of the variance of the contrast formed by summing the separate cell variances:

$$\sqrt{\mathrm{Var}(\psi)} = \sqrt{(\Sigma_j S^2_{jk}/n_{jk}) + (\Sigma_j S^2_{jk'}/n_{jk'})}.$$

The test statistic can be written as:

$$T = \frac{\sum_{j=1}^{J}\overline{X_{jk}} - \sum_{j=1}^{J}\overline{X_{jk'}}}{\sqrt{\sum_{j=1}^{J}\frac{S^2{}_{jk}}{n_{jk}} + \sum_{j=1}^{J}\frac{S^2{}_{jk'}}{n_{jk'}}}} \quad . \tag{1}$$

More generally, any hypothesis involving the linear combination of cell means can be tested using the following statistic:

$$\Sigma c_{jk}\bar{X}_{jk} \;/\; \sqrt{\Sigma(c_{jk}S^2_{jk}/n_{jk})}.$$

Where c_{jk} are the contrast coefficients and $\Sigma c_{jk} = 0$ [$(j,k) = (1,1), \ldots, (J,K)$];

$\bar{X}_{jk}$ are cell means for the j_{th} level of factor J and the k_{th} level of factor K;

S^2_{jk} is the variance of cell jk;

and n_{jk} is the number of observations in cell jk.

The T test statistic may be evaluated using several different distributions. In the present study we consider extensions of the Dunnett (1980) C and T3 procedures, the Games and Howell (1976) procedure, the Hayter (1986) modified LSD procedure, the Holland and Copenhaver (1987) adjusted Bonferroni technique, and a Scheffé type procedure using the James (1951) second-order test. The purpose of our study is to investigate whether these extensions provide valid tests, that is, control the familywise Type I error rate across all pairwise comparisons of unweighted column means in a two by four fixed effect additive nonorthogonal factorial ANOVA model. In addition we were interested in investigating the effects of sample size and degree of variance heterogeneity on the statistical power of these procedures.

Multiple Comparison Procedures

The GH, Dunnett C, and T3 Procedures

Games and Howell (1976) proposed a solution to the Behrens-Fisher problem which uses the Studentized range distribution. The critical value for the Games-Howell (GH) test based on Equation (1) is equal to $q_{\alpha, K, df_e} / \sqrt{2}$. The decision rule is to reject H_0 if $|T| > q_{\alpha, K, df_e} / \sqrt{2}$, where q_α is the Studentized range distribution for the α centile; K is the number of levels of factor K which form the family of contrasts; and df_s is the approximate degrees of freedom from Satterthwaite (1947):

$$df_s = \frac{\left[\sum_{j=1}^{J}\left(\frac{S_{jk}^2}{n_{jk}} + \frac{S_{jk'}^2}{n_{jk'}}\right)\right]^2}{\sum_{j=1}^{J}\left[\frac{S_{jk}^4}{n_{jk}^2(n_{jk}-1)} + \frac{S_{jk'}^4}{n_{jk'}^2(n_{jk'}-1)}\right]} \qquad (2)$$

Dunnett (1980) suggested two solutions. The first solution, C test, is based on the Cochran (1964) solution to the Behrens-Fisher problem and uses $q / \sqrt{2}$, where

$$q = \frac{\sum_{j=1}^{J}\left(\frac{q_{\alpha,K,n_{jk}-1}S_{jk}^2}{n_{jk}} + \frac{q_{\alpha,K,n_{jk'}-1}S_{jk'}^2}{n_{jk'}}\right)}{\sum_{j=1}^{J}\left(\frac{S_{jk}^2}{n_{jk}} + \frac{S_{jk'}^2}{n_{jk'}}\right)} \qquad (3)$$

The decision rule is if $|T| > q / \sqrt{2}$ then reject the null hypothesis.

The second solution (T3 procedure) suggested by Dunnett uses A_{α,c,df_s} as the critical test statistic. A_α is the Studentized maximum modulus distribution at the α centile. c is the number of contrasts in the family of comparisons. The df_s is defined as Equation (2). The null hypothesis is rejected if $|T| > A_{\alpha,c,df_s}$.

Several studies have investigated the three multiple comparison procedures under various conditions. The differing properties of the data sets used in the investigations have led researchers to make different recommendations as to which procedure is the "best" approach. For instance, several studies have shown the GH procedure to be robust to variance inequality (Keselman & Rogan, 1978; Games, Keselman, & Rogan, 1981), however, Tamhane (1979) indicated that the GH procedure tends to be liberal under some cases. This liberal nature of the GH procedure also was noted in Games and Howell (1976) and Dunnett (1980). Dunnett (1980) indicated that the GH procedure provides an inflated Type I error rate when a large number of groups are compared and variances are homogeneous. With a small number of groups however the empirical Type I error rate did not exceed the nominal significance level.

Stoline (1981) preferred Dunnett C procedure to the GH procedure. He cautions that the GH procedure can be used only if the researcher can bear the risk of being somewhat liberal. However, Dunnett (1980) found the C procedure to be conservative. He showed that the C procedure is comparable to the GH procedure only when degrees of freedom approached infinity. The C procedure gives tighter Type I error rate control than the GH procedure but is generally less powerful.

Dunnett (1980) concluded that the T3 procedure is also conservative. The choice between the T3 procedure and the C procedure depends on whether the degrees of freedom are large or small. For small degrees of freedom, the T3 procedure is more powerful than the C procedure. For large or moderately large degrees of freedom, the C procedure will be more powerful than the T3 procedure and is preferred (Dunnett, 1980; Hsiung & Olejnik, 1991; Stoline, 1981; Toothaker, 1991). However, there are no exact guidelines for deciding whether the degrees of freedom are small. Toothaker (1991) even indicated that the ratio of variances has effect on determining the breaking point between small and large degrees of freedom for the purpose of choosing between the C and T3 procedures.

Wilcox (1987) had investigated the three tests in a two-way ANOVA design when contrasting cell means. His results indicated that the Games-Howell procedure can exceed the nominal α level when the degrees of freedom are small. He also concluded that the T3 procedure appears to be best when the degrees of freedom are less than 50 for any treatment group. Otherwise, he recommends the C procedure.

The Bonferroni Test

Frequently, researchers have used the Bonferroni (B) adjusted *t*-test to control the overall experimentwise Type I error rate. Using the Student *t* distribution each contrast is tested at the α/c significance level (c is the number of contrasts). The decision rule for this procedure is to reject H_0 if $|T| > t_{\alpha', df_e}$, *where* $t_{\alpha'}$ is the Student range distribution at the α' ($\alpha' = \alpha/c$) centile and df_e is defined as Equation (2).

Recently several suggestions for modifying the Bonferroni approach have been made which require using a sequential hypothesis testing approach. The advantage of these procedures is an increase in statistical power while maintaining the overall experimentwise Type I error rate under the nominal significance level. The sequential procedures which have received the greatest attention include the Holm (1979) technique, the Holland & Copenhaver (1988) technique, and the Shaffer (1979, 1986) S and S1 procedures. Hsiung and Olejnik (1991) have compared these modifications of the Bonferroni inequality for a one-way ANOVA design when population variances were heterogeneous. Their study indicated that the magnitude of the power difference between the Holland-Copenhaver test and any one of the Shaffer tests were very small, although the Shaffer tests did generally give greater power. Because the Shaffer procedure is more complicated than the Holland-Copenhaver technique in determining the α' for each of the pairwise comparisons, the Holland-Copenhaver test might be preferred.

The Holland and Copenhaver (HC) (1987) procedure begins by computing a *p*-value for each contrast of interest and then orders these *p*-values from smallest to largest. The smallest *p* value is given a rank of 1 and the largest *p* value is given a rank of *c*. The significance level used to test each contrast is determined as follows:

$$\alpha'_i = 1 - (1 - \alpha)^{\frac{1}{c - r_i + 1}} \tag{4}$$

where α is the desired familywise Type I error rate, c is the total number of contrasts in the family; and r_i represents rank of the contrast being tested. The decision rule is to reject the null hypothesis if p_i value is less than α'_i. Since this is a step down procedure, the process may terminate on any step if a null hypothesis is not rejected.

The Hayter-Fisher Test

Although the Fisher LSD test was the first MCP developed, it is seldom considered by researchers in studying the properties of MCP since it gives a poor α-level protection under all configurations of the means when there are more than three populations. To overcome this limitation, Hayter (1986) suggested that, following a significant omnibus F-test, contrasts could be tested by using the Studentized range distribution instead of the Student t distribution as the reference with degrees of freedom equalling one less than the number of levels of the factor and the degrees of freedom for the error mean square. Seaman, Levin, and Serlin (1991) referred to this test as the Hayter-Fisher test.

Seaman et al. (1991) compared the Hayter-Fisher test with numerous alternatives for one-way ANOVA designs with equal sample sizes and homogeneous variances. Their results indicated that the Hayter procedure can limit the Type I error rate under the nominal α level and has comparable or greater power than the alternatives considered. Seaman et al. (1991) concluded that the Hayter (1986) procedure should be regarded as a viable alternative for pairwise comparisons.

To apply the Hayter-Fisher test in unequal variance situations, two issues need to be considered. First, a valid omnibus test must be used as a preliminary test. Several studies have recommended the James second-order test for the overall analysis when variances are unequal (e.g., Dijkstra & Werter, 1981; Oshima & Algina, 1992; Wilcox, 1988). The F test, Brown-Forsythe F test, and Welch F test have been shown to be liberal when the sample sizes are unequal and variances are severely heterogeneous.

Second, the original Hayter-Fisher test, which uses the pooled within cell variance, is likely to be invalid when cell variances are unequal. Therefore, we suggest using Equation (1) with degrees of freedom from Equation (2) to determine the critical value rather than using the original Hayter-Fisher test statistic. The new alternative is referred to as the Hayter-Welch (HW) test in the present investigation. The null hypothesis for contrast from Equation (1) is reject if

$$|T| > \frac{q_{\alpha, K-1, df_s}}{\sqrt{2}} \qquad (5)$$

where Q_α is Studentized range distribution, df_s is defined as Equation (2).

The James Second-Order Test

James (1951) proposed two solutions to the problem of heterogeneous variances. His first solution, while simpler than the second solution has been shown to lack adequate control over Type I errors (Brown & Forsythe, 1974). The second solution referred to as the James second-order test has received some attention and has been shown to provide adequate control over Type I errors. Wilcox (1989) extended the James second-order test to the two-way ANOVA model. The test statistic U_r is computed as follows:

$$U_r = \sum u_k (Z_k - \bar{Z})^2 \ . \qquad (6)$$

where $Z_k = \sum_j \bar{X}_{jk}$, $u_k = 1/(\sum_j S^2_{jk}/n_{jk})$, $\bar{Z} = \sum u_k Z_k / u_s$, $u_s = \sum u_k$.

The decision rule is to reject H_0 if $U_r > h(\alpha)$ [for detailed description of $h(\alpha)$, refer to Wilcox (1989)].

The James second-order test might be used for contrast analyses. For a pairwise contrasts the square root of Equation (7) is equivalent to Equation (1).

$$J_c = \sqrt{\sum_{k=1}^{2} u_{.k}(Z_{.k} - \overline{Z})^2}$$
$$= \frac{\sum_{j=1}^{J} \overline{X_{j1}} - \sum_{j=1}^{J} \overline{X_{j2}}}{\sqrt{\sum_{j=1}^{j=J} \frac{S_{j1}^2}{n_{j1}} + \sum_{j=1}^{J} \frac{S_{j2}^2}{n_{j2}}}} \quad (7)$$

The critical value could be based on the square root of the critical value used in the James second-order test. The decision rule is to reject H_0 if $Jc > \sqrt{h(\alpha)}$. We refer to this approach as Jc in the present paper. Using $\sqrt{h(\alpha)}$ as the critical value is a Scheffé-type approach which would allow data "snooping" of both complex and pairwise contrasts. A limitation of this approach is that it is likely to have low statistical power. Nonetheless, we felt that since the omnibus James second-order test has been shown to control the Type I error rate under the nominal level while providing adequate power, the Jc test, derived from the omnibus test, should be examined to evaluate its usefulness for contrast analysis.

The Tukey and Kramer Test

In the one-way ANOVA model when the population variances are equal and the design is balanced the Tukey HSD (1953) method is generally recommended (Toothaker, 1991, p. 91), and when the population variances are equal but the design is nonorthogonal the Kramer (1956) modification of the Tukey approach is generally recommended

(Toothaker, 1991, p. 96). The decision rule of the Tukey-Kramer test in factorial designs is if

$$\frac{\left| \sum_{j=1}^{J} \overline{X_{jk}} - \sum_{j=1}^{J} \overline{X_{jk'}} \right|}{\sqrt{MS_W \left(\sum_{j=1}^{J} \frac{1}{n_{jk}} + \sum_{j=1}^{J} \frac{1}{n_{jk'}} \right)}} > \frac{q_{\alpha, K, N-J}}{\sqrt{2}} \qquad (8)$$

then reject the null hypothesis (MS_w is the mean square within samples). Although the Tukey-Kramer (TK) test is not valid for unequal variance situations, it was included in the present study serve as a baseline to compare the other procedures for situations where populations variances are homogeneous.

Table 1 provides a summary of the computational procedures and reference distributions for the seven MCPs considered in the present study.

Insert Table 1 about here

Simulation Study

Procedure

A computer program was written using the SAS-MATRIX (1985) language to evaluate all of the procedures. Data were generated for a two by four fixed effect additive ANOVA model. The study focused on the all pairwise comparisons for the main effect having four levels. Three factors were manipulated: (a) sample size, (b) variance pattern, and (c) effect size. Three types of design were considered: (a) balanced design, (b) slightly unbalanced design, and (c) extremely unbalanced design. The total sample size for each design equaled to 80 (an average cell size of 10), 160 (an average cell size of 20), and 240 (

an average cell size of 30). Six variance patterns were considered and the coefficient of variation (CV) of the six patterns ranged from 0 to 1.4. Table 2 summarizes the patterns of sample size, variance conditions studied. Each of these patterns were associated with two patterns of mean difference, (a) $\mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4}$ and (b) $\mu_{.1} = \mu_{.2} = \mu_{.3} < \mu_{.4}$. A total of 132 conditions were included.

Insert Table 2 about here

Under each of the conditions outlined above, data were generated for the following linear model:

$$Y_{ijk} = M\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + E_{ijk}.$$

$M\mu$ is the grand mean and was set equal to 10 for the study; α_j is the effect of j_{th} level of J, β_k if the effect of k_{th} level of K, and $(\alpha\beta)_{jk}$ is the interaction effect of j_{th} level of J and k_{th} level of K in combination. For the null condition (first mean pattern), the effects of α_j, β_k, $(\alpha\beta)_{jk}$ were all set equal to zero. For the non-null condition (the second mean pattern), the effects of α_j, $\beta_{k'}$, $(k' = 1, 2, 3)$, and $(\alpha\beta)_{jk}$ were set equal to zero, and the effect of β_4 was set equal to a constant, δ. When variances were equal or slightly unequal, δ was set equal to .85, .57, .48 for the sample sizes, 80, 160, 240, respectively. When variance inequality was moderate to extreme, δ was set to 2, 1.5, 1.2 for the sample sizes, 80, 160, 240, respectively. The specific values for the δ were chosen so that for the sample sizes examined (N = 80, 160, 240), the power of the omnibus James second-order test would be generally greater than .45. E_{ijk} is the random error component normally distributed with a mean equalling 0 and variance set equal to the patterns presented in Table 2. The SAS normal random generating function RANNOR was used to generate the random error component.

The familywise nominal significance level was set equal to .05. A total of 10,000 replications was conducted for each condition. The standard error for this number of replications was .022. Based on the nominal α level, a test was interpreted as liberal if its Type I error rate was greater than .0544, and as conservative if its Type I error rate was less than .0456.

Based on the two mean patterns, the frequency of rejecting the null hypotheses was recorded and converted to five proportions to reflect for each procedure the: (a) familywise Type I error rate, (b) partial Type I error rate, (c) any-pair power which is the probability of identifying at least one true non-null contrast, (d) per-pair power, which represents the average power per-non-null contrast, and (e) all-pair power, the probability of identifying all significant non-null contrasts (Einot & Gabriel, 1975; Ramsey, 1978, 1981).

Results

Type I Error Rates

The Type I error rates for the seven pairwise multiple comparison procedures and the omnibus James second-order test for the balanced, slightly unbalanced, and extremely unbalanced designs are reported in Tables 3, 4, 5, respectively.

Insert Tables 3, 4, 5 about here

The James second-order test generally controlled the overall Type I error rate under the nominal α level. This indicates that the James second-order test is a robust test even to the severe variance heterogeneity when the populations sampled have a normal distribution.

The Tukey-Kramer (TK) test was valid in controlling the familywise Type I errors when variances were equal or slightly unequal. For the slightly unbalanced design, Table 4 shows that the TK test had Type I error rates less than the nominal when the total sample size was small (N = 80) and sample size and variance was positively related. However, for most of the unequal variance situations, the TK test was either liberal (when variance was negative related to the sample size) or conservative (when variance was positively related to the sample size) in identifying the true mean difference.

All of the Welch-type multiple comparison procedures considered here controlled the familywise Type I errors under the nominal level. The empirical Type I error rate for the Games-Howell procedure (GH) ranged between .040 to .055. Thus, for the two by four factorial design when all pairwise comparisons are made among four levels for factor K, the GH test does not appear to be liberal. This result is consistent with the findings reported by Dunnett (1980) and Hsiung and Olejnik (1991) for the one-way ANOVA model with the four group design they considered.

The Hayter-Welch approach (HW) test controlled Type I error rate under the nominal α level, ranging from .030 to .051. This indicates that the new alternative is a robust test even the variances are heterogeneous.

Type I errors for the Dunnett T3 and Holland-Copenhaver tests (HC) were nearly identical, ranging from .032 to .045. This result is similar to the findings presented by Hsiung and Olejnik (1991) for the one-way ANOVA model. The two tests are basically conservative.

As was reported in the Dunnett (1980) study, the Type I errors for the Cochran solution (C) increases as sample size increases, ranging from .017 to .041. This indicates that the C procedure tends to be very conservative in the factorial design even when the total sample sizes is large.

Results also show that the Jc test is very conservative, with Type I error rates ranging between .017 and .029. This result indicates that the Jc test behaves much like the Scheffé test and tends to be conservative.

For the pattern of means where $\mu_{.1} = \mu_{.2} = \mu_{.3} < \mu_{.4}$ the partial Type I error rate for the contrasts involving equal population means ($\mu_{.1} = \mu_{2} = \mu_{3}$) was also examined. All of the Welch-type tests had partial Type I error rates less than .05. That is, for the three null contrasts (i.e., $\mu_{.1} = \mu_{.2}$, $\mu_{.1} = \mu_{.3}$, and $\mu_{.2} = \mu_{.3}$), none of the procedures rejected at least one of these contrasts more than 5% of the time.

Statistical Power

Based on the non-null pattern ($\mu_{.1} = \mu_{.2} = \mu_{.3} < \mu_{.4}$) the any-pair power, per-pair power, and all-pair power were estimated for the 66 conditions identified in Table 2. The results are presented below by definition of power since the conclusions vary as a function of the definition.

Any-Pair Power. Tables 6, 7, 8 present the any-pair power of the seven multiple comparison procedures and the omnibus James second-order test for the balanced, slightly unbalanced, and extremely unbalanced designs, respectively. The three tables reveal somewhat similar information about the seven MCPs.

As expected that the TK test generally provided the greatest power when variances were equal. However, the magnitude of power difference between the TK and another test decreased as the sample size increased. For several situations, the GH or HW tests even had greater power than the TK test. The magnitude of power difference between the TK and GH procedures and the TK and HW procedures varied as a function of the samp'e size, ranging from -.017 to .072, but normally the difference in power ranged between .005 to .02. Thus, a researcher employing the GH procedure or the HW procedure instead of the TK test in the equal variance case will lose very little power.

Among the six Welch-type tests, the GH and HW procedures were the two most powerful tests. The GH test generally provided larger any-pair power than the HW test. The HW test, however, was more powerful than the GH test when the variance was inversely related to the sample size. For the situations involving an unbalanced design paired with equal variance, the HW test also gave greater power then the GH test when the total sample size was equal to or greater than 160.

Power for the T3 and HC procedures were nearly identical, but the T3 procedure generally had slightly greater power. The two procedures were consistently more powerful than the C and Jc tests. For the situations where the sample sizes were extremely unequal, the T3 and HC procedures were even more powerful than the HW test when the total sample size was small ($N = 80$).

Two of the least powerful tests were the C and Jc tests. Although it has been reported that the C procedure is preferred over the T3 procedure when degrees of freedom are large or moderately large on the basis of greater power (e.g. Dunnett, 1980; Toothaker, 1991), results of our study show that the C test was consistently less powerful than the T3 test even when the total sample size was large. The Jc test was more powerful than the C test when $N = 80$, but it was less powerful than the C test when $N \geq 160$.

Per-Pair Power. Per-pair (average) power of the seven MCPs in the balanced, slightly unbalanced, and extremely unbalanced designs are presented in Tables 9, 10, 11, respectively. The three tables generally provide a consistent information about the six

Welch-type tests.

Insert Tables 9, 10, and 11 about here

The results show that the TK test was no longer the most powerful test even the variances were equal. The HW test generally gave greater power than the TK test. Thus, a researcher employing the HW test rather than the TK test in the equal variance case may have greater per-pair power.

Among the six Welch-type tests, results show that the HW test generally gave the greatest per-pair power, followed by the GH, HC, T3, C, and Jc tests. The magnitude of power difference between the HW and any of the other procedures was typically between .040 to .060.

The results also show that the HC test was more powerful than the T3 test in terms of per-pair power. This result is consistent with the findings reported by Hsiung and Olejnik (1991) for the one-way ANOVA model.

The C and Jc tests, again, were the two least powerful tests. The C test was more powerful than the Jc test when the total sample size equaled to or greater than 160.

All-Pair Power. Tables 12, 13, 14 include the all-pair power for the seven MCPs for the balanced, slightly unbalanced, and extremely unbalanced designs, respectively. The results form the three tables show that the TK test was less powerful than the GH and HW tests in terms of all-pair power even the variances were equal.

Insert Tables 12, 13, and 14 about here

The three tables give similar information about the relative all-pair power for the six Welch-type tests. Among the six Welch-type tests, the results from these tables indicate that the HW test was the most powerful test, followed by the HC, GH, T3, C, and Jc tests in descending order of power. The magnitude of power difference between the HW and any other test generally exceeded 4%. This indicates that the HW test may provide a power advantage over the alternatives that is of practical importance when the identification of all true pairwise differences are of interest.

The HC test generally offered greater power than the GH and T3 tests. The magnitude of power difference between the HC and GH or T3 tests normally exceeded 2%. This result is consistent with the findings by Hsiung and Olejnik (1991) in the one-way ANOVA model.

The two least powerful tests, again, were the C and Jc tests. The Jc test was only more powerful than the C test when the total sample size equaled to 80.

Conclusions

The present study only investigated all pairwise comparisons among the four columns of a two by four factorial design. A limited number of sample size and variance combinations were considered. In addition, only one type of non-null condition was included in the study. As a result broad generalizations cannot be made. However, the conditions that were studied included many of the situations frequently encountered by the applied researcher. The results of this study are probably best viewed as an indication of the relative merits of the alternative approaches to multiple comparisons when variances differ. With these limitations in mind, the following conclusions seem justified:

1. When all pairwise contrasts among four populations are of interest and the

nominal familywise Type I error rate is set at .05, all six of the Welch-type of multiple comparison procedures considered in this study generally had empirical Type I error rates that did not exceed two standard errors of the nominal significance level. This indicates that the T statistic provides a reasonable solution to the Behrens-Fisher problem for contrast analysis in the factorial design.

2. The identification of the most powerful multiple comparison procedure for the unequal variance case depends on the definition of power. To identify at least one significant difference the GH procedure typically will provide the most sensitive test. However the difference in power between the HW procedure using the omnibus James second-order test and the GH procedure is very small.

3. To maximize the average power per contrast or to identify all significant pairwise differences, the HW procedure can be recommended. Even when the variances were equal the HW test provided greater average and all-pairs power than the TK test which uses the pooled within cell variance.

4. The Dunnett alternatives generally had lower power across all definitions of power than the GH procedure or the HW procedure. Furthermore the ext-nsion of the Cochran solution in a factorial design did not provide greater statistical power than the T3 approach even when the total sample size was large.

5. The Scheffé S-type test, the Jc test, is conservative for the pairwise contrast analysis.

References

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics, 16*, 129-132.

Dijkstra, J. B., & Werter, S. P. J. (1981). Testing the equality of several means when the population variances are unequal. *Communication in Statistics - Simulation and Computation, 10*, 557-569.

Dunnett, C. W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association, 75*, 796-800.

Einot, I., & Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association, 70*, 574-583.

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *Journal of Educational Statistics, 1*, 113-125.

Games P. A., Keselman, H. J., & Rogan, J. C. (1981). Simultaneous pairwise multiple comparison procedures for means when sample size are unequal. *Psychological Bulletin, 90*, 594-598.

Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association, 81*, 1000-1004.

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics, 104*, 145-149.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Hsiung, T. H., & Olejnik, S. (1991, April). *Power of pairwise multiple comparisons in the unequal variance case.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Keselman, H. J., & Rogan, J. 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.

Klockars, A. J., & Hancock, G. R. (1992). Power of recent multiple comparison procedures as applied to a complete set of planned orthogonal contrasts. *Psychological Bulletin, 111*, 505-510.

Kramer, C. Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics, 12*, 307-310.

Oshima, T. C., & Algina, J. (1992). Type I error rates for James's second-order test and Wilcox's Hm test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology, 45*, 255-263.

Ramsey, P. H. (1978). Rejoinder to "Comment on 'Power differences between pairwise multiple comparisons.'" *Journal of the American Statistical Association, 73*, 487.

Ramsey, P. H. (1981). Power of univariate pairwise multiple comparison procedures. *Psychological Bulletin, 90*, 352-366.

SAS Institute (1985). *SAS user's guide: Statistics.* Cary, NC: Author.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics, 2*, 110-114.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin, 110*, 577-586.

Shaffer, J. P. (1979). Comparison of means: An F test followed by a modified multiple range procedure. *Journal of Educational Statistics, 4*, 14-23.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of American Statistical Association, 81*, 826-831.

Stoline, M. R. (1981). The status of multiple comparisons simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *The American Statistician, 35*(3), 134-141.

Tamhane, A. C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association, 74*, 471-480.

Toothaker, L. E. (1991). *Multiple Comparisons for Researchers.* Newbury Park, CA: SAGE.

Tukey, J. W. (1953). *The Problem of Multiple Comparison.* Ditto: Princeton University.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29*, 350-362.

Wilcox, R. R. (1987). Within row pairwise comparisons in a two-way ANOVA design. *Communication in Statistics - Simulation and Computation, 16*, 939-955.

Wilcox, R. R. (1988). A new alternative to the ANOVA F and new results on James's second-order method. *British Journal of Mathematical and Statistical Psychology, 41*, 109-117.

Wilcox, R. R. (1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models. *Journal of Educational Statistics, 14*, 269-278.

Wilcox, R. R. (1990). Comparisons with a control in two-way and one-way designs, and determining whether the most effective treatment has been selected with probability 1-α. *British Journal of Mathematical and Statistical Psychology, 43*, 93-112.

Table 1

Reference of the seven multiple comparison procedures.

Procedure	Computed Test Statistic	Reference Distribution
Dunnett C procedure (C)	$\hat{\psi}^{a}$ / E^{b}	$q^{d}_{\alpha,K,df_w}{}^{e}/\sqrt{2}$
Dunnett T3 procedure (T3)	$\hat{\psi}$ / E	$A^{f}_{\alpha,c,df_s}{}^{g}$
Games & Howell procedure (GH)	$\hat{\psi}$ / E	$q_{\alpha,K,df_s}/\sqrt{2}$
Hayter-Welch procedure (HW)	$\hat{\psi}$ / E	$q_{\alpha,K-1,df_s}$
Holland & Copenhaver technique (HC)	$\hat{\psi}$ / E	$t_{\alpha'^{h},df_s}$
James second-order test (Jc)	$\hat{\psi}$ / E	$(\chi^2_{1-\alpha,K-1})^{1/2}$
Tukey-Kramer test (TK)	$\hat{\psi}$ / E^{c}_{pooled}	$q_{\alpha,K,df}/\sqrt{2}$

Note. ${}^{a}\hat{\psi} = \sum_j \overline{x}_{jk} - \sum_j \overline{x}_{jk'}$.

${}^{b}E = [(\sum_j S^2_{jk} / n_{jk}) + (\sum_j S^2_{jk'} / n_{jk'})]^{1/2}$.

${}^{c}E_{pooled} = [MSw / \sum_j (1/n_{jk}) + \sum_j (1/n_{jk'})]^{1/2}$.

${}^{d}q$: Studentized distribution.

${}^{e}q_{\alpha,\, K,\, df_w}$ (refer to Equation 3).

${}^{f}A$: Studentized Maximum modulus.

${}^{g}df_s$: adjusted degrees of freedom (refer to Equation 2).

${}^{h}\alpha'$: adjusted α level by Holland & Copenhaver technique (1986).

Table 2

Conditions of sample sizes, group mean patterns, and variance patterns for the two by four factorial design ($J = 2, K = 4$).

	$N = 80$	$N = 160$	$N = 240$
$n_{11}, n_{12}, n_{13}, n_{14}; n_{21}, n_{22}, n_{23}, n_{24}$	10, 10, 10, 10 ; 10, 10 ,10, 10[a] 8, 9, 12, 13 ; 8, 9, 10, 11[b] 4, 8, 12, 16 ; 3, 7, 13, 17[c]	20, 20, 20, 20 ; 20, 20, 20, 20[a] 18, 19, 22, 23 ; 18, 19, 20, 21[b] 11, 16, 24, 32 ; 9, 14, 26, 28[c]	30, 30, 30, 30 ; 30, 30, 30, 30[a] 28, 29, 32, 33 ; 28, 29, 30, 31[b] 18, 24, 36, 42 ; 20, 28, 32, 40[c]
$\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}; \mu_{21}, \mu_{22}, \mu_{23}, \mu_{24}$	0, 0, 0, 0 ; 0, 0, 0, 0 0, 0, 0, .85 ; 0, 0, 0, .85[d] 0, 0, 0, 2 ; 0, 0, 0, 2[e]	0, 0, 0, 0 ; 0, 0, 0, 0 0, 0, 0, .57 ; 0, 0, 0, .57[d] 0, 0, 0, 1.5 ; 0, 0, 0, 1.5[e]	0, 0, 0, 0 ; 0, 0, 0, 0 0, 0, 0, .48 ; 0, 0, 0, .48[d] 0, 0, 0, 1.2 ; 0, 0, 0, 1.2[e]
$\sigma^2_{11}: \sigma^2_{12}: \sigma^2_{13}: \sigma^2_{14}; \sigma^2_{21}: \sigma^2_{22}: \sigma^2_{23}: \sigma^2_{24}$	1:1:1:1 ; 1:1:1:1 1:1:2:2 ; 1:1:2:2 1:4:4:9 ; 1:4:4:9 1:4:9:16 ; 1:6:8:12 1:1:9:9 ; 1:1:9:9 1:1:1:16 ; 1:1:1:16	1:1:1:1 ; 1:1:1:1 1:1:2:2 ; 1:1:2:2 1:4:4:9 ; 1:4:4:9 1:4:9:16 ; 1:6:8:12 1:1:9:9 ; 1:1:9:9 1:1:1:16 ; 1:1:1:16	1:1:1:1 ; 1:1:1:1 1:1:2:2 ; 1:1:2:2 1:4:4:9 ; 1:4:4:9 1:4:9:16 ; 1:6:8:12 1:1:9:9 ; 1:1:9:9 1:1:1:16 ; 1:1:1:16

Note. [a]Balanced design. [b]Slightly unbalanced design. [c]Extremely unbalanced design. [d]Mean pattern for the patterns of equal variance and slightly unequal variance. [e]Mean pattern for the situations of moderately heterogeneous variance and severely heterogeneous variance.

Table 3

Proportion of familywise Type I errors for the two by four **balanced** *factorial design when all pairwise comparisons are made among levels for factor K.*

n_{jk}[a]	VP[b]	CV[c]	Jm[d]	Method[e]						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	.046	.051	.049	.045	.041	.040	.019	.027
	2	.33	.045	.056	.049	.044	.040	.038	.019	.025
	3	.64	.051	.070	.053	.049	.044	.041	.022	.028
	4	.69	.047	.064	.045	.044	.039	.037	.021	.026
	5	.80	.048	.075	.046	.045	.039	.036	.022	.026
	6	1.4	.045	.091	.044	.042	.033	.032	.017	.022
20	1	0	.049	.049	.051	.048	.043	.043	.033	.028
	2	.33	.047	.052	.049	.045	.041	.041	.031	.026
	3	.64	.050	.064	.049	.048	.041	.041	.033	.026
	4	.69	.050	.067	.049	.047	.039	.038	.031	.025
	5	.80	.049	.070	.044	.045	.036	.035	.028	.023
	6	1.4	.051	.087	.045	.046	.036	.036	.029	.025
30	1	0	.052	.050	.052	.051	.044	.043	.037	.026
	2	.33	.051	.052	.050	.049	.043	.042	.037	.027
	3	.64	.049	.057	.044	.045	.037	.036	.033	.023
	4	.69	.055	.070	.052	.051	.045	.044	.041	.030
	5	.80	.050	.066	.045	.044	.038	.038	.034	.024
	6	1.4	.048	.076	.043	.044	.034	.034	.031	.022

Note. [a]n_{jk}: The average cell size. [b]VP: Variance patterns (see Table 1).

[c]CV: Coefficient of variation. [d]Jm: the omnibus James second-order test.

[e]TK: The Tukey-Kramer test. GH: The Games-Howell test.

HW: The Hayter-Welch modified t test. T3: The Dunnett T3 test. C: The Dunnett C test. Jc: The James second-order test for contrast analysis.

Table 4

Proportion of familywise Type I errors for the two by four ***slightly unbalanced design*** *when all pairwise comparisons are made among levels for factor K.*

n_{jk}[a]	VP	CV	r[b]	Jm	Method						
					TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	048	.051	.048	.047	.042	.040	.020	.027
	2	.33	.87	.046	.041	.049	.045	.039	.038	.018	.025
	3	.64	.80	.051	.049	.052	.048	.043	.041	.021	.027
	4+	.69	.94	.047	.039	.048	.044	.040	.038	.021	.027
	4-	.69	-.89	.048	.106	.048	.045	.045	.037	.020	.027
	5	.80	.86	.046	.048	.045	.043	.037	.035	.018	.023
	6+	1.4	.67	.047	.053	.043	.045	.036	.034	.018	.024
	6-	1.4	-.67	.047	.149	.045	.044	.043	.034	.017	.025
20	1	0	ϵ	.052	.053	.053	.051	.045	.045	.034	.029
	2	.33	.87	.049	.048	.051	.048	.041	.039	.030	.025
	3	.64	.80	.051	.055	.048	.048	.039	.039	.031	.025
	4+	.69	.94	.052	.059	.051	.048	.043	.043	.036	.029
	4-	.69	-.89	.046	.079	.042	.042	.035	.035	.028	.023
	5	.80	.86	.048	.057	.045	.044	.037	.037	.030	.025
	6+	1.4	.67	.052	.063	.047	.048	.039	.039	.032	.027
	6-	1.4	-.67	.048	.108	.042	.043	.035	.034	.027	.022
30	1	0	0	.049	.050	.052	.048	.044	.043	.037	.027
	2	.33	.87	.052	.051	.052	.050	.045	.044	.040	.027
	3	.64	.80	.052	.058	.051	.048	.044	.044	.039	.026
	4+	.69	.94	.052	.058	.050	.048	.042	.042	.038	.026
	4-	.69	-.89	.050	.075	.046	.045	.039	.039	.036	.025
	5	.80	.86	.050	.056	.045	.046	.038	.037	.035	.024
	6+	1.4	.67	.049	.064	.044	.044	.037	.037	.034	.023
	6-	1.4	-.67	.048	.098	.041	.044	.034	.033	.030	.020

Note. [a]n_{jk}: The average of cell size (see Table 2 for the patterns of sample size for the slightly unbalanced design. [b]r: The correlation coefficient of sample size and variance.

Table 5

Proportion of familywise Type I errors for the two by four ***extremely unbalanced*** *factorial design when all pairwise comparisons are made among levels for factor* ***K.***

$\underline{n}_{jk}$[a]	VP	CV	r	Jm	Method TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.045	.049	.055	.044	.045	.041	.014	.027
	2	.33	.90	.041	.028	.055	.041	.045	.042	.013	.024
	3	.64	.91	.040	.019	.049	.038	.041	.038	.013	.020
	4^+	.69	.94	.037	.019	.047	.035	.036	.033	.011	.017
	4^-	.69	-.94	.045	.207	.046	.042	.038	.034	.016	.028
	5	.80	.91	.040	.024	.048	.039	.039	.038	.016	.020
	6^+	1.4	.76	.044	.013	.050	.042	.041	.038	.016	.027
	6^-	1.4	-.76	.042	.361	.040	.030	.034	.028	.013	.029
20	1	0	ϵ	.051	.048	.051	.049	.042	.041	.029	.027
	2	.33	.93	.049	.030	.051	.048	.042	.041	.029	.027
	3	.64	.87	.053	.028	.053	.051	.043	.043	.030	.026
	4^+	.69	.95	.055	.026	.054	.053	.046	.046	.035	.030
	4^-	.69	-.91	.050	.158	.046	.046	.037	.036	.026	.025
	5	.80	.94	.048	.030	.048	.045	.038	.038	.027	.023
	6^+	1.4	.72	.050	.020	.047	.047	.038	.037	.027	.024
	6^-	1.4	-.72	.047	.252	.040	.043	.033	.032	.025	.024
30	1	0	ϵ	.049	.048	.049	.047	.043	.042	.036	.027
	2	.33	.89	.049	.033	.051	.048	.044	.043	.038	.025
	3	.64	.91	.050	.033	.047	.048	.040	.040	.035	.025
	4^+	.69	.98	.052	.030	.051	.049	.043	.042	.039	.025
	4^-	.69	-.98	.051	.129	.046	.047	.037	.037	.034	.026
	5	.80	.89	.048	.037	.048	.045	.040	.039	.036	.025
	6^+	1.4	.75	.052	.027	.045	.048	.038	.038	.033	.024
	6^-	1.4	-.75	.050	.188	.043	.045	.037	.036	.033	.023

Note. [a]n_{jk}: The average cell size (see Table 2 for the patterns of sample size for the extremely unbalanced design.

Table 6

Any-pair power for the two by four ***balanced*** *factorial design when all pairwise comparisons are made among levels for factor K.*

n_{jk}[a]	VP	CV	Jm	Method						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	.727	.753	.733	.725	.703	.696	.579	.626
	2	.33	.458	.546	.474	.450	.442	.436	.326	.369
	3	.64	.596	.773	.616	.586	.581	.571	.479	.520
	4	.75	.405	.587	.414	.388	.381	.371	.298	.328
	5	.80	.604	.729	.617	.591	.580	.571	.491	.531
	6	1.4	.377	.663	.353	.351	.319	.308	.259	.289
20	1	0	.727	.730	.725	.725	.699	.699	.659	.631
	2	.33	444	.510	.456	.436	.423	.423	.384	.359
	3	.64	.696	.827	.710	.687	.683	.682	.655	.626
	4	.75	.494	.640	.499	.478	.467	.466	.435	.403
	5	.80	.711	.794	.718	.701	.692	.689	.665	.640
	6	1.4	.459	.715	.439	.432	.406	.403	.384	.361
30	1	0	.764	.763	.764	.762	.741	.739	.720	.673
	2	.33	.482	.535	.492	.473	.463	.461	.443	.396
	3	.64	.691	.814	.711	.683	.682	.681	.669	.623
	4	.75	.492	.616	.495	.475	.454	.463	.453	.402
	5	.80	.700	.773	.712	.689	.687	.686	.677	.627
	6	1.4	.451	.670	.435	.424	.409	.409	.403	.360

Note. [a]n_{jk}: see Table 3.

Table 7

Any-pair power for the two by four slightly unbalanced factorial design when all pairwise comparisons are made among levels for factor K.

n_{jk}[a]	VP[b]	CV	r	Jm	Method						
					TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.790	.804	.789	.788	.765	.760	.648	.685
	2	.33	.87	.511	.555	.535	.506	.501	.495	.378	.412
	3	.64	.80	.668	.775	.704	.664	.674	.667	.581	.596
	4^+	.69	.94	.484	.553	.510	.472	.477	.470	.386	.399
	4^-	.69	-.89	.977	.827	.925	.960	.909	.902	.856	.871
	5	.80	.86	.682	.705	.711	.675	.680	.674	.596	.611
	6^+	1.4	.67	.446	.636	.443	.424	.409	.401	.351	.357
	6^-	1.4	-.67	1	1	1	1	1	1	1	1
20	1	0	ϵ	.754	.749	.748	.752	.721	.720	.680	.648
	2	.33	.87	.479	.517	.489	.471	.458	.457	.414	.380
	3	.64	.80	.736	.829	.756	.730	.730	.729	.705	.673
	4^+	.69	.94	.540	.639	.557	.528	.525	.523	.498	.458
	4^-	.69	-.89	1	.970	.999	1	.999	.999	.999	.999
	5	.80	.86	.744	.780	.757	.737	.735	.734	.713	.679
	6^+	1.4	.67	.495	.709	.489	.471	.456	.453	.439	.401
	6^-	1.4	-.67	1	.976	1	1	1	1	1	1
30	1	0	ϵ	.778	.769	.772	.777	.748	.746	.730	.684
	2	.33	.87	.513	.553	.524	.505	.496	.494	.476	.423
	3	.64	.80	.722	.819	.739	.716	.717	.716	.707	.658
	4^+	.69	.94	.515	.615	.525	.501	.497	.497	.487	.431
	4^-	.69	.89	.988	.814	.964	.980	.955	.955	.951	.932
	5	.80	.86	.724	.770	.740	.717	.713	.711	.705	.653
	6^+	1.4	.67	.471	.690	.458	.450	.432	.431	.429	.379
	6^-	1.4	-.67	1	.964	1	1	1	1	1	1

Note. [a]n_{jk}: see Table 4. [b]For variance patterns 4 and 6, + represents variance pattern is directly paired with sample size; - represents variance parrtern is inversly paired with sample size.

Table 8

Any-pair power for the two by four extremely unbalanced design when all pairwise comparisons are made among levels for factor K.

n_{jk}[a]	VP[b]	CV	r	Jm	Method TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.767	.836	.818	.764	.792	.786	.700	.619
	2	.33	.90	.499	.536	.588	.495	.556	.547	.418	.375
	3	.64	.91	.731	.752	.819	.731	.794	.791	.670	.653
	4^+	.69	.94	.528	.507	.619	.524	.587	.582	.439	.435
	4^-	.69	-.94	.957	.957	.950	.952	.937	.934	.911	.835
	5	.80	.91	.751	.659	.842	.750	.818	.815	.734	.681
	6^+	1.4	.76	.477	.566	.616	.468	.584	.579	.516	.397
	6^-	1.4	-.76	1	.999	1	1	1	1	1	1
20	1	0	ϵ	.816	.796	.794	.810	.771	.768	.732	.693
	2	.33	.93	.550	.503	.549	.545	.519	.517	.466	.428
	3	.64	.87	.842	.828	.863	.841	.845	.845	.820	.791
	4^+	.69	.95	.654	.580	.680	.647	.652	.651	.615	.570
	4^-	.69	-.91	.994	.960	.982	.989	.976	.976	.973	.959
	5	.80	.94	.852	.728	.872	.849	.856	.855	.838	.804
	6^+	1.4	.72	.621	.666	.651	.610	.623	.623	.608	.547
	6^-	1.4	-.72	1	.999	1	1	1	1	1	1
30	1	0	ϵ	.839	.817	.820	.834	.796	.795	.783	.737
	2	.33	.89	.578	.551	.580	.573	.551	.549	.532	.470
	3	.64	.91	.806	.813	.826	.803	.809	.807	.799	.749
	4^+	.69	.98	.608	.578	.626	.599	.600	.599	.590	.529
	4^-	.69	-.98	.991	.927	.969	.983	.962	.961	.960	.939
	5	.80	.89	.821	.747	.841	.818	.824	.823	.818	.771
	6^+	1.4	.75	.579	.658	.585	.561	.561	.559	.558	.490
	6^-	1.4	-.75	1	.998	1	1	1	1	1	1

Note. [a]n_{jk}: see Table 5. [b]: see Table 7.

Table 9

Per-pair power for the two by four balanced factorial design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	Method TK	GH	HW	T3	HC	C	Jc
10	1	0	.516	.485	.551	.454	.475	.343	.382
	2	.33	.341	.281	.316	.255	.260	.172	.201
	3	.64	.618	.448	.485	.416	.434	.322	.355
	4	.69	.421	.273	.299	.248	.259	.182	.203
	5	.80	.570	.442	.479	.411	.427	.329	.360
	6	1.4	.586	.286	.314	.257	.267	.204	.231
20	1	0	.481	.473	.537	.445	.472	.405	.380
	2	.33	.311	.267	.303	.243	.259	.214	.197
	3	.64	.683	.536	.582	.508	.534	.476	.446
	4	.69	.469	.337	.370	.310	.329	.282	.257
	5	.80	.644	.540	.582	.511	.537	.482	.456
	6	1.4	.639	.361	.390	.330	.352	.312	.289
30	1	0	.514	.514	.577	.489	.513	.467	.420
	2	.33	.329	.292	.331	.271	.286	.255	.219
	3	.64	.668	.534	.576	.506	.532	.491	.444
	4	.69	.452	.332	.366	.306	.326	.295	.253
	5	.80	.616	.527	.566	.501	.527	.489	.443
	6	1.4	.623	.361	.386	.335	.358	.330	.289

Table 10

Per-pair power for the two by four slightly unbalanced factorial design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	r	Method						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.549	.517	.585	.488	.510	.374	.410
	2	.33	.87	.332	.312	.348	.285	.300	.199	.221
	3	.64	.80	.598	.518	.556	.487	.509	.393	.409
	4^+	.69	.94	.369	.330	.359	.303	.318	.230	.240
	4^-	.69	-.89	.489	.524	.602	.495	.510	.428	.459
	5	.80	.86	.527	.521	.556	.490	.512	.408	.422
	6^+	1.4	.67	.540	.359	.382	.329	.346	.273	.280
	6^-	1.4	-.67	.791	.734	.763	.728	.740	.711	.728
20	1	0	ϵ	.499	.489	.555	.460	.487	.421	.392
	2	.33	.87	.309	.285	.326	.260	.277	.230	.208
	3	.64	.80	.681	.580	.626	.550	.579	.518	.484
	4^+	.69	.94	.459	.377	.413	.350	.371	.325	.292
	4^-	.69	-.89	.748	.725	.787	.706	.734	.689	.673
	5	.80	.86	.619	.573	.613	.547	.573	.522	.488
	6^+	1.4	.67	.625	.401	.427	.370	.394	.352	.319
	6^-	1.4	-.67	.821	.774	.805	.765	.783	.758	.753
30	1	0	ϵ	.522	.524	.589	.497	.524	.477	.429
	2	.33	.87	.337	.312	.351	.291	.306	.276	.237
	3	.64	.80	.670	.562	.607	.535	.564	.521	.471
	4^+	.69	.94	.440	.353	.389	.329	.350	.320	.275
	4^-	.69	-.89	.469	.633	.703	.611	.638	.605	.563
	5	.80	.86	.609	.554	.591	.527	.522	.517	.470
	6^+	1.4	.67	.607	.378	.405	.353	.376	.351	.302
	6^-	1.4	-.67	.779	.777	.804	.767	.785	.766	.754

Table 11

Per-pair power for the two by four extremely unbalanced design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	r	Method						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.508	.455	.495	.427	.435	.329	.311
	2	.33	.90	.268	.304	.315	.279	.287	.187	.176
	3	.64	.91	.447	.591	.611	.560	.585	.402	.418
	4^+	.69	.94	.252	.395	.404	.366	.386	.243	.250
	4^-	.69	-.94	.631	.462	.520	.440	.442	.394	.375
	5	.80	.91	.359	.606	.622	.578	.601	.438	.436
	6^+	1.4	.76	.352	.483	.426	.452	.475	.354	.284
	6^-	1.4	-.76	.869	.684	.706	.680	.679	.671	.695
20	1	0	ϵ	.492	.478	.546	.450	.473	.406	.380
	2	.33	.93	.256	.301	.345	.278	.292	.240	.218
	3	.64	.87	.589	.677	.724	.651	.680	.616	.577
	4^+	.69	.95	.335	.461	.503	.435	.459	.401	.361
	4^-	.69	-.91	.649	.548	.608	.528	.545	.505	.501
	5	.80	.94	.486	.686	.726	.662	.689	.634	.592
	6^+	1.4	.72	.493	.541	.557	.512	.539	.493	.433
	6^-	1.4	-.72	.879	.714	.735	.708	.717	.699	.709
30	1	0	ϵ	.538	.533	.604	.507	.533	.489	.441
	2	.33	.89	.308	.335	.385	.312	.328	.297	.253
	3	.64	.91	.617	.642	.689	.617	.645	.604	.547
	4^+	.69	.98	.366	.420	.464	.396	.418	.386	.333
	4^-	.69	-.98	.599	.579	.642	.558	.581	.551	.548
	5	.80	.89	.536	.642	.684	.619	.644	.611	.556
	6^+	1.4	.75	.530	.483	.508	.457	.481	.455	.391
	6^-	1.4	-.75	.875	.735	.760	.730	.742	.726	.723

Table 12

All-pair power for the two by four balanced factorial design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	Method						
			TK	GH	HW	T3	HC	C	Jc
10	1	0	.272	.238	.333	.211	.257	.129	.155
	2	.33	.152	.104	.164	.088	.119	.045	.060
	3	.64	.455	.275	.353	.248	.293	.171	.195
	4	.75	.266	.135	.190	.119	.148	.073	.084
	5	.80	.405	.226	.306	.202	.244	.136	.156
	6	1.4	.507	.223	.273	.198	.230	.151	.175
20	1	0	.235	.226	.315	.202	.253	.172	.153
	2	.33	.134	.101	.153	.087	.114	.069	.061
	3	.64	.523	.349	.441	.323	.377	.289	.263
	4	.75	.302	.176	.236	.156	.193	.133	.116
	5	.80	.480	.308	.396	.281	.335	.248	.225
	6	1.4	.563	.285	.341	.258	.304	.242	.219
30	1	0	.262	.258	.351	.236	.284	.217	.180
	2	.33	.143	.111	.166	.099	.125	.089	.068
	3	.64	.508	.349	.437	.320	.376	.304	.260
	4	.75	.291	.169	.233	.152	.199	.141	.114
	5	.80	.447	.291	.377	.269	.320	.255	.213
	6	1.4	.5477	.288	.342	.264	.311	.260	.223

Table 13

All-pair power for the two by four slightly unbalanced factorial design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	r	Method						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.284	.240	.335	.212	.261	.126	.156
	2	.33	.87	.135	.117	.174	.100	.130	.054	.065
	3	.64	.80	.419	.322	.413	.293	.344	.205	.221
	4^+	.69	.94	.204	.157	.222	.138	.172	.089	.096
	4^-	.69	-.89	.153	.114	.187	.092	.122	.051	.077
	5	.80	.86	.352	.286	.371	.258	.308	.181	.192
	6^+	1.4	.67	.444	.277	.333	.252	.293	.201	.207
	6^-	1.4	-.67	.534	.211	.289	.186	.220	.134	.184
20	1	0	ϵ	.246	.230	.319	.207	.258	.177	.156
	2	.33	.87	.126	.105	.162	.090	.119	.074	.064
	3	.64	.80	.520	.388	.484	.355	.419	.320	.288
	4^+	.69	.94	.289	.199	.270	.180	.220	.157	.133
	4^-	.69	-.89	.443	.353	.471	.320	.386	.294	.268
	5	.80	.86	.450	.330	.413	.302	.355	.276	.245
	6^+	1.4	.67	.540	.316	.377	.286	.338	.269	.239
	6^-	1.4	-.67	.584	.323	.416	.295	.350	.274	.258
30	1	0	ϵ	.271	.269	.359	.244	.297	.225	.186
	2	.33	.87	.142	.120	.176	.106	.130	.097	.075
	3	.64	.80	.507	.370	.462	.342	.401	.324	.278
	4^+	.69	.94	.273	.181	.248	.164	.202	.154	.124
	4^-	.69	-.89	.134	.230	.328	.201	.257	.197	.155
	5	.80	.86	.441	.314	.394	.290	.338	.278	.236
	6^+	1.4	.67	.523	.301	.356	.278	.323	.276	.230
	6^-	1.4	-.67	.511	.331	.412	.303	.355	.297	.261

Table 14

All-pair power for the two by four extremely unbalanced design when all pairwise comparisons are made among levels for factor K.

n_{jk}	VP	CV	r	Method						
				TK	GH	HW	T3	HC	C	Jc
10	1	0	ϵ	.159	.088	.157	.072	.085	.017	.051
	2	.33	.90	.045	.059	.106	.048	.062	.011	.025
	3	.64	.91	.135	.344	.441	.310	.364	.142	.190
	4^+	.69	.94	.038	.183	.253	.159	.203	.077	.088
	4^-	.69	-.94	.237	.025	.064	.019	.020	.005	.023
	5	.80	.91	.082	.331	.430	.298	.352	.147	.171
	6^+	1.4	.76	.118	.338	.368	.310	.362	.163	.172
	6^-	1.4	-.76	.615	.052	.120	.041	.038	.014	.089
20	1	0	ϵ	.183	.159	.237	.140	.178	.102	.102
	2	.33	.93	.057	.085	.131	.073	.095	.054	.048
	3	.64	.87	.338	.459	.562	.427	.491	.385	.342
	4^+	.69	.95	.126	.238	.319	.216	.264	.189	.161
	4^-	.69	-.91	.272	.090	.148	.074	.100	.055	.065
	5	.80	.94	.259	.443	.534	.414	.473	.383	.330
	6^+	1.4	.72	.312	.424	.490	.394	.451	.371	.316
	6^-	1.4	-.72	.648	.143	.206	.124	.152	.097	.126
30	1	0	ϵ	.246	.236	.327	.211	.263	.194	.160
	2	.33	.89	.098	.120	.181	.106	.135	.096	.073
	3	.64	.91	.416	.432	.532	.403	.465	.388	.330
	4^+	.69	.98	.183	.215	.196	.239	.193	.151	.005
	4^-	.69	-.98	.229	.136	.207	.116	.155	.109	.089
	5	.80	.89	.336	.378	.471	.352	.406	.343	.286
	6^+	1.4	.75	.396	.380	.444	.357	.406	.354	.297
	6^-	1.4	-.75	.647	.207	.280	.189	.226	.177	.168