

How to Verify the Three Assumptions of a Repeated Measures ANOVA

Authored by
stats writer

December 2, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Verify the Three Assumptions of a Repeated Measures ANOVA*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103631>

The three critical statistical requirements for conducting a valid repeated measures ANOVA (RMA) are the independence of observations between subjects, the distribution of the response variable being normally distributed, and the condition of Sphericity. Meeting these core assumptions is essential; failure to do so can severely compromise the accuracy and reliability of the statistical conclusions drawn from the analysis.

A repeated measures ANOVA is a type of statistical test utilized to determine if a statistically significant difference exists between the means of three or more related groups, where the same individuals or subjects are measured under all conditions. This design is highly efficient as it minimizes inter-subject variability.

However, as a parametric test, the RMA requires that specific conditions regarding the data distribution and structure be met. Before proceeding with the analysis, researchers must ensure the following assumptions are satisfied:

Independence: Each individual observation must be independent from other subjects' observations.

Normality: The distribution of the response variable should be normally distributed at each level of the repeated measures factor.

Sphericity: The variances of the differences between all combinations of related groups must be equal.

If one or more of these assumptions are severely violated, the results of the repeated measures ANOVA may be unreliable, potentially leading to inaccurate estimations of the true population effect.

In this article, we provide a detailed explanation for each assumption, practical steps on how to determine if the assumption is met, and effective strategies for remediation if a violation is detected.

Assumption 1: Independence of Observations

The assumption of independence is a universal requirement for most standard statistical inference tests. In the context of the repeated measures design, this means that while measurements taken over time or across conditions *within* a single subject are inherently dependent, the data collected from one subject must be statistically independent of the data collected from every other subject in the study.

Failure to achieve independence between observations--for example, if participants were sampled non-randomly, or if they interacted in a way that influenced their scores--can lead to biased standard errors. This typically results in an inflated F-ratio, significantly increasing the probability of

making a Type I error (a false positive conclusion).

Assessing Independence

Verifying this assumption is primarily a methodological task rather than a statistical one. The easiest and most reliable way to check this assumption is to ensure that each individual in the dataset was recruited and sampled from the target population using a robust random sampling technique.

If the data collection process utilized proper experimental controls and guaranteed that the selection of one subject did not influence the selection or response of any other subject, then it is usually considered safe to proceed with the analysis assuming independence. This aspect relies heavily on the quality and rigor of the study design itself.

Addressing Independence Violations

When the assumption of independence is violated, it constitutes a serious threat to the validity of the statistical model. Because the dependency structure is often complex or unknown, mathematical remedies within the standard ANOVA framework are severely limited.

In cases where non-random sampling or clustering is confirmed, the most appropriate remedy often involves methodological correction: recruiting new subjects using a strict random sampling protocol. If the dependency has a clear structure (e.g., students nested within classrooms), specialized multilevel models should be considered as alternatives to the repeated measures ANOVA.

Assumption 2: Normality of the Response Variable

The second key assumption dictates that the distribution of the dependent variable for each experimental condition must follow a normal distribution. Technically, the RMA is concerned with the normality of the sampling distribution of the means, which is often satisfied even if the raw data is slightly skewed, especially when the sample size is large (as per the Central Limit Theorem).

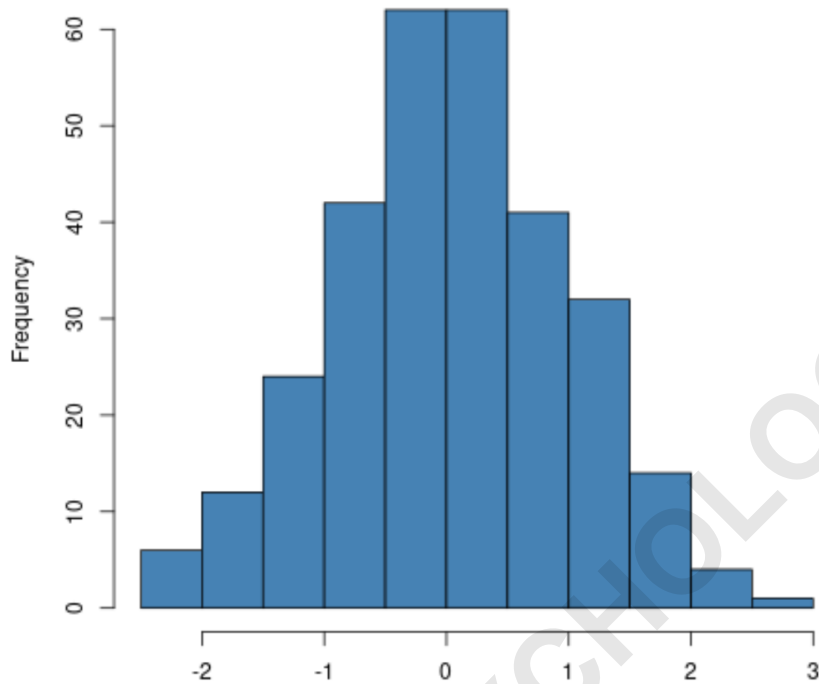
However, significant departures from Normality, particularly involving extreme skewness or kurtosis, can distort the error terms and lead to inaccurate probability calculations for the F-test. Therefore, verification of this assumption remains a crucial step in the data screening process.

Methods for Assessing Normality

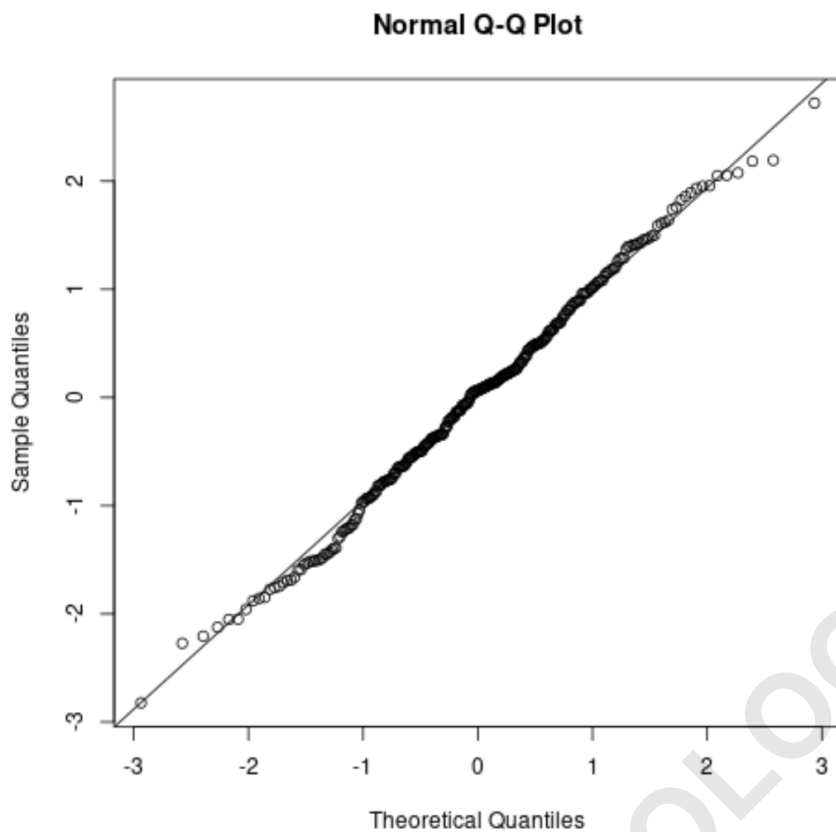
Normality is assessed through a combination of visual tools and formal statistical hypothesis testing.

1. Visual Inspection: Histograms and Q-Q Plots

Visual checks provide the most intuitive assessment. A histogram should exhibit a generally symmetrical shape, resembling a bell curve. If the visual structure is symmetrical, the normality assumption is often considered met:



Alternatively, a Q-Q plot compares the observed data quantiles against the expected quantiles of a theoretical normal distribution. If the data points lie closely along the straight diagonal line, the data is typically considered normally distributed:



2. Formal Statistical Testing

To provide quantitative evidence, researchers can conduct formal tests such as the Shapiro-Wilk test. In these tests, the null hypothesis posits that the data is normally distributed. If the resulting p-value is less than the significance threshold (e.g., $\alpha = 0.05$), the null hypothesis is rejected, implying non-normality.

However, it is crucial to understand that formal tests possess high statistical power, especially with large samples. An extremely large sample size means the Shapiro-Wilk test may detect minor, statistically significant but practically unimportant deviations from normality. For this reason, many statisticians recommend relying heavily on visual plots (histograms and Q-Q plots) to judge the practical severity of any non-normality.

Managing Normality Violations

The repeated measures ANOVA is considered robust against moderate violations of normality, provided the sample size is adequate. Nevertheless, if the assumption is severely violated, researchers have two primary options:

Data Transformation: Apply a mathematical transformation (e.g., square root, logarithm, or

reciprocal) to the response variable. This can normalize the distribution, but interpretation must then be based on the transformed scale, which can complicate reporting.

Non-Parametric Analysis: Utilize an equivalent non-parametric test that is assumption-free regarding distribution shape. For the RMA, the appropriate non-parametric alternative is the Friedman test.

Assumption 3: Sphericity

The assumption of Sphericity is unique to repeated measures designs involving three or more levels. It requires that the variances of the difference scores between all possible pairs of repeated measures conditions must be equal. For instance, the variance of (Time 1 - Time 2) must be approximately equal to the variance of (Time 1 - Time 3), and so on.

If this assumption is violated, the correlation structure between the repeated measures is uneven, resulting in a positively biased F-ratio. This inflation increases the probability of rejecting the null hypothesis when it is true (a Type I error), making the results of the repeated measures ANOVA statistically unreliable without correction.

Testing for Sphericity using Mauchly's Test

To evaluate Sphericity, researchers perform Mauchly's Test of Sphericity. This test formally assesses the homogeneity of variances of the difference scores.

The statistical hypotheses are:

H₀: The variances of the differences are equal (Sphericity holds).

H_A: The variances of the differences are *not* equal (Sphericity is violated).

If the p-value generated by Mauchly's Test is less than the predetermined alpha level (e.g., $\alpha = 0.05$), the null hypothesis is rejected, and we conclude that the assumption of Sphericity is violated. If the p-value is greater than 0.05, we fail to reject the null hypothesis and assume the condition is met.

The output of this statistical test typically appears as follows:

Within Subjects Effect	Mauchly's W	Chi-Square	df	p-value
Time	0.277	1.867	2	0.356

In the case shown, if the corresponding significance value is, for example, 0.230, we would

conclude that Sphericity is met since 0.230 is greater than 0.05.

Handling Violations of Sphericity

When Mauchly's Test forces the rejection of the null hypothesis, a statistical correction must be applied to the degrees of freedom (df) used in the ANOVA calculation. These adjustments mitigate the bias in the F-ratio by lowering the effective degrees of freedom, leading to a more conservative and reliable result.

Researchers typically select from three standard correction methods, which are based on an estimation of epsilon (ϵ), a measure of the severity of the violation:

Huynh-Feldt Correction: This is generally the least conservative correction and is often preferred when the estimated epsilon is close to 1.

Greenhouse-Geisser Correction: A more conservative correction, frequently used when the violation is substantial ($\epsilon < 0.75$).

Lower-bound Correction: This is the most conservative approach, applying the maximum possible correction.

By applying one of these corrections, the resulting p-values in the repeated measures ANOVA output are adjusted upward, accounting for the violated assumption. It is these corrected p-values that should be used to draw final conclusions about the statistical significance of the repeated measures factor.

Additional Resources for Repeated Measures ANOVA

The following tutorials provide additional information about the repeated measures ANOVA: