DOCUMENT RESUME

ED 282 917 TM 870 333

AUTHOR Robey, Randall R.; Barcikowski, Robert S.
TITLE Sphericity Tests and Repeated Measures Data.
PUB DATE Apr 87
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Error of Measurement; Hypothesis Testing; *Mathematical Models; *Monte Carlo Methods; *Multivariate Analysis; Statistical Studies
IDENTIFIERS *Repeated Measures Design; *Sphericity Tests; Univariate Analysis

ABSTRACT

The mixed model analysis of variance assumes a mathematical property known as sphericity. Several preliminary tests have been proposed to detect departures from the sphericity assumption. The logic of the preliminary testing procedure is to conduct the mixed model analysis of variance if the preliminary test suggests that the sphericity assumption is tenable, or alternatively, to conduct another analysis of variance which does not assume sphericity. This paper examines the value of basing the analysis strategy upon the outcome of the preliminary test on the sphericity assumption. The examination is divided into four parts: (1) sphericity and its relationship to the mixed model analysis are reviewed; (2) five sphericity tests are reviewed, and the appearance of these tests in three popular statistical packages is described; (3) the performances of these tests in a Monte Carlo Type 1 error study are reported; and (4) the danger of depending upon a preliminary test to make analysis decisions is emphasized using an example data set. (BAE)

ED282917

Sphericity Tests and Repeated Measures Data

Randall R. Robey

Southern Illinois University

Robert S. Barcikowski

Ohio University

A paper presented before the annual meeting of the American Educational Research Association, Washington, D. C., April 20-24, 1987.

Sphericity Tests and Repeated Measures Data

Introduction

Behavioral scientists often use one or another variation of the repeated measures research design to make decisions concerning behavioral and psychological data (Edgington, 1974; Jennings and Wood, 1976; Lana and Lubin, 1963; and Robey, 1985). In an exploratory investigation where specific a priori contrasts cannot be reasonably formulated, a researcher must depend upon an omnibus F statistic for decisions on the presence or absence of treatment effects. One of the analysis alternatives in this situation is the mixed model analysis of variance (Scheffé, 1959). This particular analysis assumes, among other things, a mathematical property known as sphericity (Huynh and Feldt, 1970). Several preliminary tests to detect departures from the sphericity assumption have been proposed (e.g., Mauchly, 1940), and advocated (e.g., Huyhn and Mandeville, 1979). The logic of the preliminary testing procedure is to conduct the mixed model analysis of variance if the preliminary test suggests that the sphericity assumption is tenable, or alternatively, to conduct another analysis of variance, (e.g., a multivariate analysis of variance) which does not assume sphericity (Huynh and Mandeville, 1979; Hertzog and Rovine, 1985; Thomas, 1983).

In this paper, we examine the value of basing the analysis strategy upon the outcome of a preliminary test on the sphericity assumption. Our examination proceeds in four parts. In the first section, we review sphericity and its relationship to the mixed model analysis. The second section contains a review of five sphericity tests and a description of the appearance of these tests in three popular statistical packages. The third section reports the performances of these tests in a Monte Carlo Type I error study. In the fourth section, an example data set is used to emphasize the danger of depending upon a preliminary test to make analysis decisions.

The Sphericity Assumption

The following discussion of sphericity proceeds in two sections. The first section relates to univariate sphericity, where one observation is collected on each unit at each occasion, while the second section addresses multivariate sphericity, where multiple measures are collected on each unit at each occasion.

Univariate Sphericity

Huynh and Feldt (1970) and Rouanet and Lépine (1970) showed that sphericity is necessary and sufficient for the ratio of mixed model variances to be distributed as F. Huynh and Feldt (1970) referred to this condition as sphericity while Rouanet and Lépine (1970) used the term circularity.

In the single group design, sphericity is held if, and only if,: $C\Sigma C' = \sigma^2 I$ where C is a k-1 x k orthonormal contrast matrix, Σ is the population variance-covariance matrix, and I is an identity matrix of order k-1. The element σ^2 is a scalar > 0 which represents the common population variance on each of the orthonormal contrasts in C (Boik, 1981). It should be noted that the value of σ^2 is the common value of the k-1 eigenvalues of $C\Sigma C'$.

For the split-plot design, the assumption is written as $C\Sigma^{*}C' = \sigma^2 I$, where Σ^{*} is the pooled variance-covariance matrix for all g groups. Pooling is indicated only when all g Σ matrices are equivalent (Huynh and Feldt, 1970).

Box (1954) and Imhof (1962) showed that when the positive non-zero eigenvalues of $C\Sigma C'$ are all equal, the single group mixed model test is distributed exactly as bF[h,h(n-1)]. Here, b is a measure of sample size heterogeneity which equals unity in the single group repeated measures design with no missing data. The quantity, h, is a ratio of $C\Sigma C'$ eigenvalues (λ) written as

$$h = \frac{\left(\sum_{i=1}^{k-1} \lambda_i \right)^2}{\sum_{i=1}^{k-1} \lambda_i^2} \quad . \qquad (1)$$

When all the k-1 eigenvalues are equal, h = k-1, and the distribution of F on k-1 and (k-1)(n-1) degrees of freedom follows.

Under a departure from sphericity (i.e., when the eigenvalues of $C\Sigma C'$ do not share a common value), a multiplicative correction factor for the usual k-1 and (k-1)(n-1) degrees of freedom to obtain an approximate distribution of F on h and h(n-1) degrees of freedom follows as

$$\epsilon = \frac{h}{k-1} = \frac{\left(\sum_{i=1}^{k-1} \lambda_i\right)^2}{(k-1)\sum_{i=1}^{k-1} \lambda_i^2} . \qquad (2)$$

Box (1954) showed that the mixed model F is approximately distributed as $F[(k-1)\epsilon, (k-1)(n-1)\epsilon]$. Note that if all the eigenvalues are equal, ϵ equals 1, and is otherwise a fraction. Geisser and Greenhouse (1958) showed the lower bound of ϵ to be 1/(k-1). However, Imhof (1962) recognized that with $C\Sigma C'$ being positive definite, ϵ only approaches 1/(k-1). Huynh (1978) showed that the mixed model test is not robust with respect to even slight departures from sphericity.

<u>Multivariate Sphericity</u>.

The multivariate mixed model analysis assumes not only sphericity of the variance-covariance matrix associated with each of the dependent variables, but that the variance-covariance structure among the dependent variables, across occasions, is the same (Timm, 1980). That is,

$$V_1 = V_2 = V_3 \ldots = V_i = V,$$

where V_i (i = 1,2,...k) is the variance-covariance matrix for the dependent variables at the i^{th} occasion, and where V is the common variance-covariance matrix of the dependent variables for all k occasions.

Testing the Sphericity Assumption

In the following paragraphs we review several tests on the sphericity assumption. A discussion of the univariate tests precedes a discussion of the multivariate test.

Univariate Tests

Mauchly's Test

Mauchly (1940) derived a test that can help a researcher determine if a sample of variates were selected from a population where the variances are all equal, and the covariances are all zero (i.e., $\sigma^2 I$). It follows that Mauchly's likelihood ratio criterion is applied to the CSC' matrix, where S is the sample estimate of Σ. Mauchly's test is defined by Huynh and Feldt (1970) as

$$w = |CSC'| \;/\; [tr(CSC') \;/\; (k-1)]^{k-1}. \qquad (3)$$

With the values

$$d = 1 - (2(k-1)^2 + (k-1) + 2) \;/\; [6(k-1)(n-1)] \qquad (4)$$

and

$$v = [(k-1)k/2] - 1 \;, \qquad (5)$$

an asymptotic test is given by -(n-1)ln(w)d which is approximately distributed as Chi-square on v degrees of freedom. Rejection of the null hypothesis indicates that the mixed model test is not valid (Rogan, Keselman, and Mendoza, 1979), and as a result, the data require an alternate test statistic, e.g., the multivariate model F or an adjusted mixed model F. Anderson (1958) described the exact distribution of w when k = 3. In 1967, Consul presented tabled critical values of w for k at 4, 5, and 7. Nargarsenker and Pillai (1973) presented tabled critical values for k-1 = 4(1)10 at the .05 and .01 alpha levels. Mathai and Rathie (1970), and Pillai and Nargarsenker (1971) among others, have described the exact distribution of the criterion w for any k. The non-null distribution of w was derived by Nargarsenker (1976).

Rogan, Keselman and Mendoza (1979) and Keselman, Rogan, Mendoza and Breen (1980) examined Mauchly's test and found it to be very sensitive to departures from normality, and to even slight departures from its null hypothesis. They concluded that rejection of the null hypothesis is so likely that one would almost always conclude that the sphericity assumption was not reasonable to maintain. Muirhead and Waternaux (1980) also found the sphericity criterion to be sensitive to departures from normality. It should be noted that Muirhead and Waternaux (1980) examined a similar criterion, V, which equals $w^{n/2}$. The value of V was evaluated as -2ln(V) which is approximately distibuted as Chi-square on [(k-1)k/2] -1 degrees of freedom.

Huyhn and Mandeville (1979) examined Mauchly's test and, among other things, their results indicated that the test performed acceptably under departures from normality in the direction of light tailed distributions. Huynh and Mandeville argued that this is the likely form of non-normality to be encountered in practice.

The criterion w is reported by PROC GLM of SAS (SAS Institute, Inc., 1985) when the PRINTE option of the REPEATED command is specified. Although the probability for w is not reported, GLM does report the Chi-square approximate with its degrees of freedom and a p value. The 2V program in BMDP (Dixon, 1983) reports a similar, but different, Chi-square approximate when the SYMMETRY option appears in the DESIGN paragraph. In a personal communication (BMDP Statistical Software; August 28, 1986), BMDP indicated that the 2V test is given in Equation 22 of Anderson (1958, p. 263). However, examination of that equation does not suggest a distinction from $-(n-1)\ln(w)d$.

Bock's Tests

Bock (1975) described an alternate two-stage method for testing the null hypothesis that $C\Sigma C' = \sigma^2 I$. In the first stage, Bock recommended using a test proposed by Bartlett (1950) which tests the null hypothesis that $C\Sigma C'$ is a diagonal matrix. The test is given by

$$-[(n-1) - (2k+5)/6]\ln|R| \qquad (6)$$

which is approximately distributed as Chi-square on k(k-1)/2 degrees of freedom under the null hypothesis. The test is applied to the sample correlation matrix derived from CSC'.

The second stage of the procedure described by Bock (1975) uses a test for homogeneity of variance proposed by H. O. Hartley and described by Winer (1971, p. 206). The test statistic, F_{max}, is a ratio of the largest of a set of variances, over the smallest of the same set of variances. Since the test is applied to the CSC' matrix, the null hypothesis concerns equivalence among the variances of the k-1 orthonormally transformed variates. Under the null hypothesis, the test is approximately distributed as F with k-1 and n-1 degrees of freedom. If both the Bartlett (1950) test, and the Hartley (Winer, 1971, p. 206) test are not rejected, Bock suggested that sphericity can be assumed.

Norusis (1985) recognized that Bock's first test statistic varies as the composition of the orthonormal coefficients in the matrix C vary. The reason for this ambiguity is explained by the fact that as the coefficients in C vary, the structure of CSC' varies. The mixed model F is invariant to this type of variation, but the Chi-square approximate in Equation 6 is not. The program described by Norusis (1985) uses a single contrast matrix to calculate w and F. As a result, w varies when the orthonormal coefficients are redefined in subsequent runs. To eliminate this ambiguity, it seems reasonable to select the orthonormal coefficients in C parsimoniously such that CSC' is diagonal. This is accomplished by post multiplying C by the matrix of eigenvectors for CSC'. The result is to reduce the structure of CSC' to its canonical form with eigenvalues along the diagonal and zeros elsewhere. This step not only reduces the ambiguity of the results of Bock's first test, it precludes the application of the test all together by insuring that the null hypothesis is true. When CSC' is diagonal, the F_{max} test becomes a ratio of the largest eigenvalue of CSC' over the smallest eigenvalue of CSC'.

The tests used in Bock's two step procedure can be obtained in the MANOVA program of SPSS-X (SPSS Inc, 1986) by specifying ERROR(COR) in the PRINT command.

<u>The John-Sugiura-Nagao Test</u>

John (1971), Sugiura (1972), and Nagao (1973) each derived a test for sphericity based upon the quantity U which equals $1/\hat{h}$, where $\hat{h}$ is obtained by substituting the eigenvalues of CSC' into Equation 1. The value U is evaluated using the statistic T which equals [(k-1)U - 1] / (k-1). Tables of critical values for T are found in John (1976).

Grieve (1984) compared the w statistic to the John-Sugiura-Nagao sphericity test using a Monte Carlo power analysis. Grieve found that the eigenvalue structure of Σ can favor one or the other test. However, Grieve's results suggested that the T statistic does not suffer as badly as does Mauchly's test under unfavorable conditions.

Neither BMDP, SAS nor SPSS-X report the T test.

Grieve's Test

Grieve (1984) also presented a sphericity test based upon the relationship of h and ϵ (i.e., $\epsilon = h/(k-1)$). Grieve defined a test on the sphericity assumption where the value of ϵ is used as the criterion which is compared to a critical value. Further, Grieve presented tables of critical values for ϵ. As Grieve contends, this test is the most intuitively appealing of all of the tests on the sphericity assumption since it uses the familiar index of sphericity as the test criterion.

Neither BMDP, SAS nor SPSS-X report the Grieve epsilon test.

Testing Multivariate Sphericity

Thomas (1983) extended Mauchly's test of sphericity for $p = 1$ to the any p case. Thomas derived the test as follows. Let $D = (1/p)E$, where E is the error sums of squares and cross products matrix for the multivariate mixed model. Further, let t_i be the natural log of the i^{th} eigenvalue of the doubly multivariate error sums of squares and cross products matrix for the doubly multivariate model, and let u_i be the natural log of the i^{th} eigenvalue of D. Then Thomas's extension of Mauchly's test is given by

$$ng[(k-1)\sum_{i=1}^{p}u_i - \sum_{i=1}^{p(k-1)}t_i], \qquad (7)$$

where g is the number of groups. This value is evaluated as Chi-square on

$$\frac{p(k-2)\ [p(k-1) + p + 1]}{2} \qquad (8)$$

degrees of freedom.

In a Monte Carlo robustness study, Robey (1985) evaluated the Type I error rate of the approximate Chi-square. He found that the test was unacceptably liberal at all levels for sample sizes likely to be encountered in the behavioral sciences. Further study by Robey (1985) revealed that the test required several hundred observations for the actual alpha level to approximate the nominal level.

A Monte Carlo Type I Error Analysis of the Univariate Tests

Although Huynh and Mandeville (1979) and Keselman et al. (1980) have examined the Type I error rate of the w criterion, to date, the remaining univariate sphericity tests have not been similarily investigated. To address this problem, a Monte Carlo robustness study was conducted on five univariate sphericity tests using multivariate normal data where $\epsilon = 1$ (i.e., the assumptions were maintained and the null hypothesis was true).

The w criterion was observed in a partial replication of the above two studies. The Chi-square approximate for w was also observed since that form of the test is more likely to be reported by practitioners. In addition, the Type I error rates of the eigenvalue F_{max}, the T test and the epsilon test were examined. The Bock Chi-square was not analyzed since it is irrelevant when considering eigenvalues.

The Monte Carlo analysis reported here was fashioned after Robey (1985) and Robey and Barcikowski (1986). The number of occasions in the design was varied at 3, 5, 7 and 10. The number of observations in the design was varied at (k-1) + 3, (k-1) + 10, (k-1) + 20, and (k-1) + 30. These sample sizes were selected to agree with those described by Davidson (1972).

The Monte Carlo dependent variable was the proportion of 6000 calculated test statistics which exceeded the critical value of a particular test. This proportion, actual alpha, was calculated for each combination of the independent variables (i.e., k and n). Decisions regarding the comparison of nominal and actual alpha levels were accomplished using the hypothesis testing approach described below. This manner of decision making satisfies the recent call for increased rigor in the design of Monte Carlo studies (Hauck and Anderson, 1984).

Data Generation and Analysis

A FORTRAN subroutine, GGNSM, from the International Mathematical and Statistical Libraries, Inc. (IMSL, 1982) was used to generate multivariate normal data for each of the variance-covariance matrices. GGNSM first generates multivariate normal vectors of random numbers, N(0,1). Then using Cholesky decomposition, the input variance-covariance matrix is decomposed to an upper triangular matrix, U, such that $UU' = \Sigma$. The N(0,1) vectors of data in some matrix, say Z, are then transformed to $N(0,\Sigma)$ through ZU'. The initial seeds for GGNSM were selected from a table of random numbers.

The analysis program was a double precision G level FORTRAN program executed on an IBM 4381 mainframe computer.

Statistical Hypotheses

The general form of the null and alternate hypotheses were

$$H_o: \quad P = \alpha \quad \text{and} \quad H_a: \quad P \neq \alpha \quad ,$$

where P represents a population proportion, and α equalled .01 and .05. Meaningful discrepancies between the nominal alpha levels and actual alpha levels were defined as departures of $\pm$ .005 from the nominal alpha of .01, and as departures of $\pm$ .025 from the nominal alpha of .05. The magnitude of these values were obtained using Bradley's (1978) robustness criterion of $\alpha \pm \alpha .5$. A two-tailed test for proportions described by Cohen (1977, p. 213) was used to analyze the results of the Monte Carlo problems. The *a priori* alpha level for all applications of the proportions test was set at .01. The desired minimal statistical power for all applications of the proportions test was set at .80. Following the method for establishing sample size described by Cohen (1977), it was determined that 5711 observations were needed for H_o: P = .01, and that 1085 observations were needed for H_o: P = .05. As a matter of convenience, 6000 observations were collected for each Monte Carlo problem. As a result, statistical power exceeded .80 in all of the analyses.

Results

The actual alpha levels for each of the sphericity tests are reported in Table 1 through Table 5. In Table 1, it can be seen that the criterion w demonstrated appropriate Type I error rates at each level of n for k = 3, 5 and 7. These results agree with the comme[illegible]ite results of Huynh and Mandeville (1979) and of Keselman et al. (1980). Unfortunately, critical values for k = 3 were not available to the authors.

The actual alpha levels found in Table 2 for the Chi-square approximation of w show an interesting pattern which is restricted to the smallest sample size (i.e., k-1+3). That is, for k at 7 and 10, the actual alpha levels become increasingly liberal. With sample sizes of k-1+10 and larger, the test demonstrated good Type I error control.

Notice that the actual alpha levels for k = 5 in the smallest sample size show a significant departure from nominal alpha at .05. This is due to the fact that the actual alpha levels are based upon 6000 observations. Since this actual alpha level (i.e., .066) is well within the critical interval of $\alpha \pm \alpha.5$ which was established *a priori*, this rejection, and others like it, will be ignored.

Table 3 contains the actual alpha levels for the F_{max} test. Recall that here F_{max} is the ratio of the largest eigenvalue of CSC' over the smallest eigenvalue of CSC'. It can be seen that the test becomes increasingly liberal as k increases. Further, the test becomes increasingly conservative as n increases. As a result, for larger n's and smaller k's, the test is excessively conservative, while for smaller n's and larger k's, the test is excessively liberal.

The results for the T test found in Table 4 are less clear. For the largest sample size (i.e., k-1+30), the actual alpha levels are all conservative and decrease as k increases. The same general trend can be seen in the results for n = k-1+20; however, at k = 3 and k = 5, the test is not conservative. Although the results for n = k-1+10 show acceptable Type I error rates for k = 3 and k = 5, the rates for k = 7 and k = 10 are both too high. In contrast, the results for n = k-1+3 are all unacceptably liberal.

The Type I error rates for the Grieve epsilon test found in Table 5 vary with sample size. In general, the test performed liberally under all levels of k in the smallest sample size. The results for n = k-1+10 represent a transition from the liberalness found at n = k-1+3 to what are generally adequate Type I error levels across all levels of k for the two largest sample sizes.

Discussion

These results suggest that the w criterion offers acceptable Type I error control which remains stable as the structure of the basic design varies. The Chi-square approximation of w also demonstrated acceptable and stable Type I error rates except where a small sample size (i.e., n exceeded k by a few) combined with larger k's (i.e., $k \geq 7$). The results for the Grieve epsilon test suggest that it provides reliable Type I error control only when n exceeds k by 20 or more.

Unfortunately, these results indicate that the Bock two-step procedure cannot be recommended under any circumstances. As shown earlier, the first step is not necessary when the CSC' is reduced to its canonical form, and is uninterpretable otherwise. For the second step, these Type I error results indicate that the interaction between k and n is sufficiently severe to make the F_{max} test useless as a decision making tool in an already difficult situation (i.e., to assume, or not to assume sphericity).

It should be kept in mind that these results were obtained using normally distributed variates. These results do not comment on the impact of non-normality upon the various sphericity tests.

The Problem with Sphericity Tests is ...

In this section we will demonstrate the potential problem with the practice of depending upon the multivariate repeated measures F when a preliminary test of the sphericity indicates that the assumption for the mixed model is not a tenable one. Consider the fi:cticious data set found in Table 6. The data set was constructed to represent an experimental situation where, following a no treatment observation prior to k_1, some intervention is introduced to a clinical sample prior to k_2 and is maintained through k_4. The last two measurement occasions are taken in a no treatment period in order to measure carry-over or extinction effects.

The means (i.e., 46.03, 45.85, 52.78, 54.0, 52.68 and 52.64) show a sustained elevation of scores which begins at k_3. The variances (i.e., 156.71, 250.78, 12.81, 10.08, 5.77 and 13.88) reflect an initial increase in the heterogeneity of the scores at k_2 with the onset of treatment. The remainder of the observations in the treatment period show a marked increase in homogeneity of the scores which is maintained through the final no treatment period. The eigenvalues of the CSC' matrix (i.e., 242.99, 108.10, 15.26, 8.25 and 7.43) yield an ϵ value of .4106.

The various sphericity tests, as well as the mixed model F corrected for $\hat{\epsilon}$ and the multivariate repeated measures F, are reported in Table 7. Each of the sphericity tests correctly indicates that the sphericity assumption is not reasonable to maintain. If it were decided to conduct only the multivariate analysis on the basis of one or more of these tests, the obvious treatment effect would be missed (F = 2.25; df = 5,15;

p = .1025). However, the corrected mixed model analysis does detect the treatment effect at the .05 level ($F = 3.57$; $df = 2.05, 39.01$; $p = .0365$). The explanation for this power differential can be found in Davidson (1972), Imhof (1962), Barcikowski and Robey (1984a) and in Robey (1985).

When deciding how best to analyze any given set of repeated meas[illegible] ata, a researcher must contend with the absence of crucial information since the structure of $C\Sigma C'$ cannot be estimated _a priori_. Without this information, the following two questions cannot be answered: 1.) Given a likely departure from sphericity, how severe is that departure? and 2.) Given a likely departure from sphericity, which analysis of variance will best be able to detect the treatment effects? This dilemma precludes an _a priori_ selection of the preferred analysis procedure.

Barcikowski and Robey (1984a, 1984b) advocated a solution to this problem based upon the fact that a departure from sphericity is very likely, and upon the fact that even slight departures from sphericity can bias the mixed model test. Their analysis strategy can be summarized in the following four points.

1. Following the procedure described by Robey and Barcikowski (1984), calculate the n neseccary to detect the smallest meaningful treatment effect with satisfactory Type I and Type II error rates.

2. Forego testing sphericity altogether, and concentrate on the value of $\hat{\epsilon}$ as a descriptive statistic which comments on the severity of the departure from sphericity.

3. Conduct the multivariate F and the mixed model F simultaneously.

   a. Routinely correct the mixed model degrees of freedom using one of the sample estimate correction factors (e.g., Greenhouse and Geisser, 1959; Huynh, 1978).

   b. Split the alpha level equally between the two analyses of variance.

This procedure leads to the most reasonable decision on the presence of treatment effects vis-a-vis Type I and Type II errors.

For the purpose of this paper, the most important step is #2. That is, since $C\Sigma C' = \sigma^2 I$ is not very likely, and since the available sphericity tests are not completely trustworthy considering the effects of non-normal data or considering fundamentally questionable Type I error control, non-sphericity might just as well be presumed and estimated using CSC'. That estimate can then be used to correct the mixed model degrees of freedom in step #3a. Conducting both tests in #3 remains important, contrary to the conclusions of O'Brien and Kaiser (1985), since one or the other analysis of variance can be more or less blind to certain patterns of treatment effects as they relate to the structure of CSC'.

Using an example data set, Robey and Barcikowski (1986) demonstrated the multivariate analog of this analysis problem. They showed that the structure of CSC' and the pattern of mean differences among the repeated measures can cause a power differential between the multivariate mixed model F and the doubly multivariate F, just as it does in univariate data. Their data set A contained an obvious treatment effect in the presence of a departure from multivariate sphericity. The treatment effect could only be detected by the multivariate mixed model F. Given this difference in the abilities of the two F's to detect certain patterns of treatment effects which directly relate to the structure of CSC', and given an unacceptable Type I error rate for the only available test of multivariate sphericity, it seems reasonable to conclude that routine preliminary testing for multivariate sphericity is a questionable practice.

Unfortunately, multivariate analogs of the sample based correction factors for the mixed model degrees of freedom have not yet been derived. As a result, it is impossible to use the multivariate mixed model with confidence when it is most needed. Robey and Barcikowski (1986) defined a conservative correction to the multivariate mixed model degrees of freedom which improves the data analyst's postion somewhat. However, much research remains to be done in this area. Although we cannot recommend the practice of preliminary testing on the multivariate sphericity assumption, we feel the issues surrounding multivariate sphericity are critical and warrant intense study.

REFERENCES

Anderson, T. W. (1958). An introduction to multivariate statistical analysis. New York: John Wiley & Sons.

Barcikowski, R. S., & Robey, R. R. (April, 1984a). Exploratory repeated measures analysis. A paper presented before the annual convention of the American Educational Research Association, New Orleans.

Barcikowski, R. S., & Robey, R. R. (1984b). Decisions in single group repeated measures analysis: statistical tests and three computer packages. The American Statistician, 38, 148-150.

Bartlett, M. S. (1950). Tests of significance in factor analysis. The British Journal of Psychology, 3, 77-85.

Bock, R. D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw-Hill, Inc.

Boik, R. J. (1981). A priori tests in repeated measures designs: effects of non-sphericity. Psychometrika, 46, 241-255.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 25, 484-498.

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

Cohen, J. (1977). Statistical power analysis for the behavioral sciences, (2nd ed.). New York: Academic Press.

Consul, P. C. (1967). On the exact distribution of the criterion w for testing sphericity in a p-variate normal distribution. Annals of Mathematical Statistics, 38, 1170-1174.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated measures experiments. Psychological Bulletin, 77, 446-452.

Dixon, W. J. (1983). BMDP statistical software. Berkley, CA: University of California Press.

Edgington, E. S. (1974). A new tabulation of statistical procedures in APA journals. American Psychologist, 29, 25-26.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 29, 885-891.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Hauck, W. W., & Anderson, S. (1984). A survey regarding the reporting of simulation studies. The American Statistician, 38, 214-216.

Hertzog, C., & Rovine, M. (1985). Repeated measures analysis of variance in developmental research: Selected issues. Child Development, 56, 787-809.

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 measurement designs have exact F-distributions. Journal of the American Statistical Association, 65, 1582-1589.

Huynh, H., & Mandeville, G. K. (1979). Validity conditions in repeated measures designs. Psychological Bulletin, 86, 964-973.

Imhof, J. P. (1962). Testing the hypothesis of no fixed main-effects in Scheffé's mixed model. The Annals of Mathematical Statistics, 33, 1085-1095.

International Mathematical and Statistical Libraries, Inc. (1982). The IMSL library (Vol. 2). Houston, TX: Author.

Jennings, J. R., & Wood, C. C. (1976). The ϵ adjustment procedure for repeated-measures analysis of variance. Psychophysiology, 13, 227-278.

John, S. (1971). Some optimal multivariate tests. Biometrika, 58, 123-127.

John, S. (1976). Fitting sampling distribution agreeing in support and moments and tables of critical values of sphericity criterion. Journal of Multivariate Analysis, 6, 601-607.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity conditions of repeated measures F tests. Psychological Bulletin, 87, 479-481.

Lane, R. E., & Lubin, A. (1963). The effects of correlation on the repeated measures design. Educational and Psychological Measurement, 23, 729-739.

Mathai, A. M., & Rathie, P. N. (1970). The exact distribution for the sphericity test. Journal of Statistical Research, 4, 140-159.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. Annals of Mathematical Statistics, 11, 204-209.

Muirhead, R. J., & Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. Biometrika, 67, 31-43.

Nagao, H. (1973). On some test criteria for covariance matrix. The Annals of Statistics, 1, 700-709.

Nargarsenker, B. N. (1976). Exact non-null distribution of the likelihood ratio criterion for covariance matrix. The Canadian Journal of Statistics, 4, 237-254.

Nargarsenker, B. N., & Pillai, K. C. S. (1973). The distribution of the sphericity criterion. Journal of Multivariate Analysis, 3, 226-235.

Norusis, M. J. (1985). SPSS-X advanced statistics guide. New York: Mc-Graw Hill.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97, 316-333.

Pillai, K. C. S., & Nargarsenker, B. N. (1971). On the distribution of the sphericity test criterion in classical and complex normal populations having unknown covariance matrices. The Annals of Mathematical Statistics, 42, 764-767.

Robey, R. R. (1985). A Monte Carlo Investigation of Type I Error in the Single Group Repeated Measures Design with Multiple Measures per Occasion. Unpublished doctoral dissertation, Ohio Univerity.

Robey, R. R., & Barcikowski, R. S. (1984). Calculating the statistical power of the univariate and the multivariate repeated measures analyses of variance for the single group case under various conditions. Educational and Psychological Measurement, 44, 137-143.

Robey, R. R., & Barcikowski, R. S. (1986). Type I error in the single group repeated measures design with multiple measures per occasion. A paper presented before the annual meeting of the American Educational Research Association, San Francisco.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measures. British Journal of Mathematical and Statistical Psychology, 32, 269-286.

Rouanet, H., & Lépine, D. (1970). Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. The British Journal of Mathematical and Statistical Psychology, 23, 147-163.

SAS Institute Inc. (1985). SAS users guide: statistics, version 5 edition. Cary, NC: Author.

Scheffé, H. (1959). The analysis of variance. New York: John Wiley and Sons.

SPSS Inc. (1986). SPSS-X user's guide, (2nd ed.). New York: Mc-Graw Hill.

Suguira, N. (1972). Locally best invariant test for sphericity and the limiting distributions. Annals of Mathematical Statistics, 43, 1312-1316.

Thomas, D. R. (1983). Univariate repeated measures techniques applied to multivariate data. Psychometrika, 48, 451-462.

Timm, N. H. (1980). Multivariate analysis of repeated measurements, in P. R. Krishnaiah (Ed.), Handbook of statistics. (Vol. 1, pp. 41-87). Amsterdam: North-Holland.

Winer, B. J. (1971). Statistical principles in experimental design, (2nd ed.). New York: McGraw-Hill, Inc.

TABLE 1

Actual Alpha Values of the w Criterion

n	k = 3	k = 5	k = 7	k = 10
k-1+3	NA	.004*	.010	.012
	NA	.046	.054	.051
k-1+10	NA	.010	.009	.009
	NA	.046	.043	.050
k-1+20	NA	.008	.009	.010
	NA	.047	.048	.047
k-1+30	NA	.010	.009	.010
	NA	.049	.051	.046

Note. The double entries for each Monte Carlo problem represent nominal alpha at .01 (top), and at .05 (bottom). An asterisk indicates a significant ($p \leq .01$) departure from nominal alpha. NA indicates that critical values were not available.

TABLE 2

Actual Alpha Values of the w Chi-Square Approximate

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.010 .050	.016* .066*	.029* .097*	.049* .155*
k-1+10	.012 .051	.012 .048	.011 .051	.014* .067*
k-1+20	.010 .053	.008 .048	.010 .050	.012 .052
k-1+30	.008 .049	.010 .049	.009 .054	.010 .048

Note. The double entries for each Monte Carlo problem represent nominal alpha at .01 (top), and at .05 (bottom). An asterisk indicates a significant ($p \leq .01$) departure from nominal alpha.

TABLE 3

Actual Alpha Values of the Bock F_{max} Test

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.092* .294*	.795* .973*	.998* 1.000*	1.000* 1.000*
k-1+10	.017* .110*	.510* .894*	.981* 1.000*	1.000* 1.000*
k-1+20	.000* .030*	.201* .702*	.886* .997*	1.000* 1.000*
k-1+30	.000* .004*	.065* .492*	.694* .987*	1.000* 1.000*

Note. The double entries for each Monte Carlo problem represent nominal alpha at .01 (top), and at .05 (bottom). An asterisk indicates a significant ($p \leq .01$) departure from nominal alpha.

TABLE 4

Actual Alpha Values of the John T Test

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.106*	.099*	.096*	.108*
k-1+10	.058*	.064*	.077*	.183*
k-1+20	.054	.065*	.016*	.010*
k-1+30	.023*	.009*	.003*	.000*

Note. Actual alpha equals .05 only. John (1976) does not provide critical values for the .01 level. An asterisk indicates a significant ($p \leq .01$) departure from nominal alpha.

TABLE 5

Actual Alpha Values of the Grieve Epsilon Test

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.028* .109*	.024* .099*	.021* .099*	.027* .111*
k-1+10	.018* .068*	.012 .066*	.029* .075*	.017* .082*
k-1+20	.011 .059*	.004* .061*	.014* .066*	.014* .065*
k-1+30	.010 .052	.022* .059*	.014* .063	.014* .062*

Note. The double entries for each Monte Carlo problem represent nominal alpha at .01 (top), and at .05 (bottom). An asterisk indicates a significant ($p \leq .01$) departure from nominal alpha.

TABLE 6

Example Data Set

Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
19.28	46.13	48.62	56.46	5[illegible].43	48.33
45.87	52.51	49.22	53.33	54.92	48.95
30.04	53.80	57.78	52.66	54.07	51.33
52.67	57.08	56.16	49.37	54.94	49.93
64.72	35.06	55.64	55.87	48.58	54.24
40.16	34.92	47.96	52.52	51.80	51.92
65.18	50.34	55.69	57.81	48.83	53.50
65.55	44.98	48.16	55.83	52.64	48.82
29.07	46.57	49.21	53.78	57.56	59.45
44.18	29.78	51.89	56.40	53.69	56.90
59.71	16.85	53.54	55.10	53.08	55.53
49.63	60.16	55.87	56.62	52.95	58.62
48.59	18.00	54.49	55.26	52.38	44.31
42.94	68.58	51.48	47.06	52.14	52.88
41.22	81.40	59.65	58.96	49.60	49.62
57.78	53.30	55.29	47.50	54.32	56.34
40.65	50.30	52.49	52.65	54.67	53.83
39.42	30.02	52.50	54.99	52.40	53.68
47.26	49.86	46.97	52.38	52.53	52.71
36.65	37.38	53.09	55.52	48.16	51.88

TABLE 7

Statistics for the Example Data Set

Test	Statistic	df	p
W	0.0094	5	<.01
Chi-square$_W$	79.74	14	<.0001
F_{max}	32.68	5,19	<.0001
T	0.3588	5	<.05
Epsilon	0.4106	5	<.01
Mixed Model F	3.57	2.05,39.01	.0365
Multivariate F	2.25	5,15	.1025

Note. The degrees of freedom for the mixed model F were corrected for $\hat{\epsilon}$.