

The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated

David M. Lane^a, 

^aRice University

Abstract ■ Violating the assumption of sphericity in repeated-measures analysis of variance leads to an inflated Type I error rate. The first portion of this article provides a thorough yet non-technical description of the sphericity assumption and explains why violations of sphericity lead to an inflated Type I error rate. The second portion describes univariate and multivariate approaches for addressing the problem of an inflated Type I error rate. The univariate approach involves estimating the parameter ε that reflects the degree to which sphericity is violated and then reducing the degrees of freedom by multiplying them by the estimate of ε . Two estimates of ε , $\hat{\varepsilon}$ and $\tilde{\varepsilon}$, have been recommended. The former has lower power than the latter whereas the latter fails to fully control the Type I error rate under some circumstances. The multivariate approach does not assume sphericity and therefore does not have an inflated Type I error rate. A decision tree for deciding among $\hat{\varepsilon}$, $\tilde{\varepsilon}$, and the multivariate approach based on a review of previously published simulations is presented along with a JavaScript program to automate the navigation of the decision tree.

Keywords ■ analysis of variance, sphericity, epsilon correction, repeated measures.

 Lane@rice.edu

 DML: 0000-0002-6364-9945

 10.20982/tqmp.12.2.p114

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ Two anonymous reviewers.

Introduction

Between-subjects ANOVA assumes independence of observations, normality within groups, and homogeneity of variance. Repeated measures ANOVA has an additional assumption called *sphericity*. Sphericity has been defined both in terms of the variances of difference scores (Kim, 2015; Myers, Well, & Lorch, 2010; Phakiti, 2015; Warner, 2012; Weinfurt, 1995) and in terms of the variances and covariances of orthogonal contrasts (Bijleveld et al., 1999; Bock, 1975; Crowder & Hand, 1990; Stevens, 2000; Winer, Brown, & Michaels, 1991). Although both definitions are technically correct, the definition of sphericity in terms of orthogonal contrasts is generally superior because it (a) generalizes to multi-factor designs, (b) reveals why violating sphericity increases the Type I error rate, and (3) makes clear why the univariate approach for correcting for sphericity violations reduces the degrees of freedom.

It is noteworthy that matrix algebra is not used in any of the above sources that defined sphericity in terms of

variances of difference scores but was used in all that defined sphericity in terms of orthogonal contrasts. Although matrix algebra provides a powerful and compact way to present the formulas related to sphericity, it is not difficult to explain sphericity in terms of orthogonal contrasts without using matrix algebra. To make this article more accessible to researchers not well versed in matrix algebra, the few places in which matrix algebra is used can be skipped without loss of continuity.

The first section of this article presents definitions of sphericity in terms of both variances of difference scores and variances and covariances of orthogonal contrasts. As will be seen, the consequences of violating sphericity are easy to deduce from the latter definition. The second section describes procedures for correcting for violations of sphericity in repeated-measures ANOVA and the third describes multivariate analyses that do not assume sphericity. The final section provides guidance for choosing among the methods presented in the second section.

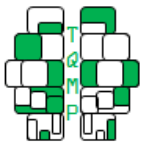


Table 1 ■ Population Showing Sphericity.

	T_1	T_2	T_3	$T_3 - T_2$	$T_3 - T_1$	$T_2 - T_1$
	9.102	7.821	9.174	1.353	0.072	-1.280
	8.422	8.673	4.317	-4.356	-4.105	0.252
	12.861	9.716	10.374	0.658	-2.488	-3.146
	11.887	18.386	13.767	-4.620	1.880	6.500
	11.083	9.788	10.858	1.069	-0.225	-1.295
	7.782	8.711	13.017	4.306	5.235	0.929
	8.416	10.495	7.073	-3.422	-1.342	2.080
	12.509	12.304	14.653	2.349	2.144	-0.205
	4.810	6.773	6.196	-0.577	1.386	1.963
	8.210	7.940	8.180	0.240	-0.030	-0.270
	16.808	12.038	12.194	0.156	-4.614	-4.770
	8.110	7.353	10.198	2.845	2.087	-0.757
Variance	10.00	10.00	10.00	8.00	8.00	8.00

Definitions of Sphericity

As described in the introductory section, sphericity can be defined in terms of variances of difference scores and in terms of variances and covariances of orthogonal contrasts. The two definitions are identical in that if the sphericity assumption is met for one definition it will be met for the second. However, the definition in terms of orthogonal contrasts is more general and more revealing about the consequences of violating sphericity.

Sphericity as Equal Variances of Difference Scores

In a one-way within-subjects design with three treatments, sphericity is met if the variances of the three pairwise difference scores ($T_3 - T_2, T_3 - T_1, T_2 - T_1$) are equal. To see this concretely, consider the small population of 12 people with scores on three treatments shown in Table 1. Note that the variances of all three pairwise differences are equal so sphericity is met.

A limitation of this definition is that it does not generalize to interactions in more complex designs such as in an $A(3) \times B(3)$ design. From a pedagogical point of view, this approach suffers from the fact that it does not provide an intuitive understanding of why sphericity must be assumed or the consequences of not meeting this assumption.

Sphericity in Terms of Orthogonal Contrasts

It is common to think of orthogonal contrasts as a method to partition the sum of squares of an ANOVA into components. Alternatively, but mathematically equivalent, one can consider an ANOVA as a way to test a set of orthogonal

comparisons with a single test. This section shows how the formulas for calculating ANOVA by combining orthogonal contrasts clarifies why the assumption of sphericity is necessary. Subsequently, it is shown that the population data in Table 1 that meet the definition of sphericity in terms of variances of differences also meet the definition in terms of orthogonal contrasts.

The fictitious data in Table 2 are from a one-way within-subjects design. An ANOVA on these data will be done by computing the mean square numerator and mean square denominator for two orthogonal contrasts and averaging them to get the mean squares for the ANOVA. Trend coefficients also known as *polynomial trends* are used here but any set of orthogonal contrasts would give the same final results.

The first step in computing a mean square for a contrast for a within-subjects variable is to create a new variable by applying the coefficients for the contrast to the raw data. This is done separately for each subject. This new variable for Subject i on contrast c will be denoted by l_{ic} where l stands for *linear contrast*. In Table 2, contrast c_1 has coefficients -0.7071, 0, and 0.7071, which are the coefficients for the linear component of trend scaled so that their sum of squares is 1. When coefficients are scaled so that their sum of squares is 1, they are said to be *normalized to 1*.¹

The value of l_{11} is:

$$l_{11} = -0.7071 \times 3 + 0.0000 \times 5 + 0.7071 \times 7 = 2.828. \quad (1)$$

The value of l_{i2} is calculated in like manner using the contrast c_2 with coefficients -1, 2, -1 normalized to 1 which are -0.4083, 0.8165, and -0.4083. These coefficients for c_2

¹Normalizing to 1 is done by dividing each element in a contrast by the square root of the sum of squared coefficients of that contrast. For example, the coefficients for the linear component of trend for three levels in their typical integer form are -1, 0, and 1. The sum of these coefficients squared is $-1^2 + 0^2 + 1^2 = 2$. Since the square root of 2 is 1.4142, the normalized coefficients are -1/1.4142, 0/1.4142, and 1/1.4142 which equal -.7071, 0, .7071.

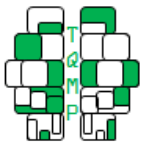


Table 2 ■ Fictitious Data for Three Treatments ($T_1, T_2,$ and T_3) from Five Subjects and Contrasts Representing the Linear (l_1) and Quadratic (l_2) Components of Trend.

	T_1	T_2	T_3	l_1	l_2
	3	5	7	2.828	0
	2	4	4	1.414	-0.817
	4	8	7	2.121	-2.041
	6	5	9	2.121	2.041
	4	3	5	0.707	1.225
Mean	3.8	5	6.4	1.838	0.082
Variance	2.2	3.5	3.8	0.65	2.617

are the coefficients for the quadratic component of trend. Note that the two sets of coefficients are "orthogonal" in the sense that the sum of cross products is 0. It will be shown in a later section that the contrasts are not truly orthogonal unless the assumption of sphericity is met.

More generally, letting Subject i 's score on trial j be denoted as X_{ij} and the coefficient for contrast c on trial j as w_{cj} , Subject i 's score on contrast c , l_{ic} , can be expressed as:

$$l_{ic} = \sum_{j=1}^t w_{cj} X_{ij} \tag{2}$$

where t is the number of treatments.

Linear Contrasts in Matrix Notation (Optional). The following description using matrix notation provides a compact way of expressing formulas but can be skipped without loss of continuity or understanding. Let \mathbf{X} be the $N \times t$ matrix of raw scores and \mathbf{W} be a $t \times (t - 1)$ matrix with each column containing the t coefficients for a contrast. Then \mathbf{L} computed as:

$$\mathbf{L} = \mathbf{XW} \tag{3}$$

is an $N \times (t - 1)$ matrix in which each row is a subject and each column is a vector of scores on a contrast.

For the data in Table 2:

$$\mathbf{L} = \begin{pmatrix} 3 & 5 & 7 \\ 2 & 4 & 4 \\ 4 & 8 & 7 \\ 6 & 5 & 9 \\ 4 & 3 & 5 \end{pmatrix} \begin{pmatrix} -0.7071 & -0.4803 \\ 0.0000 & 0.8165 \\ 0.7071 & -0.4803 \end{pmatrix} \tag{4}$$

$$= \begin{pmatrix} 2.828 & 0.000 \\ 1.414 & -0.817 \\ 2.121 & -2.041 \\ 2.121 & 2.041 \\ 0.707 & 1.225 \end{pmatrix}$$

Mean Squares and F. The means and variances of each of the l 's are shown in the last two rows of the last two columns of Table 2. The mean square numerator for contrast c (MSN_c) is computed as:

$$MSN_c = Nm_c^2 \tag{5}$$

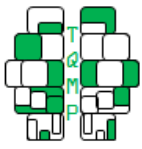
where N is the number of subjects and m_c is the mean for all subjects on l_c . If the null hypothesis that the population mean of $l_c = 0$ is true, then MSN_c is an estimate of population variance of its associated l_c with one degree of freedom.

For the data in Table 2, the mean square numerators for the two contrasts are $5 \times 1.8382 = 16.891$ and $5 \times 0.0822 = 0.034$. The mean square for treatments ($MS_{\text{treatments}}$) is the mean of these two mean squares and is equal to 8.462. Since each of the two contrasts provides an estimate of variance based on 1 degree of freedom, $df_{\text{treatments}} = 2$.

The variance of each contrast is an estimate of population variance of its associated l_c with $N - 1 = 4$ degrees of freedom. The mean square error (MS_e) is the mean of the two variances and is equal to $(0.652 + 2.617)/2 = 1.633$. Since each of the two values of MS_e is an estimate of variance based on $N - 1 = 4$ degrees of freedom error, df_e for the effect of treatments is $2 \times 4 = 8$. Therefore, $F(2, 8)$ is $MS_{\text{treatments}}/MS_e = 8.462/1.633 = 5.18$.

The astute reader will have noticed that two assumptions are necessary for the above calculations to be valid. First, since the numerator is the mean of two estimates of variance, it must be assumed that the two estimates are estimating the same population parameter. In other words, it must be assumed that the population variance of l_1 is equal to the population variance of l_2 . This same assumption must be made to average the two estimates for the denominator. Second, the degrees of freedom for the estimates are added in both the numerator and denominator. Since the degrees of freedom is the number of independent estimates, adding the degrees of freedom is only valid if the contrasts l_1 and l_2 provide independent estimates. This would be true only if the correlation between l_1 and l_2 when computed for every subject in the population is 0.

In more general terms, a repeated-measures ANOVA assumes that the variances of the orthogonal contrasts are all equal and that the correlations (and therefore the covariances) among the contrasts are 0. That is, if the coefficients



for all $t - 1$ contrasts were applied to the raw data to create $t - 1$ new l 's, the variances of these l 's would be equal and their covariances would be 0. This means that the covariance matrix of all l 's is a diagonal matrix with the common variance on the diagonal and 0's off the diagonal. This assumption is the assumption of sphericity.

The data in Table 1 are reproduced in Table 3 replacing the pairwise differences with l_1 and l_2 . The variances of l_1 and l_2 are both 4.00 and the covariance between them is

0.00. This indicates that the sphericity assumption is met. *Definition of Sphericity in Matrix Algebra (Optional).* In terms of matrix algebra, Σ^* , the covariance matrix among the l 's can be computed as:

$$\Sigma^* = W' \Sigma W \tag{6}$$

where Σ is the covariance matrix among the raw scores. For the data in Table 3,

$$\Sigma^* = \begin{pmatrix} -0.707 & 0.000 & 0.707 \\ -0.408 & 0.816 & -0.408 \end{pmatrix} \begin{pmatrix} 10 & 6 & 6 \\ 6 & 10 & 6 \\ 6 & 6 & 10 \end{pmatrix} \begin{pmatrix} -0.707 & -0.480 \\ 0.000 & 0.816 \\ 0.707 & -0.480 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \tag{7}$$

Since Σ^* is a diagonal matrix with equal values on the diagonal, the assumption of sphericity is met.

More Complex Designs. The definition of sphericity in terms of orthogonal contrasts generalizes easily from one-factor designs to multi-factor designs because, as in one-way designs, sphericity for each effect is determined by the orthogonal contrasts associated with the effect. Table 4 illustrates how to define sphericity for the interaction in a Days (3) x Treatments (3) within-subject design. The first two columns are orthogonal contrasts among the levels of Days, the next two are orthogonal contrasts among the levels of Treatments, Columns 5-8 represent the interaction and are formed by multiplying the associated columns for days and treatments. Finally, the last four columns are Columns 5-8 normalized to 1. Normalizing is done by dividing each element in a column by the square root of the sum of squared coefficients in that column. For Column 5, the sum of squared coefficients is 4 so Column 9 is computed by dividing each element of Column 5 by the square root of 4 = 2.

Sphericity for the interaction is determined by the variances and covariances among four contrasts computed by applying the coefficients in the last four columns to the raw data. If, in the population, all variances are equal and all covariances are 0, then the assumption of sphericity is met. Naturally, this method generalizes to designs with any number of factors.

Violating Sphericity in ANOVA

If sphericity is violated, then, because the contrasts are correlated, the degrees of freedom used to calculate probability values from F's are too large. As a result, the p values are too low and, depending on the degree to which sphericity is violated, the Type I error rate can be severely inflated.

The parameter ϵ reflects the degree to which sphericity is violated (Box, 1954). The worst case for a violation of sphericity is that all contrasts are perfectly correlated. In this case, the proper degrees of freedom is the degrees of freedom for a single contrast since all the others are completely redundant. This occurs for the minimum possible value of ϵ which is $1/(t - 1)$. The value of ϵ is 1.0 when the assumption of sphericity is met.

To correct for violations of sphericity, the degrees of freedom for both the numerator and the denominator should be reduced by multiplying by ϵ . Since the value of ϵ is not known, it has to be estimated from sample data. Two estimates of ϵ are commonly used: $\hat{\epsilon}$ Geisser and Greenhouse (1958) and $\tilde{\epsilon}$ (Huynh & Feldt, 1976). The former is often referred to as the *G-G correction* and the latter as the *H-F correction*. The estimate $\hat{\epsilon}$ is computed in a sample using the same formula as would be used in the population and does a good job controlling the Type I error rate. However, it has a slight negative bias which makes it a conservative correction. The negative bias can be seen in the case in which sphericity is met, since in that case the population value of ϵ is 1 but every sample value of $\hat{\epsilon}$ will be less than 1. The estimate $\tilde{\epsilon}$ is less conservative and therefore has greater power. However, when the violation of sphericity is severe, $\tilde{\epsilon}$ can result in a somewhat inflated Type I error rate.

Quintana and Maxwell (1994) used simulations to investigate several approaches to estimating ϵ and recommended using $\tilde{\epsilon}$ if $\tilde{\epsilon}$ is greater than 0.75 and $\hat{\epsilon}$ otherwise. This procedure has more power than using $\hat{\epsilon}$ exclusively while controlling the Type I error rate adequately.

It is important to be aware that the original formula for $\tilde{\epsilon}$ (Huynh & Feldt, 1976) has an error that affects the calculation when there are two or more groups (Lecoutre,

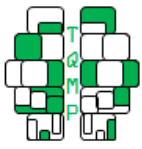


Table 3 ■ Demonstrating Sphericity in terms of Orthogonal Contrasts on the Data From Table 1.

	T_1	T_2	T_3	l_1	l_2
	9.102	7.821	9.174	0.051	-1.075
	8.422	8.673	4.317	-2.903	1.881
	12.861	9.716	10.374	-1.759	-1.553
	11.887	18.386	13.767	1.329	4.539
	11.083	9.788	10.858	-0.159	-0.966
	7.782	8.711	13.017	3.702	-1.379
	8.416	10.495	7.073	-0.950	2.246
	12.509	12.304	14.653	1.516	-1.043
	4.810	6.773	6.196	0.980	1.037
	8.210	7.940	8.180	-0.021	-0.208
	16.808	12.038	12.194	-3.263	-2.011
	8.110	7.353	10.198	1.476	-1.471
Variance	10.00	10.00	10.00	4.00	4.00

Note. The covariances among T_1 , T_2 , and T_3 are all 6 and the covariance between l_1 and l_2 is 0.

Table 4 ■ Contrasts for a 2×3 Design

D_1	D_2	T_1	T_2	D_1T_1	D_1T_2	D_2T_1	D_2T_2	D_1T_1	D_1T_2	D_2T_1	D_2T_2
-1	-1	-1	-1	1	1	1	1	0.500	0.289	0.289	0.167
-1	-1	0	2	0	-2	0	-2	0.000	-0.577	0.000	-0.333
-1	-1	1	-1	-1	1	-1	1	-0.500	0.289	-0.289	0.167
0	2	-1	-1	0	0	-2	-2	0.000	0.000	-0.577	-0.333
0	2	0	2	0	0	0	4	0.000	0.000	0.000	0.667
0	2	1	-1	0	0	2	-2	0.000	0.000	0.577	-0.333
1	-1	-1	-1	-1	-1	1	1	-0.500	-0.289	0.289	0.167
1	-1	0	2	0	2	0	-2	0.000	0.577	0.000	-0.333
1	-1	1	-1	1	-1	-1	1	0.500	-0.289	-0.289	0.167

Note. D_1 and T_1 are linear contrasts; D_2 and T_2 are quadratic contrasts. The contrasts in the first eight columns are not normalized to one whereas those in the last four are normalized to 1.

1991). The error leads to values of $\tilde{\epsilon}$ that are too high and therefore to an inflated Type I error rate (Chen & Dunlap, 1994; Quintana & Maxwell, 1994). Since even the major statistical programs may be using the incorrect formula, it is advisable to use the file sphericity.html associated with this article to make sure that $\tilde{\epsilon}$ is calculated correctly.

Multivariate Approach

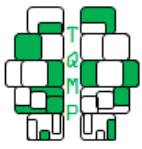
The multivariate approach is an alternative to using an ϵ correction. In short, the multivariate approach starts with a set of $t - 1$ new variables created by contrasts among the original variables. These contrasts can be orthogonal contrasts, differences between treatments, or any other set of contrasts as long as no contrast can be predicted perfectly from the other contrasts. The multivariate test is then conducted by doing a multivariate analysis of variance (MANOVA) on these contrasts. In practice, all modern statistical analysis programs make up the contrasts implicitly so there is no need to actually compute new vari-

ables. The multivariate approach makes no assumptions about the form of the covariance matrix so there is no assumption of sphericity. However, the multivariate approach does assume homogeneity of covariance matrices for designs with between-subjects variables and that the residuals are multivariate normal.

Choosing Between ANOVA and the Multivariate Approach

Both ANOVA using the ϵ correction and the multivariate approach control the Type I error rate adequately. The choice between these approaches, therefore, depends on their relative power. The relative power is complexly determined and is affected by sample size, the number of variables, and the covariances among the measures. A discussion of how and why these factors affect relative power follows.

Recall that in a repeated-measures ANOVA, each orthogonal contrast is counted equally. The multivariate ap-



proach, on the other hand, uses the sample data to weight the contrasts so as to maximize differences among conditions. For example, the linear and quadratic contrasts for the data in Table 2 would be weighted 0.633 and 0.069 respectively. Creating a new variable using these weights and testing whether it differs significantly from zero results in an $F(1, 4)$ of 27.33 and a p of 0.006. However, this is not a valid test since it does not account for the fact that these coefficients were determined *post hoc*. The multivariate test takes this into account and results in an $F(2, 3)$ of 10.25 and a p of 0.046. As noted previously, any set of contrasts among conditions would result in the same multivariate test result. For example, if difference-score variables were created for differences between Variables 2 and 1 and between Variables 3 and 2, the weights would be 0.419 and 0.476 respectively. Applying these weights would result in the same variable as the one created by applying the weights 0.633 and 0.069 to the linear and quadratic trend variables.

Since the multivariate test involves *post hoc* weighting, it is able to take advantage of chance. The larger the sample size, the less opportunity for taking advantage of chance and therefore the lower the penalty for choosing the weights *post hoc*. As a result, the power of the multivariate approach relative to repeated measures ANOVA increases with sample size. Similarly, the more repeated measures variables, the more *post hoc* weighting can take advantage of chance. Therefore, the power of the multivariate approach relative to repeated measures ANOVA decreases as the number of repeated measures variables increases. The pattern of the variances and covariances of the repeated measures variables is a critical determinant of relative power. As pointed out by Davidson (1972), there are some patterns that lead to vastly more power for the multivariate approach. As an example, consider the data in Table 5. The univariate test using $\tilde{\epsilon}$ (which equals 0.86 for these data) results in an $F(2.58, 23.20) = 2.86, p = 0.066$. By contrast, the multivariate test results in an $F(3, 7) = 9.89, p = 0.007$. The multivariate test has such a low p value because it was able to locate a small but very consistent difference between conditions. Specifically, although the difference between T_3 and T_1 is smaller than the difference between either of the other two treatments and T_1 , the difference is very consistent across subjects. The variance of the differences between T_3 and T_1 is only 0.46 as compared to the variance of 5.73 for the difference between T_2 and T_1 and of 8.00 for the difference between T_4 and T_1 . Repeated-measures ANOVA is unable to exploit this small but consistent difference because it weights all differences equally. The consistency of the difference between T_3 and T_1 has little impact when averaged with other sources of error.

When sphericity is met, the variances of all pairwise difference scores are equal and therefore there is no advantage to weighting some differences more than others. Therefore, when sphericity is met or nearly met, repeated-measures ANOVA tends to be more powerful than the multivariate approach. Note that this relative power advantage of ANOVA is not because ANOVA is more powerful when sphericity is met but, rather, because the multivariate approach is less powerful when sphericity is met.

Algina and Kesselman (1997) investigated the relative power of repeated-measures ANOVA and the multivariate approach over a range of sample sizes (N), number of repeated-measures variables (t), and values of ϵ . Sample size was defined relative to t such that sample sizes of $t + 10, t + 20, t + 30$, and $t + 40$ were examined for each combination of two values of t (4 and 8) and 5 values of ϵ (0.57, 0.75, 0.85, 0.90, and 0.96). For each value of ϵ , they examined a wide range of non-spherical covariance matrices. Algina and Kesselman (1997) presented the minimum, median, and maximum power advantage of the multivariate approach for the various values of N, t , and ϵ across covariance matrices for the relevant ϵ .

The power of the multivariate approach relative to repeated-measures ANOVA increased as a function of N , decreased as a function of t , and decreased as ϵ increased. Importantly, the relative power varied greatly among the covariance matrices even with the other factors held constant. Based on their results, Algina and Kesselman (1997) concluded that the multivariate test should be used if either (a) $t \leq 4, N \geq t + 15$, and $\tilde{\epsilon} < 0.90$ or (b) $t \geq 5, N \geq t + 30$, and $\tilde{\epsilon} < 0.85$.

Although this is sound advice and its simplicity makes it easy for a researcher to follow, the fact that it is a decision model based on multiple cutoffs means that a very small difference in one variable can result in a change in the recommendation method. For example, if $t = 3, N = 35$, and $\epsilon = 0.91$, repeated measures ANOVA would be recommended whereas if $t = 4, N = 19$, and $\epsilon = 0.89$, the multivariate approach would be recommended. Clearly a smaller value of t together with a much larger N in the former example should compensate for the very small difference in $\tilde{\epsilon}$.

An Algorithm for Choosing the Statistical Method

In order to develop a compensatory decision system, a multiple regression model was developed to predict the median power advantage of the multivariate approach as determined by Algina and Kesselman (1997). In the first model, the 32 values of median power were predicted predicted by N, t , and ϵ . The model fit well, $R^2 = 0.95$, but a model that also included the crossproduct of t , and ϵ fit even better, $R^2 = 0.97$. Even with this high level of fit,

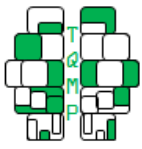


Table 5 ■ Fictitious Data Used to Compare Methods

T_1	T_2	T_3	T_4
4	8	5	9
3	5	5	7
5	9	6	8
4	7	6	6
3	4	5	8
5	7	6	7
4	4	5	8
5	8	7	3
4	7	5	4
5	1	5	2

there were a few instances in which the relative power of the multivariate test was overestimated. The biggest overestimate was for $N = 18$, $\varepsilon = 0.85$, and $t = 8$ for which the model predicted the power of the multivariate test would be greater by 0.04 whereas the median power was lower by 0.11. The second highest overestimate was for $N = 18$, $\varepsilon = 0.75$, and $t = 8$ for which the model predicted the power of the multivariate test would be greater by 0.06 whereas the median power was lower by 0.02. These two values were based on the smallest difference between N and t and suggest that the model may be less accurate for small N . Moreover, Algina and Kesselman's (1997) results were based on the population values of ε whereas researchers have to depend on an estimated value and estimated values will be relatively inaccurate with small sample sizes. To be prudent and following the recommendation of Algina and Kesselman (1997), it is recommended to use repeated-measures ANOVA in any case in which $N + t$ is not greater than or equal to 15. Considering only these cases, $R^2 = 0.98$ and the regression equation is

$$\text{Relative Power} = 51.07 - 52.40 \times \tilde{\varepsilon} + 4.17 \times t + 0.22 \times N - 6.75 \times t \times \tilde{\varepsilon} \quad (8)$$

where relative power is the power advantage of the multivariate test times 100. To account for cases in which there is a between-subjects variable, $g - 1$ should be subtracted from N where g is the number of groups. Naturally, $g = 1$ if there are no between-subjects variables.

The multivariate test is recommended when the relative power is greater than zero. Figure 1 shows the application of this formula for selected values of N , t , and $\tilde{\varepsilon}$. The extent to which the multivariate test becomes more powerful as N increases, t decreases, and $\tilde{\varepsilon}$ decreases is apparent.

The recommended decision tree is shown in Figure 2. The file sphericity.html included with this article implements this decision tree. As previously noted, the relative power varied greatly among the covariance matrices

even with the other factors held constant. For example, Algina and Kesselman (1997) found that for $N = 19$, $t = 4$, and $\varepsilon = 0.85$, the power difference between the multivariate approach and repeated-measures ANOVA ranged from a power advantage of 0.20 for the multivariate approach to a power advantage of 0.11 for repeated measures ANOVA depending on the covariance matrix. Since the form of the covariance matrix will, in general not be known in advance, the recommendation system presented here does not guarantee that the more powerful method will be recommended.

Summary

The power of ANOVA with the $\hat{\varepsilon}$ or the $\tilde{\varepsilon}$ correction relative to the power of the multivariate test is complexly determined. Smaller sample sizes favor the ε corrections with $\tilde{\varepsilon}$ being more powerful than $\hat{\varepsilon}$. However, using $\tilde{\varepsilon}$ leads to an inflated Type 1 error rate with low population values of ε . Finally, low population values of ε are associated with higher power for the multivariate test relative to ANOVA with either of the ε corrections.

References

Algina, J. & Kesselman, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods*, 2, 208–218. doi:10.1037/1082-989X.2.2.208

Bijleveld, C. C., Leo, J. T., Leo, J., Mooijaart, A., Van Der Vand Der, W. A., Van Der Leeden, R., & Van Der Burg, E. (1999). *Longitudinal data analysis: designs, models and methods*. London, UK: Sage Publications.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. NY, NY: McGraw Hill.

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 correlation between errors in the two-way classification. *Annals of*

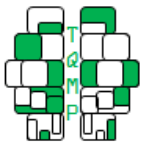
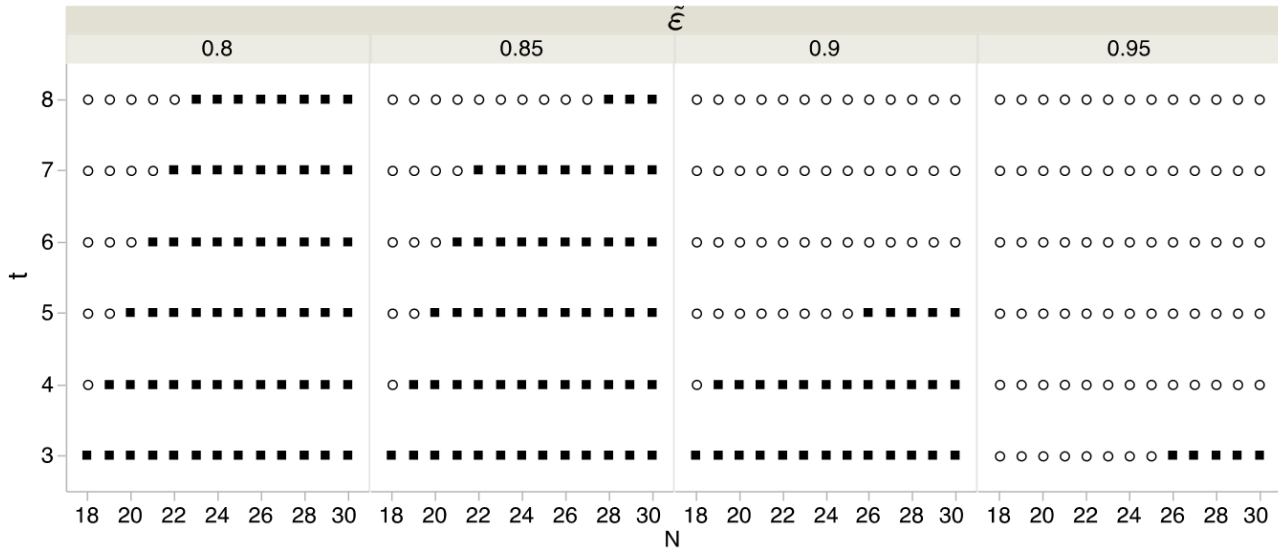


Figure 1 ■ Recommendations as a function of N, t, and ε̃. Values for which the multivariate test is recommended are indicated by filled squares whereas values for which repeated-measures ANOVA is recommended are indicated by open circles.



Mathematical Statistics, 10(1214), 290–302. doi:[4484-498](https://doi.org/10.1037/h0032674)

Chen, R. S. & Dunlap, W. P. (1994). A monte carlo study on the performance of a corrected formula for suggested by lecoure. *Journal of Educational and Behavioral Statistics*, 19, 119–126. doi:[10.3102/10769986019002119](https://doi.org/10.3102/10769986019002119)

Crowder, M. J. & Hand, D. J. (1990). *Analysis of repeated measures (vol. 41)*. Chapman and Hall/CRC Press.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated-measures experiments. *Psychological Bulletin*, 77, 446–452. doi:[10.1037/h0032674](https://doi.org/10.1037/h0032674)

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*, 25, 484–498.

Huynh, H. & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in the randomized block and split-plot designs. *Journal of Educational Statistics*, 1, 69–82. doi:[10.3102/10769986001001069](https://doi.org/10.3102/10769986001001069)

Kim, H. Y. (2015). Statistical notes for clinical researchers: a one-way repeated measures anova for data with repeated observations. *Restorative Dentistry & Endodontics*, 40, 91–95. doi:[10.5395/rde.2015.40.1.91](https://doi.org/10.5395/rde.2015.40.1.91)

Lecoutre, B. (1991). A correction for the approximate test in repeated measures designs with two or more independent groups. *Journal of Educational and Behavioral Statistics*, 16, 371–372. doi:[10.3102/10769986016004371](https://doi.org/10.3102/10769986016004371)

Myers, J. L., Well, A. D., & Lorch, R. F. (2010). *Research design and statistical analysis (third edition)*. London: Routledge.

Phakiti, A. (2015). *Experimental research methods in language learning*. Bloomsbury Publishing.

Quintana, S. M. & Maxwell, S. E. (1994). A monte carlo comparison of seven e-adjustment procedures in repeated measures designs with small sample sizes. *Journal of Educational and Behavioral Statistics*, 19, 57–71. doi:[10.3102/10769986019001057](https://doi.org/10.3102/10769986019001057)

Stevens, J. P. (2000). *Applied multivariate statistics for the social sciences*. NY, NY: Taylor & Francis.

Warner, R. M. (2012). *Applied statistics from bivariate through multivariate techniques*. Thousand Oaks, CA: Sage Publications.

Weinfurt, K. P. (1995). Multivariate analysis of variance. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics*. Washington: American Psychological Association DC.

Winer, B. J., Brown, D. R., & Michaels, K. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.

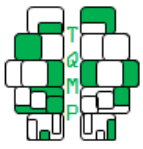
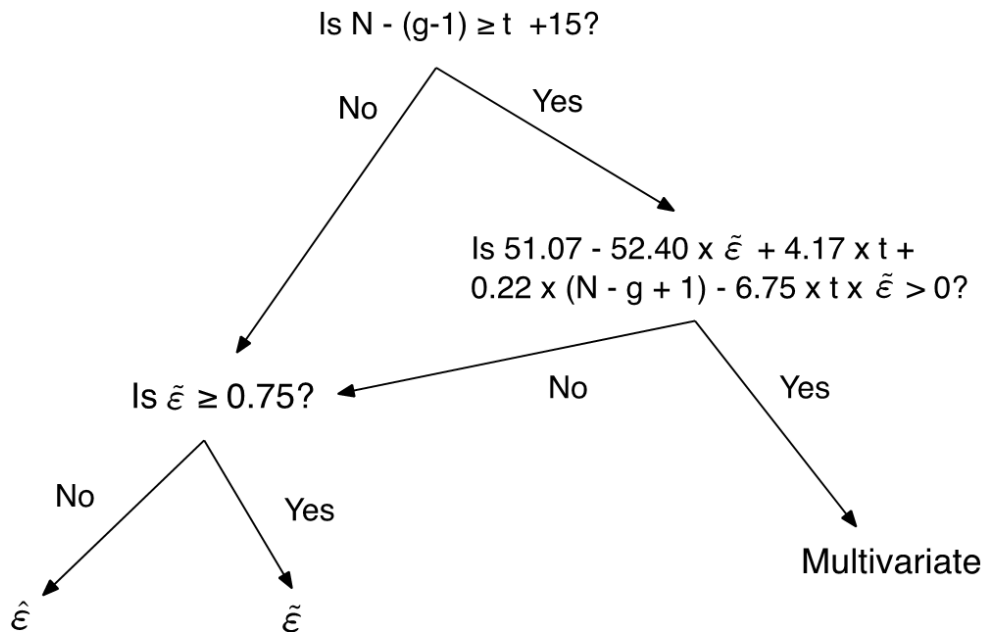


Figure 2 ■ Decision tree for deciding among analyses.



Open practices

The Open Material badge was earned because supplementary material(s) are available on the journal's web site.

Citation

Lane, D. M. (2016). The assumption of sphericity in repeated-measures designs: what it means and what to do when it is violated. *The Quantitative Methods for Psychology*, 12(2), 114–122. doi:10.20982/tqmp.12.2.p114

Copyright © 2016, Lane . This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 02/12/2015 ~ Accepted: 25/02/2016