

Understanding the Repeated Measures ANOVA: Checking Key Assumptions

Authored by
Mohammed looti

November 1, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding the Repeated Measures ANOVA: Checking Key Assumptions*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7824>

A [Repeated Measures ANOVA](#) (RM-ANOVA) is a highly effective statistical tool utilized to determine if there are statistically significant differences among the means of three or more related groups. This method is specifically designed for within-subjects designs, meaning the **same subjects** are measured repeatedly across every condition or time point.

However, the validity and reliability of the inferences drawn from an RM-ANOVA fundamentally rely on meeting several critical statistical assumptions. If these prerequisites are not satisfied, the resulting inferential statistics--such as the [F-ratio](#) and the corresponding [p-value](#)--may be inaccurate, potentially leading to erroneous conclusions about treatment effects.

Researchers must rigorously assess the data against the three fundamental assumptions required for a valid repeated measures ANOVA:

Independence: Observations gathered from one subject must be statistically independent of observations gathered from all other subjects.

Normality: The distribution of the dependent variable (or the residuals) must approximate a [normal distribution](#) for each level of the within-subjects factor.

Sphericity: The variances of the differences between all possible pairs of related measurements must be equal.

This article provides a detailed explanation of each assumption, outlines the standard procedures for testing their validity using statistical software, and suggests appropriate corrective measures to take if a significant violation is detected.

Assumption 1: Independence of Observations

The first and perhaps most crucial assumption dictates that each [observation](#) included in the dataset must be statistically independent of every other observation, excluding the inherent relationship that exists among measurements taken from the same subject. This means that Subject A's data points must not in any way influence or be influenced by Subject B's data points. While the RM-ANOVA specifically handles the dependency within a single subject across conditions, dependency **between** subjects is fatal to the analysis.

Violations of independence are rarely detectable through post-hoc statistical testing; rather, they are typically a consequence of poor study methodology or flawed data collection procedures. Common scenarios that compromise independence include testing subjects in groups where they can communicate or influence one another's responses, or utilizing non-random sampling techniques that introduce systematic bias or clustering among participants.

Because the assumption of independence is foundational to nearly all parametric statistics, its violation is considered the most serious issue. If subjects are related or systematically influenced

by external factors not accounted for, the standard error estimates produced by the ANOVA become biased, fundamentally corrupting the statistical foundation of the model and leading to invalid hypothesis testing.

How to Determine if Independence is Met

Verification of the independence assumption relies primarily on a methodological and logical review of the experimental design, rather than a statistical test performed on the data itself. The most reliable way to ensure that observations are independent across subjects is to confirm that participants were selected from the target [population](#) using a robust and unbiased [simple random sample](#) method.

If the researcher can confirm that a meticulous, unbiased random sampling procedure was followed during recruitment, and that data collection procedures minimized interaction or shared environments among participants, it is generally safe to conclude that the observations are independent across subjects.

What to Do if Independence is Violated

Unfortunately, standard statistical remedies for dependency are extremely limited once the data has been collected. If the assumption of independence is severely violated due to a fundamental flaw in the experimental design--such as non-random assignment or unavoidable subject interaction--the most appropriate and often only ethical remedy is to recognize the limitations of the current data. In many cases, the only path forward is to redesign the study and recruit a new sample of individuals using rigorous random sampling and controlled data collection to guarantee true independence.

Assumption 2: Normality of the Residuals

The second core assumption of the [RM-ANOVA](#) requires that the distribution of the dependent variable be [normally distributed](#) within the population from which the samples were drawn. More precisely, the assumption applies to the residuals (the differences between the observed values and the predicted values of the model).

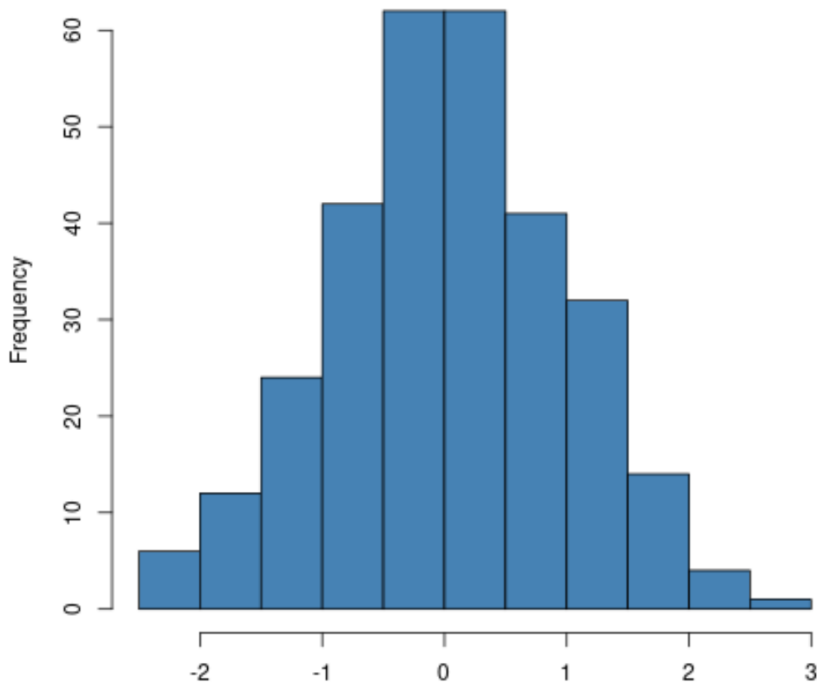
While the RM-ANOVA is recognized for being reasonably **robust** against minor deviations from normality, particularly when the sample size is large (due to the Central Limit Theorem), extreme non-normality--such as severe skewness or high kurtosis--can distort the results, increasing the risk of Type I or Type II errors. Therefore, researchers must always assess the distribution both visually and statistically.

How to Determine if Normality is Met

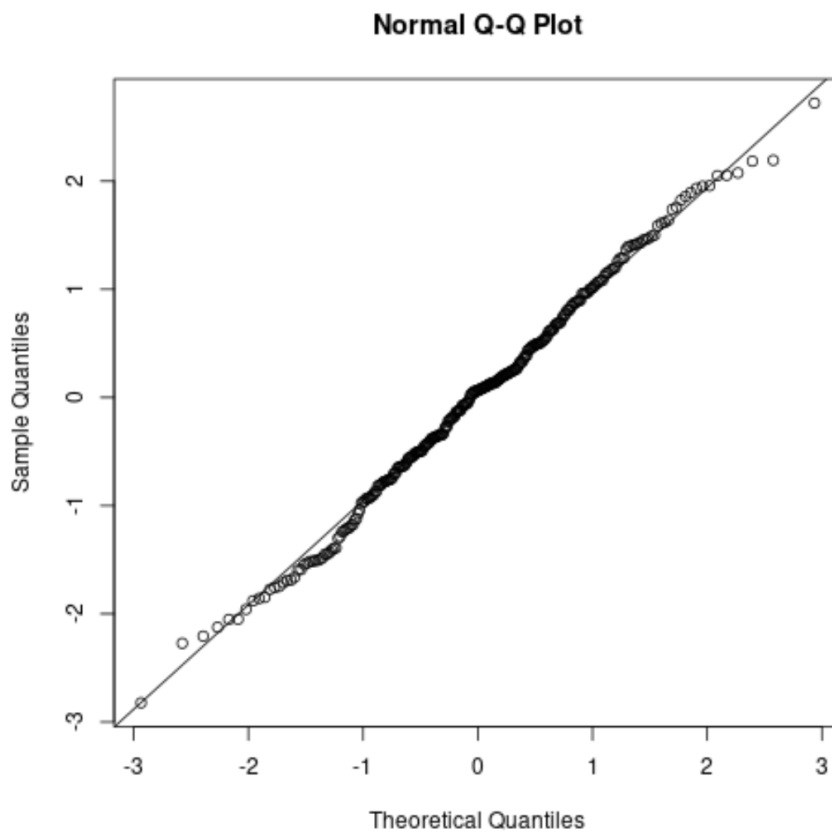
Researchers typically employ a combination of visual inspection and formal statistical testing to verify the normality assumption across all levels of the within-subjects factor.

1. Visual Assessment (Histogram or Q-Q Plot)

Visual checks offer an intuitive understanding of the data's shape. Creating a [histogram](#) allows for a quick verification of whether the response variable follows the characteristic symmetrical "bell" shape of a normal distribution. If the distribution appears roughly symmetrical and unimodal, the assumption is often considered acceptable.



Alternatively, a [Q-Q plot](#) provides a more direct comparison by plotting the quantiles of your observed data against the theoretical quantiles of a perfect normal distribution. If the majority of the data points closely align along the straight diagonal line, the assumption of normality is met:



2. Formal Statistical Testing

To provide an objective, quantifiable measure, formal statistical tests for normality, such as the [Shapiro-Wilk test](#) or the Kolmogorov-Smirnov test, can be performed. In these tests, the null hypothesis (H_0) states that the data is normally distributed. If the resulting [p-value](#) is less than the predetermined [significance level](#) (e.g., $\alpha = .05$), the null hypothesis is rejected, indicating that the data is non-normal.

It is important to interpret formal normality tests cautiously, especially with extremely large sample sizes. Large samples can cause these tests to flag even minor, statistically insignificant deviations from perfect normality. For this reason, many statisticians advocate for prioritizing the visual inspection of graphs like histograms and Q-Q plots over strict reliance on the p-value of a formal test.

What to Do if Normality is Violated

If the violation of normality is severe enough to risk compromising the analysis, the researcher has two principal corrective paths to ensure the validity of the final results:

Perform [data transformation](#): Applying mathematical functions (e.g., logarithmic, square root, or

reciprocal transformations) can often adjust the distribution of the residuals so that it more closely approximates normality.

Utilize a non-parametric alternative: If transformation is unsuccessful or inappropriate, the researcher should employ an equivalent non-parametric statistical test, such as the [Friedman Test](#). Non-parametric methods are distribution-free, meaning they do not assume that the data follows a specific distributional shape.

Assumption 3: Sphericity

The assumption of [sphericity](#) is unique to within-subjects or repeated measures designs. It is a critical requirement stating that the variances of the differences between all possible pairs of within-subject conditions must be equal. Essentially, if you calculate the difference score between condition A and B, B and C, and A and C, the variances of these difference scores must be homogenous.

When this assumption is violated (a condition known as non-sphericity), the standard RM-ANOVA calculation becomes biased, leading to an inflated Type I error rate. This means the calculated F-ratio is overestimated, increasing the likelihood that the researcher will incorrectly reject the null hypothesis and conclude that a treatment effect exists when it does not.

How to Determine if Sphericity is Met

To formally test the homogeneity of variances of the differences, researchers must perform [Mauchly's Test of Sphericity](#). This test evaluates the data matrix against the sphericity requirement using the following set of hypotheses:

H₀ (Null Hypothesis): The variances of the differences are equal (Sphericity is met).

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

If the p-value resulting from Mauchly's test is less than the chosen alpha level (typically $\alpha = .05$), we must reject the null hypothesis and conclude that the assumption of sphericity has been violated. Conversely, if the p-value is greater than or equal to the significance level, we fail to reject the null hypothesis, confirming that the data satisfies the sphericity assumption.

Statistical software output for this test commonly displays results similar to the following example:

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

In this specific scenario, since the p-value is greater than .05, the researcher would fail to reject the null hypothesis, confirming that the assumption of sphericity holds and no correction is necessary.

What to Do if Sphericity is Violated

If Mauchly's test indicates a significant violation of sphericity, the standard procedure is to apply a correction factor to the degrees of freedom used in the RM-ANOVA calculation. These corrections adjust the degrees of freedom downward, which results in a more conservative F-test and a higher p-value, thereby mitigating the risk of Type I error caused by non-sphericity.

Three common correction methods are implemented by statistical software, based on the severity of the violation (measured by the epsilon value):

[Huynh-Feldt](#) correction: Often recommended when the estimated epsilon value is close to 1 (generally the least conservative adjustment).

[Greenhouse-Geisser](#) correction: Typically employed when the epsilon value is less than .75, as it provides a robust, though more conservative, adjustment.

Lower-bound correction: This is the most conservative adjustment and is used when the violation is extremely severe.

Researchers must utilize the corrected p-values provided by these methods to make the final, valid determination regarding the rejection or failure to reject the null hypothesis of the [RM-ANOVA](#).

Additional Resources for Repeated Measures ANOVA

For further exploration of this essential statistical technique, the following tutorials offer additional detailed information and practical implementation guides: