Assumptions in Repeated-measures Designs Analysis

Assumptions in Repeated-measures Designs Analysis

Normality, Random Selection, Homogeneity of variance, Sphericity, Compound Symmetry, and Additivity are common assumptions for repeated-measures designs analysis

1. **Normality (or at least symmetry)**

Each set of data has a Normal distribution (or at least symmetrical distribution).

2. **Independent and Random Selection**

Selection of subjects has been independent and random from the population of interest.

3. **Homogeneity of variance**

Different sets of measurements have homogeneous variances.

4. **Sphericity**

Sphericity is an important assumption of a repeated-measures ANOVA. Differences in measurements between any two variables are similar to differences between any other two. The sphericity assumption states that the variance of the difference scores in a within-subjects design (the in a paired t-test) are equal across all the groups. The sphericity assumption (sometimes called the "**circularity**" assumption) assumes that the variances of each of these sets of difference scores are not statistically different from one another. When this assumption is violated, there will be an increase in Type I errors, because the critical values in the F-table are too small. One could say that the F-test is *positively biased* under these conditions.

5. **Compound Symmetry**

Another assumption of within-subjects ANOVA is the "compound symmetry" assumption. The compound symmetry assumption is a stricter assumption than the sphericity assumption. Not only do the variances of the difference scores need to be equal for pairs of conditions, but their correlations (technically, the assumption concerns covariances—the unstandardized version of correlation) must also be equal. A violation of this stricter compound symmetry assumption does not necessarily mean the sphericity assumption will be violated. SPSS does not currently provide a test of this assumption. Sphericity refers to the equality of variances of the *differences* between treatment levels. Whereas compound symmetry concerns the covariation between treatments, sphericity is related to the variance of the differences between treatments. Sphericity is a less restrictive form of compound symmetry.

6. Additivity

The error term for within-subjects is the interaction term, **S** x **A**. In other words, we assume that any variation in differences between levels of the independent variable is due to error variation. It is possible, however, that the effect of the independent variable **A** is different for different subjects, and thus there is truly an interaction between **S** and **A**. Thus, some of what we consider to be error when we calculate **S** x **A**, is really an interaction of subject and treatment and not error variation. For example if Factor A represents program groups, then an **S** x **A** interaction suggests the program is not equally effective for each subject. This is the so-called "**additivity**" assumption that there is no interaction between A and S that is not error (interactions are multiplicative or "non additive"). Because non additivity (a violation of this assumption) implies heterogeneous variances for the difference scores, the sphericity assumption will be violated if non additivity occurs. SPSS provides a test for this assumption, the Tukey test for non additivity, under the scale reliability command.

What is Sphericity? (In detail)

Sphericity (denoted by ε and sometimes referred to as circularity) is a more general condition of compound symmetry. Imagine you had a population covariance matrix Σ, where

$$\Sigma = \begin{bmatrix} s_{11}^2 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{2} & s_{22}^2 & \alpha_{23} & \cdots & \alpha_{2n} \\ \alpha_{3} & \alpha_{32} & s_{33}^2 & \cdots & \alpha_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n} & \alpha_{n2} & \alpha_{n3} & \cdots & s_{nn}^2 \end{bmatrix}$$

Equation 1

This matrix represents two things: (1) the off-diagonal elements represent the covariances between the treatments 1 ... *n* (you can think of this as the unstandardized correlation between each of the repeated measures conditions); and (2) the diagonal elements signify the variances within each treatment. As such, the assumption of homogeneity of variance between treatments will hold when:

$$s_{11}^2 \approx s_{22}^2 \approx s_{33}^2 \approx \ldots \approx s_{nn}^2$$

Equation 2

Compound Symmetry holds when there is a pattern of constant variances along the diagonal (i.e. homogeneity of variance, see Equation 2) and constant covariances off of the diagonal (i.e. the covariances between treatments are equal, see Equation 3). While compound symmetry has been shown to be a sufficient condition for conduction ANOVA on repeated measures data, it is not a necessary condition.

$$\alpha_{12} \approx \alpha_{13} \approx \alpha_{23} \approx \ldots \approx \alpha_{1n} \approx \alpha_{2n} \approx \alpha_{3n} \qquad \textbf{Equation 3}$$

Imagine a situation where there are four levels of a repeated measures treatment (A, B, C, D). For sphericity to hold, one condition must be satisfied:

$$s^2_{A-B} \approx s^2_{A-C} \approx s^2_{A-D} \approx s^2_{B-C} \approx s^2_{B-D} \approx s^2_{C-D} \qquad \textbf{Equation 4}$$

Sphericity is violated when the condition in Equation 4 is not met (i.e. the differences between pairs of conditions have unequal variances).

How is Sphericity Measured?

The simplest way to see whether or not the assumption of sphericity has been met is to calculate the differences between pairs of scores in all combinations of the treatment levels. Once this has been done, you can simply calculate the variance of these differences.

Example. Following table shows data from an experiment with 3 conditions (for simplicity there are only 5 scores per condition). The differences between pairs of conditions are calculated for each subject. The variance for each set of differences is then calculated. We saw above that sphericity is met when these variances are roughly equal. For this data, sphericity will hold when:

$$s^2_{A-B} \approx s^2_{A-C} \approx s^2_{B-C}$$

Condition A	**Condition B**	**Condition C**	**A-B**	**A-C**	**B-C**
10	12	8	-2	2	5
15	15	12	0	3	3
25	30	20	-5	5	10
35	30	28	5	7	2
30	27	20	3	10	7

Variance:	**15.7**	**10.3**	**10.3**

Here $s^2_{A-B} = 15.7$, $s^2_{A-C} = 10.3$, $s^2_{B-C} = 10.3$

So there is, at least some deviation from sphericity because the variance of the differences between conditions A and B is greater than the variance of the differences between conditions A and C, and between B and C. However, we can say that this data has **local circularity** (or **local sphericity**) because two of the variances are identical. This means that for any multiple comparisons involving these differences, the sphericity assumption has been met (for a discussion of local circularity see Rouanet and Lépine, 1970). The deviation from sphericity in the data in Table 1 does not seem too severe (all variances are *roughly* equal). This raises the issue of how we assess whether violations from sphericity are severe enough to warrant action.

Assessing the Severity of Departures from Sphericity (Mauchley's test)

Mauchly's test of sphericity (developed in 1940 by John W. Mauchly) is a popular test to evaluate whether the sphericity assumption has been violated. The null hypothesis of sphericity and alternative hypothesis of non-sphericity in the above example can be mathematically written in terms of difference scores.

$$\mathbf{H_0}: \sigma_{\text{Tx A} - \text{Tx B}} = \sigma_{\text{Tx A} - \text{Tx C}} = \sigma_{\text{Tx B} - \text{Tx C}}$$
$$\mathbf{H_1}: \sigma_{\text{Tx A} - \text{Tx B}} \neq \sigma_{\text{Tx A} - \text{Tx C}} \neq \sigma_{\text{Tx B} - \text{Tx C}}$$

Interpreting Mauchly's test is fairly straightforward. When the probability of Mauchly's test statistic is greater than or equal to $\boldsymbol{\alpha}$ (i.e. $p \geq \boldsymbol{\alpha}$, with $\boldsymbol{\alpha}$ commonly being set to .05), we fail to reject the null hypothesis that the variances are equal. Therefore, we could conclude that the assumption has not been violated. However, when the probability of Mauchly's test statistic is less than or equal to $\boldsymbol{\alpha}$ (i.e. $p < \boldsymbol{\alpha}$), sphericity cannot be assumed and we would therefore conclude that there are significant differences between the variances of the differences. It should be noted that sphericity is always met for two levels of a repeated measure factor and is, therefore, unnecessary to evaluate.

SPSS produces Mauchley's test, which tests the hypothesis that the variances of the differences between conditions are equal. If Mauchley's test statistic is significant (i.e. has a probability value less than 0.05) we must conclude that there are significant differences between the variance of differences, so the condition of sphericity has not been met. If, however, Mauchley's test statistic is non significant (i.e. has a probability value greater than or equal to 0.05) then it is reasonable to conclude that the variances of differences are not significantly different (i.e. they are roughly equal). So, in short, if Mauchley's test is significant then we must be wary of the F-ratios produced by the computer.

Mauchly's test is a test of the sphericity assumption using the chi-square test. Unfortunately, this test is not a very useful one. For small sample sizes, it tends to have too many Type II errors (i.e., it misses sphericity violations), and for large sample sizes, it tends to be significant even though the violation is small in magnitude (Type I error; e.g., Kesselman, Rogan, Mendoza, & Breen, 1980). Thus, the Mauchly's test may not be much helpful and may be misleading when deciding if there is a violation of the sphericity assumption.

Mauchley's test statistic ε (Epsilon) provides a measure of departure from sphericity. By evaluating epsilon, we can determine the degree to which sphericity has been violated. If the variances of differences between all possible pairs of groups are equal and sphericity is exactly met, then epsilon will be exactly 1, indicating no departure from sphericity. If the variances of differences between all possible pairs of groups are unequal and sphericity is violated, epsilon will be below 1. The further epsilon is from 1, the worse the violation.

Criticisms

While Mauchly's test is one of the most commonly used to evaluate sphericity, the test fails to detect departures from sphericity in small samples and over-detects departures from sphericity in large samples. Consequently, the sample size has an influence on the interpretation of the results. In practice, the assumption of sphericity is extremely unlikely to be exactly met so it is prudent to correct for a possible violation without actually testing for a violation.

Effect of Violating the Assumption of Sphericity?

If sphericity is violated, a decision must be made as to whether a univariate or multivariate analysis is selected. If a univariate method is selected, the repeated-measures ANOVA must be appropriately corrected depending on the degree to which sphericity has been

Multivariate test statistics (MANOVA)

Another alternative procedure is using the multivariate test statistics (MANOVA) since they do not require the assumption of sphericity. However, this procedure can be less powerful than using a repeated measures ANOVA, especially when sphericity violation is not large or sample sizes are small. O'Brien and Kaiser suggested that when you have a large violation of sphericity (i.e., epsilon < .70) and your sample size is greater than a + 10 (i.e., the number of levels of the repeated measures factor + 10), then a MANOVA is more powerful; in other cases, repeated measures design should be selected. Additionally, the power of MANOVA is contingent upon the correlations between the dependent variables, so the relationship between the different conditions must also be considered. SPSS provides an F-ratio from four different methods: Pillai's trace, Wilks' lambda, Hotelling's trace, and Roy's largest root. In general, Wilks' lambda has been recommended as the most appropriate multivariate test statistic to use.

Davidson (1972) compared the power of adjusted univariate techniques with those of Hotellings T^2 (a MANOVA test statistic) and found that the univariate technique was relatively powerless to detect small reliable changes between highly correlated conditions when other less correlated conditions were also present.

Mendoza, Toothaker and Nicewander (1974) conducted a Monte Carlo study comparing univariate and multivariate techniques under violations of compound symmetry and normality and found that "as the degree of violation of compound symmetry increased, the empirical power for the multivariate tests also increased. In contrast, the power for the univariate tests generally decreased".

As a general rule it seems that when you have a large violation of sphericity ($\varepsilon < 0.7$) and your sample size is greater than (a + 10) then multivariate procedures are more powerful whilst with small sample sizes or when sphericity holds ($\varepsilon > 0.7$) the univariate approach is preferred (Stevens, 1992).

It is also worth noting that the power of MANOVA increases and decreases as a function of the correlations between dependent variables (Cole *et al*, 1994) and so the relationship between treatment conditions must be considered also.