DOCUMENT RESUME

ED 229 389 TM 830 136

AUTHOR	Berger, Dale E.; Selhorst, Susan C.
TITLE	A Simulation Comparison of Univariate and Multivariate Analyses of a Multi-Factor Repeated Measures Design.
PUB DATE	[81]
NOTE	14p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Comparative Analysis; *Multivariate Analysis; Research Design; Research Methodology; *Research Problems; *Simulation; Test Bias
IDENTIFIERS	*Repeated Measures Design; *Univariate Analysis

ABSTRACT

Although it is widely known that special assumptions are needed for univariate analysis of repeated measures data, researchers seldom examine their data for violation of these assumptions. This paper reviews ways in which repeated measures analyses are usually handled and describes limitations of these methods. A design with two within subject factors (3x3) was tested with a computer simulation of 1,000 such experiments (each with 30 subjects) to examine the bias of alternate test procedures with data similar to that which might reasonably be observed. Two data structures were used, with small and large violations of the univariate assumptions. Four methods of analysis were compared: unadjusted univariate, Geiser-Greenhouse conservative test, epsilon correction, and multivariate analysis. The multivariate test was the only procedure for which the empirical alpha error rate did not differ reliably from the nominal alpha for any effect tested here. It is recommended that multivariate procedures should be used for analysis of repeated measures designs when sample size permits. (Author/PN)

A Simulation Comparison of Univariate and Multivariate Analyses of a Multi-factor Repeated Measures Design

Dale E. Berger and Susan C. Selhorst, Claremont Graduate School

Running head: Simulation of Repeated Measures

Abstract

Although it is widely known that special assumptions are needed for univariate analysis of repeated measures data, researchers seldom examine their data for violation of these assumptions. In this paper, we review the ways in which repeated measures analyses are usually handled and describe limitations of these methods. A design with two within subject factors (3x3) was tested with a computer simulation of 1000 such experiments, each with 30 subjects. Two data structures were used, with small and large violations of the univariate assumptions. Four methods of analysis were compared: unadjusted univariate, Geiser-Greenhouse conservative test, epsilon correction, and multivariate analysis. The multivariate test was the only procedure for which the empirical alpha error rate did not differ reliably from the nominal alpha for any effect tested here. Our recommendation is that multivariate procedures should be used for analysis of repeated measures designs when sample size permits.

A Simulation Comparison of Univariate and Multivariate Analyses of a Multi-factor Repeated Measures Design[1]

Dale E. Berger and Susan C. Selhorst, Claremont Graduate School

In this paper we will review the ways in which repeated measures analyses are usually handled, describe limitations of these methods, present the results of an empirical comparison of four procedures, and make recommendations for the selection of a test.

Assumptions for Univariate Analysis

It is widely known that the conventional univariate test of significance for a within subjects factor may be positively biased when the data violate the "symmetry" assumption. This assumption is satisfied if a) the population variances within each treatment level on the repeated factor are homogeneous, and b) the covariances between these treatment levels are homogeneous. Together, these assumptions are also called "symmetry of covariance matrices" (Kirk, 1968), or "compound symmetry" (Scheffé, 1959). In addition, if there is a between-groups factor (or factors), it is necessary to assume that the covariance matrices for the repeated measures are identical for all groups.

It is less well known that the symmetry assumption specifies sufficient but not necessary conditions for the conventional univariate tests of the repeated factor to be valid. The actual necessary and sufficient condition is that any set of k-1 orthogonal normalized (orthonormal) contrasts on the repeated factor with k levels should generate a covariance matrix that has a "sphericity" pattern (i.e., equal variances and zero covariances) (Huynh & Feldt, 1970). This pattern is also called "circularity."

In practice, researchers seldom examine repeated measures data for violations of assumptions. Jennings and Wood (1976) tabulated the use of repeated measures analysis in the 1975 volume of Psychophysiology. They found that of 56 articles with repeated measures analysis, 47 ignored the sphericity (or symmetry) assumption; and 34 reported at least one F ratio that would not have been significant if the conservative Geisser-Greenhouse test had been used. The severity of violation of sphericity can be very substantial, as we discovered in our own longitudinal data from a developmental study. On reflection, it seems obvious that departures from sphericity should be expected with longitudinal data since covariances are naturally larger between measures close together in time than between measures taken farther apart.

The degree of sphericity of a covariance matrix of <u>k</u> levels can be indexed by a coefficient epsilon, which varies from $\varepsilon=1$ for perfect sphericity, to a lower limit of $1/(k-1)$ (Box, 1954). Simulation studies have shown that the standard unadjusted univariate tests are usually reasonably accurate when epsilon is greater than 0.7 (e.g., Rogan, Keselman, & Mendoza, 1979). Box (1950, 1954) developed methods for testing whether epsilon differs from 1.0, but there are at least two reasons why these tests are not likely to be useful: (1) the tests are not very powerful for small samples, which is where the bias in the univariate test is greatest; and (2) the test of epsilon is about as complex as using it to adjust the degrees of freedom, so one might just as well make the epsilon adjustment and bypass the test of significance.

<u>Analysis of Repeated Measures</u>

Several approaches have been suggested to avoid positive bias in repeated measures designs. In some applications, it may be possible to avoid heterogeneity among correlations between treatment levels by randomizing or counterbalancing the order of presentation. However, this procedure will not remove correlations between similar treatment conditions, and it is not even possible in longitudinal studies.

The sphericity assumption can be avoided altogether with nonparametric tests, such as the Friedman two-way "ANOVA." Nonparametric tests are not attractive substitutes because they test different hypotheses, and they do not make efficient use of data.

Perhaps the simplest procedure is the Geisser-Greenhouse conservative test wherein the degrees of freedom for the F test are multiplied by the smallest possible value of epsilon, 1/(1-k). Although this procedure has received wide endorsement, and it certainly does have the advantage of simplicity, its disadvantage is that it is much too conservative when the sphericity assumption is approximately satisfied. Thus, routine application of the Geisser-Greenhouse conservative adjustment is inappropriate.

A more accurate procedure is to multiply the degrees of freedom for the F test by an estimate of the population value of epsilon based on the sample variance-covariance matrix for the repeated factor (Box, 1954). This procedure is quite accurate, although it does not have good reliability when the number of observations per group is less than 15 (Collier, Baker, Mandeville, & Hayes, 1967). Probably the main reason the epsilon correction has not been used more is that computation of epsilon is not easily done by hand, and it has not been provided by popular statistical computer packages.

A three-step approach to testing significance of an F ratio for a repeated measures factor was first proposed by Greenhouse and Geisser (1959), and this approach has been endorsed by some standard textbooks on analysis of variance (e.g., Kirk, 1968). If the F ratio is not significant with the unadjusted degrees of freedom, the procedure is stopped and the null hypothesis is not rejected, since the null hypothesis would not be rejected with reduced degrees of freedom. If the F ratio is significant both with unadjusted degrees of freedom and with the conservative Geisser-Greenhouse adjustment by the smallest possible value of epsilon, then the null hypothesis can be rejected without further testing. If the F ratio is significant with unadjusted degrees of freedom, but not significant with the conservative test, then epsilon should be estimated from the data and used to adjust the degrees of freedom.

The final approach to be discussed here is multivariate analysis. In this approach, the k measures from a given individual are recast into a set of (k-1) contrasts. These contrasts are used as multiple dependent measures for each individual to test the multivariate null hypothesis that the mean of each contrast in the set is zero. No assumptions need to be made about the form of the variance-covariance matrix, although the form is assumed to be the same for all treatment groups. The method provides an exact test, even for complex designs. A limitation is that the multivariate test is less powerful than the univariate tests when the sphericity assumption is satisfied, especially when the sample size is small. In fact, the multivariate test cannot be calculated at all if the sample size does not exceed the number of measurements plus the number of treatment groups by at least one. Else, the pooled within subjects variance-covariance matrix does not have an inverse, and multivariate computations are not possible.

It should be noted that it is not appropriate to make unqualified statements about the relative power of the multivariate and univariate approaches, since the null hypotheses are not the same. For different sets of data, the multivariate test may be more powerful than the unadjusted univariate test, or less powerful than the conservative Geisser-Greenhouse test (Romaniuk, Levin, & Hubert, 1977). It is not the case that the multivariate test will always produce a probability value between the extremes of these two univariate test procedures.

Method

Relatively little empirical work has been done with complex designs with two or more factors with repeated measures. In the current study, we used computer simulation to examine the bias of alternate test procedures with data similar to that which might reasonably be observed. We assumed a research situation in which measures are taken under three conditions (C factor) at each of three different times (T factor), with sample N = 30. We constructed two data structures, with small and large departures from sphericity. Correlations were constructed to vary inversely with separation in time, and were higher between conditions 1 and 2 than between 1 and 3 or 2 and 3. The epsilon values for the factors ranged from .96 to .64, and for the interactions were .71 and .54. The epsilon values for the interactions were computed as products of the epsilons for the corresponding factors (McHugh, Sivanich, & Geisser, 1961).

A total of 1000 computer simulations were done on each data set, using four different methods:

1) unadjusted univariate;

2) Geisser-Greenhouse conservative test;

3) epsilon correction using sample values of epsilon; and

4) multivariate analysis, using MANOVA.

The actual population means were all equal, so that an unbiased test would produce (false) "significant" results at the rate set by alpha.

Results and Conclusions

The results are summarized in Table 1, and can be described as follows:

1) The unadjusted univariate test was too liberal, especially for epsilon values below .7.

2) The Geisser-Greenhouse test was too conservative, especially for epsilon larger than .7.

3) The epsilon correction was reasonably unbiased except for very small values of epsilon. The tabled values are for rounded degrees of freedom; when degrees of freedom were truncated, the test procedure became much too conservative.

4) The multivariate test was accurate throughout.

It should be noted that about a third of the tests that were significant with the multivariate test were not significant with the unadjusted univariate test. We interpret this to mean that one does not have to construct highly artificial data to find a case where the multivariate procedure is more sensitive than the most powerful univariate test.

If freedom from bias is desired, neither the unadjusted univariate test nor the Geisser-Greenhouse conservative test is appropriate. The former can be much too liberal, while the latter is generally much too conservative. The epsilon adjustment is a great improvement, but it may be too liberal for very small values of epsilon. The multivariate procedure did not depart from the nominal alpha at any level of epsilon tested here.

With the recent addition of MANOVA to SPSS, access to the multivariate test should no longer be an obstacle. Our recommendation is that multivariate procedures routinely be used for analysis of repeated measures designs when sample size permits.

References

Box, G. E. P. Problems in the analysis of growth and wear curves. Biometrics, 1950, 6, 362-389.

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

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

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

Huynh, H. S., & Feldt, L. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, 1970, 65, 1582-1585.

Jennings, J. R., & Wood, C. C. The ϵ-adjustment procedure for repeated-measures analyses of variance. Psychophysiology, 1976, 13, 277-278.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.

McHugh, R. B., Sivanich, F., & Geisser, S. On the evaluation of personality changes as measured by psychometric test profiles. Psychological Reports, 1961, 9, 335-344.

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

Romaniuk, J. G., Levin, J. R., & Hubert, L. J. Hypothesis-testing procedures in repeated measures designs: On the road map not taken. Child Development, 1977, 48, 1757-1769.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Footnote

[1]The computer simulation was conducted by the second author as part of a master's thesis under the supervision of the first author. Portions of this paper were presented at the meeting of the Western Psychological Association in Los Angeles, 1981.

Table 1

Empirical Type I Error Rates for Each of Four Test Procedures[a]

	Unadjusted Univariate		Conservative Adjustment		Epsilon Adjustment		MANOVA	
Alpha:	.05	.01	.05	.01	.05	.01	.05	.01
Epsilon[b]								
.960	.049	.007	.014**	.003*	.048	.007	.044	.010
.845	.053	.017*	.029**	.004	.049	.015	.047	.008
.736	.063	.017*	.032**	.004	.043	.004	.045	.007
.707	.074**	.028**	.014**	.000**	.047	.012	.049	.013
.636	.082**	.028**	.048*	.004	.051	.009	.055	.006
.537	.107**	.055**	.035*	.005	.067*	.026**	.055	.013

[a] Each entry is based on 1000 replications; the design is 3x3 repeated measures with 30 subjects.

[b] Epsilon values for factors Time (T), Condition (C), and TxC, respectively, are .736, .960, and .707 for Data Set 1, and .636, .845, and .537 for Data Set 2.

*Empirical probability outside 95% confidence interval.

**Empirical probability outside 99% confidence interval.