DOCUMENT RESUME

ED 306 269 TM 013 100

AUTHOR	Robey, Randall R.; Barcikowski, Robert S.
TITLE	Type I Error for the Simultaneous Application of Two Tests for Repeated Measures Data.
PUB DATE	Mar 89
NOTE	25p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Algorithms; *Computer Simulation; *Error of Measurement; Hypothesis Testing
IDENTIFIERS	Null Hypothesis; *Repeated Measures Design; Robustness; Sphericity Tests; *Type I Errors

ABSTRACT

In analyzing exploratory repeated measures data with more than two measures, two competing tests must be administered simultaneously if one is to make an efficient and effective decision regarding the tenability of the null hypothesis of no differences among measurement means. Obviously, such a procedure is not without a cost vis-a-vis Type I error control. This study represents a measure of that cost. The Type I error properties for the simultaneous application of the mixed model test and the multivariate test were estimated. The results of this Monte Carlo robustness study suggest that a single rule of thumb designed to control Type I error (i.e., split the alpha or, alternatively, do not split the alpha) is not practical under all circumstances. A more dynamic method for satisfactory management of Type I errors, which deals with departures from sphericity and the related effects on correlation of the two tests, is outlined. A 28-item list of references and five data tables are included. (Author/TJH)

Type I Error for the Simultaneous Application of Two Tests for 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, San Francisco, March, 1989

Abstract

In analyzing exploratory repeated measures data with more than two measures, two competing tests must be administered simultaneously if one is to make an efficient and effective decision regarding the tenability of the null hypothesis of no differences among the measurement means. Obviously, such a procedure is not without a cost vis-a-vis Type I error control. This study represents a measure of that cost. The simulation results reported here suggest that a single rule of thumb designed to control Type I error (i.e., split the α or, alternatively, don't split the α) is not practical under all circumstances. A more dynamic method for the satisfactory management of Type I error is reported.

Type I Error for the Simultaneous Application of Two Tests for 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). In an exploratory investigation where specific a priori contrasts cannot be reasonably formulated, a researcher must depend upon an omnibus F statistic for a decision regarding the presence or absence of a treatment effect. An analysis alternative 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).

Huynh and Feldt (1970) and Rouanet and Lepine (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. Departures from sphericity are indexed by the value of ε which is well known (Box, 1954; Geisser and Greenhouse, 1958; Imhof, 1962).

Unfortunately, the mixed model test is not robust with respect to even small departures from sphericity (Huynh, 1978). Departures from sphericity cause the test to be positively biased. Moreover, behavioral and psychophysiological data almost certainly depart from sphericity (Keselman and Rogan, 1980). Several correction factors have been proposed to remedy this problem. In general, the correction factors are fractions which when applied to the mixed model degrees of freedom cause the test to approximate its theoretical null distribution. The correction factor used in this paper ($\hat{\varepsilon}$) was proposed by Greenhouse and Geisser (1959).

A second analysis alternative for this type of data does not assume sphericity. This test evaluates the same null hypothesis, but is conducted as a multivariate analysis of variance (Bock, 1975). The two algorithms, the adjusted mixed model and the multivariate model, differ sufficiently to cause them not to be interchangeable. In what follows, these differences are reviewed.

The Statistical Power Differential

The general form of the multivariate null hypothesis for a design with *g* groups, *k* repeated measures, and one dependent variable per occasion is:

$$H_o: ABC' = D$$

where A is a g-1 x g contrast matrix representing the between group hypothesis, B is a g x k matrix of cell means, and C is a k-1 x k contrast matrix representing the within factor hypothesis (Timm, 1975). In the single group repeated measures design there is no between group hypothesis, therefore, the matrix A is a scalar set at unity. The matrix D is typically a null matrix.

The multivariate sums of squares and cross products matrices for the hypothesis (H) and error (E) are given by:

$$H = CB'(X'X)^{-1}BC' \quad \text{, and} \tag{1}$$

$$E = C(Y-\hat{Y})'(Y-\hat{Y})C' \tag{2}$$

where B is a 1 x $\underline{k}$ vector of the treatment means over the $\underline{k}$ occasions, and X is a design matrix. Here, X is an $\underline{n}$ x 1 vector of ones, where $\underline{n}$ is the number of subjects. The matrix C is the $\underline{k}$-1 x $\underline{k}$ contrast matrix, and Y is an $\underline{n}$ x $\underline{k}$ matrix containing the $\underline{n}$ vectors of observations. The matrix $\hat{Y}$ is defined as $\hat{Y} = XB$.

The omnibus multivariate repeated measures hypothesis can be tested using

$$\Lambda = |E| \; / \; |E + H| \tag{3}$$

where Λ is Wilks's likelihood ratio criterion (Wilks, 1932) with $\underline{k}$-1, 1, and $\underline{n}$-$\underline{k}$+1 degrees of freedom, and $|E|$ denotes the determinant of the matrix E. A multivariate $\underline{F}$ statistic can be obtained by

$$\underline{F} = [(1- \Lambda)/ \Lambda] \; (\underline{v}_1/\underline{v}_2) \tag{4}$$

where $\underline{v}_1$ are the $\underline{k}$-1 hypothesis degrees of freedom, and $\underline{v}_2$ are the $\underline{n}$-$\underline{k}$+1 error degrees of freedom.

The $\underline{F}$ statistic for the multivariate approach to repeated measures given in Equation 4 is invariant to the linear contrasts in the matrix C. That is, any combination of $\underline{k}$-1 linearly independent, contrasts in the matrix C will yield the same test statistic. If the contrasts in C are row-wise orthonormal, the omnibus mixed model test statistic can be obtained from Equations 1 and 2 as well. The omnibus mixed model test statistic obtained through the multivariate model is given by

$$\underline{F} = [\mathrm{tr}(H) \;/\; \underline{q}_1 \;] \;/\; [\mathrm{tr}(E) \;/\; \underline{q}_2 \;] \quad , \qquad (5)$$

where tr is the trace of a matrix, $\underline{q}_1$ are the $\underline{k}$-1 hypothesis degrees of freedom, and $\underline{q}_2$ are the ($\underline{k}$-1)($\underline{n}$-1) error degrees of freedom.

When $\underline{k}$ = 2 the mixed model and multivariate model tests are identical. However, when $\underline{k}$ > 2, the two tests can differ substantially. When $\underline{k}$ > 2 and $\varepsilon = 1$, the mixed model test is always more powerful considering its greater number of error degrees of freedom. When $\underline{k}$ > 2 and $\varepsilon \neq 1$, the power of the two tests is determined by the pattern of mean differences relative to the structure of the variance-covariance matrix.

Consider the following explanation. Recall that B is a row vector of the $\underline{k}$ means. The elements in the vector given by BC' are the sum of the differences among the repeated measures found by each contrast in C. The $\underline{k}$-1 elements of this vector are the contrast effects. When they are squared, the contrast effects are referred to with the following notation $\psi^2{}_{\underline{i}}$, ($\underline{i}$ = 1, 2, ... $\underline{k}$-1).

Let L be a matrix with component column vectors which are the eigenvectors of CΣC', where Σ is the $\underline{k}$ x $\underline{k}$ variance-covariance matrix. When the orthonormal contrast matrix C is premultiplied by by L', CΣC' will be a diagonal matrix with the eigenvalues (λ) along the diagonal. In other words, CΣC' is reduced to its canonical form with uncorrelated contrast variances along the diagonal. These variances are referred to with the notation $\sigma^2_{\psi_i}$, where $\sigma^2_{\psi_i} = C\Sigma C'(i,i)$, ($i$ = 1, 2, ... k-1). Note that each contrast effect, ψ^2_i, is associated with its contrast variance, $\sigma^2_{\psi_i}$.

Building upon Imhof (1962) and upon Davidson (1972), Barcikowski and Robey (1984a) explained the statistical power difference between the mixed model test and the multivariate test by examining their respective noncentrality parameters. A brief review follows.

With the above restrictions on C, the noncentrality parameter for the mixed model (δ^2) is given by the sum of the k-1 mixed model contrast noncentrality parameters. The noncentrality parameter for the ith mixed model contrast (d^2_i) is given by:

$$d^2_i = \frac{n(k-1)\psi^2_i}{\sum_{j=1}^{k-1} \sigma^2_{\psi_j}} \qquad (6)$$

Similarly, the multivariate noncentrality parameter (γ^2) is given by the sum of the k-1 multivariate contrast noncentrality parameters. The noncentrality parameter for the ith multivariate model contrast (g^2_i) is given by:

$$g^2_i = \frac{n\psi^2_i}{\sigma^2_{\psi_i}} \qquad (7)$$

When one (or some) combination(s) of the $\underline{k}$ means (i.e., $\psi^2_{\underline{i}}$) represents most of the treatment effect and when that same combination accounts for a relatively large portion of the treatment by subjects interaction effect (i.e., $\sigma^2_{\psi_{\underline{i}}}$), the mixed model test dominates in terms of power. On the other hand, when that same contrast accounts for a relatively small portion of the treatment by subjects interaction effect, the multivariate test dominates in terms of power. That is, the pooling of errors in the denominator of δ^2 will wash out an isolated treatment effect when substantial experimental error exists that is not associated with that treatment effect.

Davidson (1972) found that the difference in power between the two tests was most noticable when small treatment effects were coupled with small $\underline{n}$'s (i.e., $\underline{n}$ exceeds $\underline{k}$ by no more than a few.) Barcikowski and Robey (1984a) noted that small $\underline{n}$'s and small treatment effects were likely occurrences in exploratory investigations.

Exploratory investigations are those experiments conducted when a researcher has little prior information, experience or both with which to design a study. In a repeated measures research design, this means that a researcher has no way of identifying a priori which of the two repeated measures tests will be more powerful when applied to the observations. That is, in an exploratory situation, it is impossisble to make valid estimations of the above elements in order to complete the necessary calculations for determining the test of choice. It is this research design dilemma that motivated the the present study.

Purpose

Barcikowski and Robey (1984a, 1984b) and Robey and Barcikowski (1984, 1987) addressed this dilemma by advocating simultaneous application of the adjusted mixed model test and the multivariate model test to evaluate the tenability of the omnibus null hypothesis in exploratory experiments. While this advice is sound with respect to Type II error control, it is problematic with respect to the maintenance of Type I error control. As a further step, Barcikowski and Robey (1984a, 1984b and elsewhere) have suggested splitting the Type I error tolerance (α) equally between the two tests. This is probably a conservative procedure given that the Type I error rate for two simultaneously conducted independent tests is $1-(1-\alpha)^2$, or approximately 2α. Others have informally suggested that since the two tests are not independent, one might as well use the same α for both tests.

The purpose of the present investigation was to estimate the Type I error properties for the simultaneous application of the mixed model test and the multivariate test. The mixed model test was here adjusted for $\hat{\varepsilon}$. The research design selected for examination was the single group repeated measures design. A Monte Carlo technique was employed because the mathematical derivation of the distribution for this procedure is precluded by the complexity resulting from two separate error terms and two separate error degrees of freedom.

Methods

The independent variables in this Monte Carlo robustness study were: the magnitude of the departure from sphericity; the number of measurement occasions ($\underline{k}$); and the number of observation units ($\underline{n}$). The magnitude of the departure from sphericity was varied at the following levels: no departure (e.i., $\varepsilon = 1$), slight departure (i.e., $\varepsilon = .9$), moderate departure (i.e., $\varepsilon = .75$), severe departure (i.e., $\varepsilon = .5$), and a maximal departure (i.e., $\varepsilon = 1/(\underline{k}-1)$). The number of occasions in a single group repeated measures design was varied at 3, 5, 7 and 10. The number of observations in the design was varied at $(\underline{k}-1) + 3$, $(\underline{k}-1) + 10$, $(\underline{k}-1) + 20$, and $(\underline{k}-1) + 30$. Thus, the research design under investigation was represented by reasonable ranges of departures from each of sphericity, of design size and of sample size.

The dependent variable in this experiment was the proportion ($\hat{\pi}$) of incorrect rejections of the null hypothesis as indicated by at least one of the two algorithms.

A FORTRAN subroutine, DRNMVN, from the International Mathematical and Statistical Libraries, Inc. (IMSL, 1987) was used to generate multivariate normal data for each variance-covariance matrix. The analysis program was a double precision G level VS FORTRAN program.

Statistical Hypotheses

The general form of the null and alternate hypotheses were H_o: $\pi = \alpha$ and H_o: $\pi \neq \alpha$. Here, π represents the population proportion of tests which exceed a critical value, and α represents nominal Type I error. Nominal Type I error was examined at .01 and .05. Bradley's (1978) robustness criterion of $\alpha \pm \alpha(0.5)$ was adopted as a definition for adequate Type I error performances. Thus, departures of $\pm$ .005 from α = .01, and departures of $\pm$ .025 from α = .05 were defined as meaningful. A two-tailed test for proportions described by Cohen (1977, p. 213) was used to analyze the obtained results. The *a priori* α 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 to evaluate H_o: π = .01, and that 1085 observations were needed to evaluate H_o: π = .05.

Results

The results of this experiment can be found in Tables 1 through 5. In Table 1 it can be seen that when the sphericity assumption is met, the simultaneous test procedure yields acceptable, albeit slightly elevated, Type I error rates. In general, the same pattern can be seen in Table 2 which contains the observed Type I error rates for the slight departure from sphericity (i.e., $\varepsilon = .9$) condition.

Table 3 contains the $\hat{\pi}$ values observed under the moderate departure from sphericity (i.e., $\varepsilon = .75$) condition. This table contains seven unacceptably high values at the $\alpha = .01$ level and one unacceptably high value at the $\alpha = .05$ level. While these eight values exceed their respective α values, none of them represents a drastic departure. The situation is more serious under a severe departure from sphericity (i.e., $\varepsilon = .5$). More than half of the observed Type I error rates in Table 4 exceed α. The mean of the unacceptable $\hat{\pi}$'s is approximately 1.6α at both levels.

Table 5 contains the results for the condition of maximal departure from sphericity (i.e., $\varepsilon = 1/(\underline{k}-1)$). Every observed Type I error rate reported in Table 5 exceeds α. In fact, as $\underline{k}$ increases, $\hat{\pi}$ approaches 2α. Note that the values reported in Table 5 for the $\underline{k}=3$ condition are the same values reported in Table 4. That is, $\varepsilon = .5$ in both cases.

Discussion

The pattern characterizing these results across the various departures from sphericity is clear. As the value of ε decreases the values of $\hat{\pi}$ increase. This pattern is probably best explained by the correlation between the two tests. When the sphericity assumption is met or nearly met the two tests are maximally correlated. It can be seen that when $\varepsilon = 1$, the two noncentrality parameters are equal. As the magnitude of the departure from sphericity increases, the two tests become less well correlated. As can be seen in Table 5, $\hat{\pi}$ approaches $1-(1-\alpha)^2$ when $\varepsilon = 1/(\underline{k}-1)$, particularly for the greater levels of $\underline{k}$.

The question to be clarified by these results is, "What magnitude of Type I error tolerance should be split equally between the two repeated measures tests?" To date, answers to this question have ranged from α (i.e., give each test one half of α) to 2α (i.e., give each test the full measure of α).

The results reported here suggest that the former tack is reasonable when the sphericity assumption is, or is nearly, met. However, when the data are characterized by a very severe departure from the sphericity assumption, the latter approach is indicated. Neither tack is particularly appropriate under a moderate departure from sphericity. If one assumes on the basis of Huynh and Feldt (1976) that behavioral data usually do not fall below $\varepsilon = .75$, then the liberal decision appears the more attractive in terms of general merit.

The dilemma, however, remains. The fact that the magnitude of the departure from sphericity cannot be estimated _a priori_ in an exploratory study precludes prudent application of either tack. For this reason we recommend the following.

1. Treat the value of $\hat{\varepsilon}$ as a descriptive statistic.

2. Set α for the simultaneous test procedure somewhere between 1α and 2α on the basis of the magnitude of the departure from sphericity.

3. That error tolerance should then be divided equally between the two tests.

Accordingly, the Type I error tolerance for the simultaneous test procedure would approach 2α as the value of $\hat{\varepsilon}$ increases. Under the worst of all circumstances vis-a-vis sphericity, the Type I error tolerance for the simultaneous test procedure would be set at, or near, 1α.

An aspect of this strategy which deserves comment is the fact that a sample estimate becomes the linchpin for a decision regarding an inferential test. While this practice is often cause for concern, here we note the results of several simulation studies that have compared the $\hat{\varepsilon}$ adjusted mixed model test to the ε adjusted mixed model test in terms of Type I and Type II error performance (Collier, Baker, Mandeville and Hayes, 1967; Mendoza, Toothaker, and Nicewander, 1974; Stollof, 1970; and Wilson, 1975). In each of these studies, the estimate perfomed very much like the parameter. As a result, practitioners now routinely adjust the mixed model degrees of freedom with $\hat{\varepsilon}$ rather than ε. It would seem then that $\hat{\varepsilon}$ represents relatively stable and unbiased estimate.

Subsequent simulations have indicated that setting the value in step #2 at 1.7 has the desired effect in terms of Type I error control when $\varepsilon = .75$. Moreover, values of 1.25 and 1 have worked well for $\varepsilon = .5$ and $\varepsilon = 1/(\underline{k}-1)$, respectively. Researchers who prefer a more stringent definition of Type I error tolerance than that used here may want to set the value in step #2 at 1.8 or so, even when the sphericity assumption is nearly met.

Conclusion

Barcikowski and Robey (1984a, 1984b) and Robey and Barcikowski (1984, 1987) have related a strategy for analyzing exploratory repeated measures data which effectively manages Type II error. Their strategy affords the application scientist confidence with respect to the detection of false omnibus null hypotheses. The results reported here provide the application scientist with similar confidence vis-a-vis the management of Type I error when analyzing repeated measures data.

REFERENCES

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.

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

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: Effects of inequality of 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.

Collier, R. O., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 32, 339-353.

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

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.

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., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in random blocks and split-plot designs. Journal of Education Statistics, 1, 69-82.

Imhof, J. P. (1954). Testing the hypothesis of no fixed main-effects in Scheffe'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.

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.

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

Mendoza, J. L., Toothaker, L. E., & Nicewander, W. A. (1974). A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. Multivariate Behavioral Research, 9, 165-178.

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. (April, 1987). Sphericity tests and repeated measures data. A paper presented before the annual convention of the American Educational Research Association, Washington, D.C.

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.

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

Stoloff, P. H. (1959). Correcting for heterogeneity of covariances for repeated measures designs of the analysis of variance. Educational and Psychological Measurement, 30, 909-924.

Timm, N. H. (1980). Multivariate analysis with applications in education and psychology. Monterey, CA: Brooks/Cole Publishing Co.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

Wilson, K. (1975). The sampling distribution of the conventional, conservative and corrected F ratios in repeated measures designs with heterogeneity of covariance. Journal of Statistical Computation and Simulation, 3, 201-215.

TABLE 1

Observed Type I Error Rates for the Simultaneous Application of Both Tests When Sphericity Exists

n	*k* = 3	*k* = 5	*k* = 7	*k* = 10
k-1+3	.013 .073	.010 .064	.011 .068	.010 .065
k-1+10	.013 .064	.010 .063	.015 .067	.014 .071
k-1+20	.013 .062	.014 .063	.012 .064	.016* .070
k-1+30	.010 .060	.016* .066	.012 .064	.014 .064

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

TABLE 2

Observed Type I Error Rates for the Simultaneous Appication of Both Tests When $\varepsilon = .9$

$\underline{n}$	$\underline{k} = 3$	$\underline{k} = 5$	$\underline{k} = 7$	$\underline{k} = 10$
$\underline{k}$-1+3	.014 .073	.014 .070	.012 .068	.012 .064
$\underline{k}$-1+10	.015 .068	.014 .073	.013 .071	.015 .072
$\underline{k}$-1+20	.016* .062	.015 .069	.016* .070	.015 .064
$\underline{k}$-1+30	.017* .067	.014 .067	.013 .063	.012 .065

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

TABLE 3

Observed Type I Error Rates for the Simultaneous Appication of Both Tests When ε = .75

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.016* .071	.015 .077*	.012 .064	.011 .068
k-1+10	.016* .068	.016* .075	.016* .067	.014 .072
k-1+20	.018* .072	.018* .072	.013 .072	.015 .071
k-1+30	.015 .071	.017* .070	.015 .069	.014 .068

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

TABLE 4

Observed Type I Error Rates for the Simultaneous Appication of Both Tests When ε = .5

n	k = 3	k = 5	k = 7	k = 10
k-1+3	.017* .079*	.013 .071	.015 .080*	.013 .077*
k-1+10	.016* .086*	.014 .074	.012 .074	.015 .079*
k-1+20	.017* .075	.017* .077*	.016* .074	.016* .079*
k-1+30	.018* .081*	.016* .074	.016* .080*	.014 .075

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

TABLE 5

Observed Type I Error Rates for the Simultaneous Appication of Both Tests Under a Maximal Departure from Sphericity

n	*k* = 3	*k* = 5	*k* = 7	*k* = 10
k-1+3	.017*	.018*	.019*	.020*
	.079*	.088*	.090*	.098*
k-1+10	.016*	.017*	.019*	.019*
	.086*	.095*	.088*	.094*
k-1+20	.017*	.019*	.020*	.020*
	.075*	.086*	.087*	.089*
k-1+30	.018*	.018*	.019*	.020*
	.081*	.085*	.091*	.090*

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