DOCUMENT RESUME

ED 361 345 TM 020 373

| | |
|---|---|
| AUTHOR | Lix, Lisa M.; Keselman, H. J. |
| TITLE | Interaction Contrasts in Repeated Measures Designs. |
| PUB DATE | Apr 93 |
| NOTE | 33p.; 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 | Comparative Analysis; Computer Simulation; Equations (Mathematics); *Estimation (Mathematics); *Interaction; *Mathematical Models; *Monte Carlo Methods; *Research Design; *Sample Size |
| IDENTIFIERS | Bonferroni Procedure; Omnibus Tests; *Repeated Measures Design; Scheffes Contrast Test; Sphericity Tests; Type I Errors |

ABSTRACT

Current omnibus procedures for the analysis of interaction effects in repeated measures designs which contain a grouping variable are known to be nonrobust to violations of multisample sphericity, particularly when group sizes are unequal. An alternative approach is to formulate a comprehensive set of contrasts on the data which probe the specific nature of the interaction. Six interaction contrast procedures are compared via Monte Carlo methods. Two test statistics are considered, one relying on an estimate of the standard error of the contrast formed by pooling across levels of the groups and trials factors, and the other employing a nonpooled estimate based on only that data used in defining a contrast. A Scheffe, Studentized maximum modulus, and Hochberg step-up Bonferroni critical value are paired with each statistic. Only the Studentized maximum modulus and Hochberg nonpooled procedures provided acceptable rates of familywise Type I error control under departures from multisample sphericity when the data was nonnormal. Six tables present data from the analyses. (Contains 23 references.) (Author)

# Interaction Contrasts in Repeated Measures Designs

Lisa M. Lix

H. J. Keselman

Department of Psychology

University of Manitoba

Paper presented at the annual meeting of the

American Educational Research Association

Atlanta, GA

April 12 - 16, 1993

## Abstract

Current omnibus procedures for the analysis of interaction effects in repeated measures designs which contain a grouping variable are known to be nonrobust to violations of multisample sphericity, particularly when group sizes are unequal. An alternative approach is to formulate a comprehensive set of contrasts on the data which probe the specific nature of the interaction. Six interaction contrast procedures are compared via Monte Carlo methods. Two test statistics are considered, one relying on a estimate of the standard error of the contrast formed by pooling across levels of the groups and trials factors, and the other employing a nonpooled estimate based on only that data used in defining a contrast. A Scheffe, Studentized maximum modulus, and Hochberg step-up Bonferroni critical value are paired with each statistic. Only the Studentized maximum modulus and Hochberg nonpooled procedures provided acceptable rates of familywise Type I error control under departures from multisample sphericity when the data was nonnormal.

Interaction Contrasts in Repeated Measures Designs

A common experimental design in the fields of education and psychology is one in which successive measurements are obtained from the same subjects under each of K treatments. Frequently a grouping factor is employed in the design, so that $n_j$ subjects in J independent groups ($\Sigma n_j = N$) are exposed to all levels of the within-subjects variable. The observations $X_{ijk}$, $i = 1, \ldots, n_j$; $j = 1, \ldots, J$; $k = 1, \ldots, K$ are assumed to be independent, normal, and identically distributed within each level j of the grouping factor, with common mean vector $\mu_j$ and covariance matrix $\Sigma_j$.

It is well known that unless the data from such a groups by trials design satisfy certain parametric conditions the ANOVA F-test will produce invalid results for tests of hypotheses on the within-subjects trials and interaction effects. Specifically, it is assumed that a set of orthonormalized contrasts on the K trials exhibit a common variance. This assumption is known as sphericity or circularity (Huynh & Feldt, 1970). Furthermore, this variance must be constant across all levels of the between-subjects factor. Jointly, these assumptions are referred to as multisample sphericity (Huynh, 1978).

Recent evidence suggests that educational and psychological research will rarely, if ever, conform to the multisample sphericity assumption (Green & Barcikowski, 1992). Consequently, it is recommended that researchers adopt either an adjusted degrees of freedom (df) univariate F-test or a multivariate test for examining the trials main effect.

However, no such recommendations are forthcoming from the literature with respect to the interaction effect. Recent research conducted by Belli (1988) and Keselman and Keselman (1990) indicates that these univariate and multivariate procedures are sensitive to violations of multisample sphericity, particularly when group sizes are unequal. Both studies found that in situations where there was a positive relationship between group sizes and covariance matrices, in which the group with the largest sample size also exhibited a covariance matrix with the largest element values, the statistical tests produced empirical Type I error rates which were less than the nominal level of significance, $\alpha$. Furthermore, this conservatism increased as the degree of group size inequality and/or covariance matrix heterogeneity increased for these positive pairings. In the case of a negative pairing, where the group with the largest sample size also exhibited a covariance matrix with the smallest element values, the empirical Type I error rates were excessively large, attaining values as high as .40 for a nominal $\alpha$ of .05. As well, the liberalness of these statistical procedures increased as the degree of group size inequality and/or covariance heterogeneity increased.

At present then, there are neither valid nor robust parametric tests of the interaction effect in mixed designs when the data do not conform to the assumption of multisample sphericity and the design is unbalanced. This represents a significant gap in methods available for the analysis of mixed designs, as researchers are typically most interested in testing for the presence of an interaction effect. The use of existing univariate and multivariate methods may lead researchers to draw unreliable, and even erroneous conclusions about their data.

The purpose of this study was to examine, via Monte Carlo methods, alternatives to existing parametric tests of the interaction effect in mixed repeated measures designs, in order to identify procedures which may be robust to departures from multisample sphericity in unbalanced designs.

## Definition of Test Statistics

Two techniques frequently adopted for examining interaction effects are simple main effect tests and tetrad contrasts. The latter are considered to be a better choice for two reasons: (a) the null hypothesis under consideration for a particular contrast is consistent with that of an omnibus test of the interaction effect (Boik, 1975; Marascuilo & Levin, 1970), and (b) tetrad contrasts may provide the researcher with specific information regarding the combinations of factor levels which contribute to the interaction effect (Gabriel, Putter, & Wax, 1973).

A tetrad contrast is defined as

$$\psi = (\mu_{jk} - \mu_{jk'}) - (\mu_{j'k} - \mu_{j'k'}) \quad , \tag{1}$$

where $j \neq j'$, $k \neq k'$, and $\mu_{jk}$ is estimated by $\bar{X}_{jk}$, the *jk*th sample mean. In other words, a tetrad contrast involves testing for the presence of an interaction between rows and columns in a 2 x 2 submatrix of the J x K data matrix. Given the likelihood of an incorrect conclusion regarding an omnibus test of the interaction in mixed designs, one may address specific interaction questions by adopting a data analysis strategy of bypassing the omnibus test in favour of conducting all possible tetrad contrasts.

Two choices of a test statistic exist for performing these interaction contrasts in the mixed design. One statistic employs an estimate of the standard error of the contrast formed using $MS_{KxS/J}$, the error mean square for the omnibus F-test of the interaction. The test statistic is

$$t = \frac{(\bar{X}_{jk} - \bar{X}_{jk'}) - (\bar{X}_{j'k} - \bar{X}_{j'k'})}{\sqrt{MS_{K \times S/J}\left[\frac{2}{n_j} + \frac{2}{n_{j'}}\right]}}, \tag{2}$$

which follows Student's t distribution with df $\upsilon = (K - 1)(N - J)$. However, if multisample sphericity is not present in the data, the error term for this statistic, which involves pooling over both the within-subjects and between-subjects factors, will produce a biased estimate of the standard error of the contrast.

An alternate statistic employs a standard error derived only from that data used in forming the contrast and is defined as

$$t = \frac{(\bar{X}_{jk} - \bar{X}_{jk'}) - (\bar{X}_{j'k} - \bar{X}_{j'k'})}{\sqrt{\frac{c^T S_j c}{n_j} + \frac{c^T S_{j'} c}{n_{j'}}}}, \tag{3}$$

where $S_j$ is the sample covariance matrix for the *j*th group, and **c** is a K x 1 vector of coefficients which contrasts the *k*th and *k'*th levels of the within-subjects factor. In other words, the standard error of the tetrad contrast is formed using only four cells of the J x K data matrix. The nonpooled statistic does not follow a t distribution, but can be approximated by Student's t with Satterthwaite (1946) estimated df

$$\upsilon_s = \frac{\left[\frac{c^T S_j c}{n_j} + \frac{c^T S_{j'} c}{n_{j'}}\right]^2}{\frac{[c^T S_j c / n_j]^2}{n_j - 1} + \frac{[c^T S_{j'} c / n_{j'}]^2}{n_{j'} - 1}} . \quad (4)$$

Keselman, Keselman, and Shaffer (1991) examined the robustness of four simultaneous procedures for pairwise contrasts on repeated measures means in a mixed design when the multisample sphericity assumption was violated and group sizes were unequal. The procedures all relied on a test statistic which employed a nonpooled error term [i.e. equation (3)], but utilized different critical values. The authors demonstrated that the probability of committing one or more Type I errors, otherwise known as the familywise error rate (FWR), could be controlled for pairwise contrasts when this test statistic was used in combination with either a Bonferroni, Studentized range, or Studentized maximum modulus critical value.

To limit the FWR for the set of all possible tetrad contrasts, Marascuilo and Levin (1970) suggest adopting a Scheffe (1953) critical value (CV), $\upsilon_1 F[\alpha; \upsilon_1, \upsilon_2]$, where $\upsilon_1 = (J - 1)(K - 1)$ and $\upsilon_2$ is the error df. However, given that Scheffe's method is likely to produce conservative results, Gabriel et al. (1973) recommend a Bonferroni CV (Dunn, 1961), $t[\alpha/(2C); \upsilon_2]$, where

$C = J^*K^*$, $J^* = J(J - 1)/2$, and $K^* = K(K - 1)/2$. Furthermore, Hochberg and Tamhane (1987, p. 299) suggest that a Studentized maximum modulus CV, $M[\alpha; C, \upsilon_2]$, is more likely to maintain the FWR at its upper bound than a Bonferroni CV.

An alternative to adopting one of these simultaneous multiple comparison procedures (MCPs), is to select a stepwise procedure. Hochberg (1988) developed a step-up procedure which is based on the Bonferroni inequality, and hence can provide control of the FWR. However, since a different CV is used at each stage of hypothesis testing, Hochberg's method may provide greater power to detect treatment effects than the Dunn-Bonferroni method.

Hochberg's (1988) procedure is also an attractive choice because it is one of the simpler stepwise procedures available. Under this method, one begins by rank ordering the p-values corresponding to the statistics used for testing the hypotheses $H_{(1)}, \ldots, H_{(C)}$ (i.e. $\psi = 0$), so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(C)}$ represent the ordered p-values. Testing begins with the hypothesis corresponding to the largest p-value, $p_{(C)}$. If $p_{(C)} \leq \alpha$, all C hypotheses are rejected; if not, $H_{(C)}$ is retained and testing moves to $H_{(C-1)}$. If $p_{(C-1)} \leq \alpha/2$, $H_{(C-1)}$ is rejected, as are all remaining hypotheses; if not $H_{(C-1)}$ is also retained, and $p_{(C-2)}$ is compared to $\alpha/3$, and so on. This continues, if all previous hypotheses have been retained, until $p_{(1)}$ is compared to $\alpha/C$.

In summary, six procedures were selected to investigate the viability of conducting tetrad contrasts in mixed designs. Since no information currently exists comparing the behaviour of a test statistic employing a pooled error term [equation (2)] versus one which is based on a nonpooled error term [equation (3)] under violations of multisample sphericity, both were utilized in the current study. The pooled and nonpooled statistics were considered in combination with either a Scheffe, Studentized maximum modulus, or Hochberg step-up Bonferroni CV. Although a Scheffe CV will be larger than either of the other two CVs, this method was selected in case the other two procedures could not limit the number of Type I errors under violations of the multisample sphericity assumption. Furthermore, Jaccard, Becker, and Wood (1984) found that psychologists often adopt Scheffe's method for pairwise comparisons of means. Thus, it was considered desirable to determine if this is an acceptable choice for conducting interaction contrasts. Finally, Hochberg's (1988) procedure was selected instead of the Dunn-Bonferroni method for the reasons previously enumerated.

## Monte Carlo Study

Although some results could be obtained analytically, the complexity and number of conditions to be considered necessitated a Monte Carlo study. The six procedures for testing interaction hypotheses were compared for a two-way mixed design containing a single between-subjects grouping factor and a single within-subjects factor. The number of levels of the between-subjects factor was held constant at three across all conditions. Total sample size was fixed at 30.

Seven variables were manipulated to investigate the behaviour of the selected statistical procedures. These were: (a) the number of levels of the within-subjects factor, (b) the sphericity pattern, (c) equality/inequality of the between-subjects covariance matrices, (d) group size equality/inequality, (e) the nature of the pairings of unequal covariance matrices and unequal group sizes, (f) population shape, and (h) the nature of the null hypothesis.

Keselman and Keselman (1990) found that the adjusted df univariate and multivariate procedures for interaction tests in mixed designs were more sensitive to violations of multisample sphericity as the number of levels of the within-subjects factor increased. Hence, the six procedures were studied when the number of within-subjects levels was set at four and eight.

Box's (1954) correction factor, $\epsilon$, was used to quantify the degree of departure from the assumption of sphericity. When sphericity is satisfied, $\epsilon$ attains a value of one. With increasing departures from sphericity, $\epsilon$ decreases in value to a minimum of $1/(K - 1)$. Without loss of generality, a covariance matrix formed by pooling across levels of the between-subjects factor contained element values of ten and five on the diagonal and off-diagonal respectively when sphericity was satisfied ($\epsilon = 1.00$). Matrices with $\epsilon$ values of .75 and .40 were chosen to represent nonspherical conditions. The elements of these pooled covariance matrices were chosen such that the average variance and covariance were equal to ten and five respectively, in order to achieve comparability across the simulation conditions. The pooled matrices for the $K = 4$ and $K = 8$ conditions are contained in Tables 1 and 2 respectively.

---

Insert Tables 1 and 2 about here

---

The effects of heterogeneity of between-subjects level covariance matrices was investigated by creating two sets of matrices. For one set, a given element in a particular between-subjects level matrix was equal to the corresponding element in each of the other two matrices, so that the elements of the matrices were in a 1:1:1 ratio. For the second set, corresponding elements in the between-subjects level covariance matrices were not equal to one another. Each element in the $j = 2$ covariance matrix was three times that of the $j = 1$ matrix and each element in the $j = 3$ matrix was five times that of the $j = 1$ matrix. Thus the elements in these between-subjects covariance matrices were in a 1:3:5 ratio.

The six procedures were investigated when the number of observations per between-subjects level were equal or unequal. When sample sizes were equal, there were ten observations per group. Two cases of group size imbalance were considered: $n_j = 8, 10, 12$ and $n_j = 6, 10, 14$. The coefficient of group size variation is .163 for the former condition, and .327 for the latter.

For those conditions involving both unequal group sizes and unequal covariance matrices, both positive and negative pairings of these group sizes and covariance matrices were investigated. In the former case, the largest $n_j$ was associated with the covariance matrix containing the largest element values; while in the latter, the largest $n_j$ was associated with the covariance matrix containing the smallest element values.

To summarize, six pairings of covariance matrices and group sizes were investigated: (a) equal $n_j$; equal $\Sigma_j$, (b) equal $n_j$; unequal $\Sigma_j$, (c/c′) unequal $n_j$; unequal $\Sigma_j$ (positively paired), and (d,d′) unequal $n_j$; unequal $\Sigma_j$ (negatively paired). The c′/d′ conditions denotes the more disparate unequal group sizes case, while the c/d conditions designates the less disparate unequal group sizes case.

Micceri (1989) investigated the distributional properties of 440 educational and psychological data sets, and found that few could be characterized as following a normal distribution. Thus, it was deemed important to examine the operating characteristics of the selected procedures when the underlying population distribution was normal and nonnormal. For the normal distribution, pseudorandom observation vectors $X_{ij}^T = [X_{ij1}, X_{ij2}, \ldots, X_{ijK}]$ with mean vector $\mu_j^T = [\mu_{j1}, \mu_{j2}, \ldots, \mu_{jK}]$ and covariance matrix $\Sigma_j$ were generated using the International Mathematical and Statistical Library (IMSL) subroutine GGNSM (IMSL, 1989).

Sawilowsky and Blair (1992) investigated the robustness of Student's t-statistic, for both independent and correlated samples, using eight nonnormal distributions identified by Micceri (1989) as representative of those found in educational and psychological data. They found that the Type I error rates for the t-statistics were affected only under conditions of extreme skewness, that is when $\gamma_1 = 1.64$. Therefore, the nonnormal data for the current study were obtained from a $\chi^2$ distribution with 3 df, for which $\gamma_1 = 1.63$ and $\gamma_2 = 4.00$. The IMSL subroutine GGCHS (IMSL, 1989) was used to generate deviates following a univariate $\chi^2$ distribution, which were then standardized to have a mean of 0 and a variance of 1. The corresponding multivariate observations were obtained by a triangular decomposition of $\Sigma_j$ [often referred to as the Cholesky factorization or the square root method (Harman, 1976)], that is,

$$X_{ij} = \mu_j + L Z_{ij} \quad , \tag{5}$$

where L is a lower triangular matrix satisfying the equality $\Sigma_j = LL^T$ and $Z_{ij}$ is an K x 1 vector of $\chi^2$ variates.

Finally, empirical familywise Type I error rates for the six procedures were obtained under a complete null hypothesis, when all of the $\mu_{jk}$'s were equal, and under a partial null hypothesis, when not all $\mu_{jk}$'s were equal. The familywise error rate was defined as the probability that at least one tetrad contrast was statistically significant when the corresponding population contrast was null. Seaman, Levin, and Serlin (1991) investigated the FWRs for a number of independent sample MCPs procedures, and found that the error rates were generally lower under a partial null than under a complete null hypothesis. Keselman (1993) reported similar findings for repeated measures mean comparisons in a mixed design. Since a researcher can never know the nature of the null hypothesis under investigation for a given set of data, it is advisable to select a procedure which can maintain the FWR close to its upper bound across all population mean configurations.

Five thousand replications of each condition were performed using a .05 significance level.

## Results

In discussing the results of the Monte Carlo study, Bradley's (1978) liberal criterion will be used to help identify robust tetrad contrast procedures. According to this criterion, a test may be considered robust if its empirical Type I error rate ($\hat{\alpha}$) falls within the range $.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$. Hence, for the .05 significance level selected for this study, a robust procedure is defined as one having an empirical familywise Type I error rate between .025 and .075.

The Type I error rates for the six procedures under a complete null hypothesis are reported in Tables 3 through 6. Those values which do not satisfy Bradley's (1978) criterion are marked with an asterisk (*). Herein, the Scheffe, Hochberg, and Studentized maximum modulus pooled and nonpooled procedures will respectively be referred · as the SP, SNP, HP, HNP, MP, and MNP procedures.

---

Insert Tables 3 and 4 about here

---

Tables 3 and 4 contain the results for the six procedures when the number of levels of the within-subjects factor was set at four. The results for the normal data found in Table 3, reveal that those procedures which utilized a test statistic based on a pooled estimated of error variance for a contrast (i.e. SP, HP, and MP) could not control the rate of Type I errors under violations of the multisample sphericity assumption. This finding holds for balanced (conditions a and b), as well as unbalanced designs (conditions c, c′, d, and d′), with the discrepancy between the nominal and empirical values being greatest for unbalanced designs. Both the HP and MP procedures were very liberal under extreme departures from sphericity ($\epsilon$ = .40), attaining Type I error rates as high as .24.

The error rates for those procedures which employed a nonpooled estimate of error variance in estimating the standard error of a contrast (i.e. SNP, HNP, and MNP) were never liberal. The empirical values for the

Studentized maximum modulus procedure were consistently larger than those obtained for the Hochberg procedure. However, both procedures occasionally resulted in conservative values, most notably when unequal covariance matrices were positively paired with unequal sample sizes (conditions c and c'). On the other hand the error rates for the Scheffe nonpooled procedure, with few exceptions, were conservative.

The results for the nonnormal data, contained in Table 4, reveal generally lower Type I error rates for all six procedures, as compared to the values obtained for normal data when K = 4. The pooled procedures produced similar results for nonnormal and normal data, and hence will not be considered further. In terms of the nonpooled procedures, the SNP values were conservative for all 18 of the conditions investigated. The HNP and MNP procedures provided Type I error control for $\epsilon > .40$. Under extreme departures from sphericity ($\epsilon = .40$) both procedures were largely conservative, only exceeding the lower bound of Bradley's (1978) criterion for negative pairings of sample sizes and covariance matrices. The minimum values attained by the HNP and MNP procedures under nonnormality were .0132 and .0158 respectively.

Insert Tables 5 and 6 about here

The K = 8 results for the six tetrad contrast procedures are contained in Tables 5 and 6. Consistent with the Table 3 results, none of the pooled procedures could control the Type I error rate under violations of sphericity the data was sampled from a normal distribution. Moreover, the empirical values were more extreme than those obtained when there were four levels of the repeated measures factor.

The SNP procedure was very conservative under all of the conditions investigated for normal data, with a mean FWR of .0045. For the most part, the HNP and MNP procedures provided good Type I error control when the repeated measures factor had eight levels. However, the latter was liberal for all values of $\epsilon$ when the more disparate group sizes were negatively paired with unequal covariance matrices (condition d′), attaining a maximum value of .1056. The empirical FWR for the HNP procedure was only slightly greater than the upper bound of Bradley's (1978) criterion when $\epsilon = 1.0$ for this same negative pairing condition (i.e. $\hat{\alpha} = .0756$).

The Type I error rates for the six procedures were generally lower for nonnormal data than for normal data when K = 8, consistent with the findings for K = 4. The MNP procedure was slightly liberal for negative pairings of the more disparate group sizes and unequal covariance matrices when $\epsilon < 1.0$, attaining a maximum value of .0810, but otherwise provided good FWR control. The HNP procedure was slightly conservative for all values of $\epsilon$ when the design was balanced and covariance matrices were equal, and for positive pairings of unequal group sizes and unequal covariance matrices. The minimum HNP value obtained was .0200.

By comparing the results found in Tables 3 and 4 to those in Tables 5 and 6 it is apparent that the familywise Type I error rates for both the HNP and MNP procedures increased as the number of levels of the within-subjects factor increased from four to eight. For normal data, the average values were .0362 and .0419 respectively for the HNP and MNP procedures when K = 4, and rose to .0417 and .0571 for K = 8. For the nonnormal data, the corresponding values were .0300 and .0353 when K = 4, and .0315 and .0436 for K = 8.

The data obtained for the six tetrad contrast procedures under a partial null hypothesis has not been tabled, since the empirical Type I error rates for the Hochberg and Studentized maximum modulus nonpooled procedures were consistently lower than the .05 level of significance. As expected, and consistent with previous findings for partial null hypotheses (Keselman, 1993, Seaman et al., 1991), the error rates were generally either less than the lower bound of Bradley's (1978) criterion, or approached it in value. However, in contrast with the findings for the complete null hypothesis, the HNP values were marginally larger than the MNP values across several conditions, for both normal and nonnormal data when K = 4. This pattern was not evident when K = 8, for which the Studentized maximum modulus procedure always produced larger Type I error rates than the Hochberg procedure.

## Discussion

As anticipated, those procedures which employed a pooled estimate of error variance could not control the FWR to $\alpha$ under departures from multisample sphericity, particularly when the design was unbalanced. As well, the procedure which relied on a nonpooled test statistic in combination with a Scheffe critical value was predictably conservative. For the most part, the nonpooled procedures which utilized either a Hochberg step-up Bonferroni or Studentized maximum modulus critical value provided good control of the familywise Type I error rate under violations of multisample sphericity in unbalanced designs, even when the data came from a nonnormal distribution. However, both of these procedures became quite conservative under the combined effects of nonnormality, large departures from sphericity, and small values of K. This was particularly evident when the configuration of the population means was such that not all of the means were equal. Nonetheless, under the majority of the conditions investigated these procedures provided reasonably acceptable rates of familywise Type I error control, thereby providing researchers with robust alternatives to omnibus tests of interaction effects in mixed repeated measures designs.

References

Belli, G. M. (1988, April). Type I error rates of MANOVA of repeated measures under group heterogeneity in unbalanced designs. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Boik, R. J. (1975). Interactions in the analysis of variance: A procedure for interpretation and a Monte Carlo comparison of univariate and multivariate methods for repeated measures designs. Dissertation Abstracts International, 36, 2908B. (University Microfilms No. 75-27,837)

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and correlation between errors in the two-way classification. Annals of Mathematical Statistics, 25, 290-302.

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Dunn, O. J. (1961). Multiple comparisons among means. Journal of the American Statistical Association, 56, 52-64.

Gabriel, K. R., Putter, J., & Wax, Y. (1973). Simultaneous confidence intervals for product-type interaction contrasts. Journal of the Royal Statistical Society, Series B, 35, 234-244.

Green, S., & Barcikowski, R. S. (1992, April). Sphericity in the repeated measures univariate mixed-model design. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Harman, H. H. (1967). Modern factor analysis (2nd ed.). Chicago: University of Chicago Press.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800-802.

Hochberg, Y., & Tamhane, A. C. (1987). Multiple comparison procedures. New York: Wiley.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. Psychometrika, 43, 161-175.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F distributions. Journal of the American Statistical Association, 65, 1582-1589.

IMSL. (1989). IMSL library (9th ed). Houston, TX: Author.

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise multiple comparison procedures: A review. Psychological Bulletin, 96, 589-596.

Keselman, H. J. (1993). Stepwise multiple comparisons of repeated measures means under violations of multisample sphericity. In F. M. Hoppe (Ed.), Multiple comparisons, selection, and applications in biometry (pp. 167-186). New York: Marcel Dekker.

Keselman, H. J., & Keselman, J. C. (1990). Analyzing unbalanced repeated measures designs. British Journal of Mathematical and Statistical Psychology, 43, 265-282.

Keselman, H. J., Keselman, J. C., & Shaffer, J. P. (1991). Multiple pairwise comparisons of repeated measures means under violation of multisample sphericity. Psychological Bulletin, 110, 162-170.

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of type IV errors. American Education Research Journal, 7, 397-421.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 154-166.

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

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. Psychological Bulletin, 111, 352-360.

Scheffe, H. (1953). A method for judging all contrasts in analysis of variance. Biometrika, 40, 87-104.

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.

Table 1. Pooled covariance matrices for K = 4

| | | | | |
|---|---|---|---|---|
| | 18.0 | 8.0 | 6.0 | 4.0 |
| | | 8.0 | 5.0 | 4.0 |
| $\epsilon = .75$ | | | 7.0 | 3.0 |
| | | | | 7.0 |
| | 23.8 | 11.9 | 6.4 | 0.9 |
| | | 9.5 | 5.7 | 2.6 |
| $\epsilon = .40$ | | | 3.9 | 2.5 |
| | | | | 2.8 |

Table 2. Pooled covariance matrices for K = 8

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 18.0 | 8.0 | 7.0 | 7.0 | 6.0 | 5.0 | 5.0 | 5.0 |
| | 12.0 | 8.0 | 7.0 | 6.0 | 5.0 | 5.0 | 2.0 |
| | | 10.0 | 6.0 | 6.0 | 5.0 | 5.0 | 2.0 |
| $\epsilon$ = .75 | | | 10.0 | 5.0 | 5.0 | 4.0 | 4.0 |
| | | | | 9.0 | 5.0 | 5.0 | 3.0 |
| | | | | | 8.0 | 4.0 | 4.0 |
| | | | | | | 7.0 | 1.0 |
| | | | | | | | 6.0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 28.8 | 12.8 | 10.1 | 9.8 | 8.3 | 7.3 | 6.0 | 1.8 |
| | 17.4 | 8.1 | 7.4 | 6.9 | 4.1 | 3.4 | -1.0 |
| | | 9.9 | 7.7 | 6.5 | 5.7 | 3.4 | 1.4 |
| $\epsilon$ = .40 | | | 8.3 | 5.6 | 4.3 | 3.9 | 2.4 |
| | | | | 5.6 | 4.4 | 2.6 | 1.4 |
| | | | | | 4.3 | 2.4 | 1.4 |
| | | | | | | 3.2 | 1.9 |
| | | | | | | | 2.5 |

Table 3. Empirical Percentages of Type I Error Under a Complete Null Hypothesis for Six Interaction Contrast Procedures when K = 4 (Normal)

| $\epsilon$ | | SP | SNP | HP | HNP | MP | MNP |
|---|---|---|---|---|---|---|---|
| | a | .0076* | .0130* | .0368 | .0380 | .0382 | .0432 |
| | b | .0140* | .0132* | .0544 | .0386 | .0566 | .0450 |
| | c | .0070* | .0112* | .0306 | .0400 | .0310 | .0464 |
| 1.0 | d | .0276 | .0188* | .0908* | .0466 | .0940* | .0572 |
| | c′ | .0034* | .0112* | .0230* | .0366 | .0246* | .0428 |
| | d′ | .0546 | .0298 | .1506* | .0546 | .1544* | .0632 |
| | a | .0242* | .0112* | .0710 | .0380 | .0720 | .0422 |
| | b | .0362 | .0158* | .0942* | .0454 | .0968* | .0532 |
| | c | .0182* | .0120* | .0580 | .0342 | .0594 | .0408 |
| .75 | d | .0576 | .0194* | .1278* | .0462 | .1316* | .0522 |
| | c′ | .0120* | .0110* | .0364 | .0364 | .0372 | .0416 |
| | d′ | .0730 | .0308 | .1662* | .0524 | .1704* | .0628 |
| | a | .0702 | .0068* | .1506* | .0220* | .1540* | .0250 |
| | b | .0734 | .0058* | .1460* | .0194* | .1470* | .0226* |
| | c | .0540 | .0072* | .1098* | .0228* | .1124* | .0252 |
| .40 | d | .1110* | .0138* | .1900* | .0280 | .1924* | .0308 |
| | c′ | .0402 | .0064* | .0828* | .0202* | .0850* | .0216* |
| | d′ | .1392* | .0176* | .2370* | .0318 | .2408* | .0382 |

NOTE: SP = Scheffe, pooled error; SNP = Scheffe, nonpooled error; HP = Hochberg, pooled error; HNP = Hochberg, nonpooled error; MP = Studentized maximum modulus, pooled error; MNP = Studentized maximum modulus, nonpooled error; a = pairings of equal covariance matrices and equal sample sizes; b = pairings of unequal covariance matrices and equal sample sizes; c/c′ = positive pairings of unequal covariance matrices and unequal group sizes [c: $n_j$ = 8, 10, 12; c′: $n_j$ = 6, 10, 14]; d/d′ = negative pairings of unequal covariance matrices and unequal group sizes [d: $n_j$ = 12, 10, 8; d′: $n_j$ = 14, 10, 6]

Table 4. Empirical Percentages of Type I Error Under a Complete Null Hypothesis for Six Interaction Contrast Procedures when K = 4 (Nonnormal $\chi^2$)

| $\epsilon$ | | SP | SNP | HP | HNP | MP | MNP |
|---|---|---|---|---|---|---|---|
| | a | .0064* | .0084* | .0410 | .0286 | .0424 | .0324 |
| | b | .0126* | .0098* | .0518 | .0316 | .0552 | .0368 |
| | c | .0060* | .0084* | .0314 | .0286 | .0330 | .0334 |
| 1.0 | d | .0246* | .0174* | .0826* | .0412 | .0848* | .0494 |
| | c′ | .0026* | .0074* | .0152* | .0288 | .0156* | .0336 |
| | d′ | .0522 | .0240* | .1404* | .0474 | .1442* | .0586 |
| | a | .0214* | .0072* | .0714 | .0264 | .0740 | .0326 |
| | b | .0284 | .0092* | .0786* | .0274 | .0810* | .0326 |
| | c | .0168* | .0088* | .0582 | .0310 | .0598 | .0354 |
| .75 | d | .0480 | .0140* | .1204* | .0336 | .1240* | .0416 |
| | c′ | .0100* | .0082* | .0348 | .0286 | .0360 | .0342 |
| | d′ | .0770* | .0200* | .1708* | .0474 | .1744* | .0550 |
| | a | .0610 | .0050* | .1310* | .0132* | .1344* | .0158* |
| | b | .0696 | .0068* | .1418* | .0202* | .1450* | .0234* |
| | c | .0516 | .0058* | .1084* | .0184* | .1112* | .0208* |
| .40 | d | .1060* | .0136* | .1924* | .0320 | .1958* | .0364 |
| | c′ | .0394 | .0066* | .0826* | .0186* | .0846* | .0208* |
| | d′ | .1378* | .0200* | .2278* | .0362 | .2314* | .0426 |

NOTE: See Table 3 notes

Table 5. Empirical Percentages of Type I Error Under a Complete Null Hypothesis for Six Interaction Contrast Procedures when K = 8 (Normal)

| $\epsilon$ | | SP | SNP | HP | HNP | MP | MNP |
|---|---|---|---|---|---|---|---|
| | a | .0000* | .0016* | .0362 | .0372 | .0378 | .0498 |
| | b | .0008* | .0044* | .0680 | .0426 | .0702 | .0554 |
| | c | .0000* | .0018* | .0380 | .0364 | .0390 | .0504 |
| 1.0 | d | .0026* | .0030* | .1154* | .0474 | .1180* | .0722 |
| | c′ | .0002* | .0020* | .0200* | .0352 | .0206* | .0486 |
| | d′ | .0074* | .0136* | .2078* | .0756* | .2110* | .1056* |
| | a | .0016* | .0010* | .0656 | .0354 | .0670 | .0476 |
| | b | .0042* | .0030* | .1014* | .0458 | .1028* | .0610 |
| | c | .0010* | .0014* | .0536 | .0328 | .0546 | .0468 |
| .75 | d | .0076* | .0054* | .1450* | .0470 | .1478* | .0668 |
| | c′ | .0004* | .0012* | .0370 | .0362 | .0378 | .0476 |
| | d′ | .0228* | .0120* | .2132* | .0688 | .2180* | .0974* |
| | a | .0160* | .0012* | .1678* | .0290 | .1702* | .0364 |
| | b | .0278 | .0008* | .1754* | .0324 | .1778* | .0432 |
| | c | .0130* | .0014* | .1230* | .0278 | .1246* | .0368 |
| .40 | d | .0418 | .0040* | .2346* | .0392 | .2366* | .0528 |
| | c′ | .0106* | .0010* | .1024* | .0270 | .1044* | .0346 |
| | d′ | .0706 | .0106* | .3076* | .0550 | .3114* | .0754* |

NOTE: See Table 3 notes

Table 6. Empirical Percentages of Type I Error under a Complete Null Hypothesis for Six Interaction Contrast Procedures when K = 8 (Nonnormal $\chi^2$)

| $\epsilon$ | | SP | SNP | HP | HNP | MP | MNP |
|---|---|---|---|---|---|---|---|
| | a | .0006* | .0004* | .0384 | .0200* | .0398 | .0288 |
| | b | .0010* | .0008* | .0688 | .0272 | .0698 | .0366 |
| | c | .0010* | .0008* | .0352 | .0236* | .0364 | .0376 |
| 1.0 | d | .0040* | .0018* | .1198* | .0416 | .1232* | .0570 |
| | c′ | .0004* | .0006* | .0242* | .0214* | .0248* | .0310 |
| | d′ | .0138* | .0070* | .2078* | .0500 | .2138* | .0730 |
| | a | .0018* | .0006* | .0764* | .0216* | .0774* | .0288 |
| | b | .0024* | .0008* | .0954* | .0292 | .0978* | .0398 |
| | c | .0016* | .0002* | .0566 | .0288 | .0582 | .0396 |
| .75 | d | .0074* | .0036* | .1442* | .0396 | .1468* | .0542 |
| | c′ | .0008* | .0012* | .0342 | .0212* | .0350 | .0298 |
| | d′ | .0198* | .0070* | .2284* | .0544 | .2326* | .0776* |
| | a | .0180* | .0010* | .1652* | .0206* | .1678* | .0268 |
| | b | .0192* | .0022* | .1764* | .0286 | .1786* | .0388 |
| | c | .0132* | .0018* | .1328* | .0228* | .1352* | .0294 |
| .40 | d | .0362 | .0042* | .2364* | .0398 | .2396* | .0508 |
| | c′ | .0054* | .0014* | .0914* | .0168* | .0928* | .0250 |
| | d′ | .0662 | .0106* | .3082* | .0606 | .3118* | .0810* |

NOTE: See Table 3 notes