

How to Check the Three Assumptions for a Paired T-Test

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Check the Three Assumptions for a Paired T-Test*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103293>

The paired t-test is a fundamental statistical tool used when researchers need to compare the means of two related groups or measurements taken from the same individuals at different times (e.g., before and after an intervention). This powerful hypothesis test analyzes the differences between the paired observations, rather than the raw scores themselves.

For the results derived from the paired samples t-test to be statistically valid and reliable, several underlying assumptions regarding the nature and distribution of the data must be rigorously satisfied. Failure to meet these criteria can lead to misleading conclusions, potentially invalidating the entire study. Understanding and verifying these assumptions is therefore an essential step in conducting any statistical analysis involving dependent samples.

The three core assumptions center on the relationship between observations, the shape of the distribution of the difference scores, and the presence of unusual data points. This detailed guide will explore each assumption, explaining why it is necessary, how to confirm if your data meets the requirement, and the appropriate remedial actions to take if a violation occurs.

The Critical Role of Statistical Assumptions

A paired samples t-test is specifically designed to compare the means of two samples where each observation in one sample is intrinsically linked, or paired, with an observation in the other sample. Common scenarios include measuring performance pre-treatment versus post-treatment, or comparing husband and wife scores on an anxiety scale.

Since the t-test relies on parametric methods, it makes specific demands on the underlying data distribution. When these conditions are met, the test statistic follows the theoretical t-distribution, allowing for accurate p-value calculations. The primary assumptions governing this type of analysis are detailed below:

Independence of Paired Differences: The difference score calculated for one pair must be independent of the difference score calculated for any other pair.

Normality of Differences: The population distribution of the difference scores between the pairs must be approximately normally distributed.

Absence of Extreme Outliers: There should be no extreme outliers present within the difference scores, as these can drastically skew the mean and standard deviation.

If one or more of these fundamental assumptions are violated, particularly those concerning independence or severe non-normality in small samples, the statistical inference drawn from the paired samples t-test may be fundamentally unreliable or highly misleading. The subsequent sections provide an in-depth explanation of how to assess and manage each of these critical conditions.

Assumption 1: Independence of Observations

While the paired observations within a single unit (e.g., pre-test score and post-test score for Subject A) are inherently dependent, the core requirement of the first assumption is that each individual **pair of observations** must be independent of every other pair. In practical terms, Subject A's difference score must not influence Subject B's difference score. This independence ensures that we are accurately measuring the variability inherent in the population, rather than effects caused by interconnected or correlated experimental units.

Violation of independence is one of the most serious errors in statistical testing, as it fundamentally undermines the calculation of standard error, which is crucial for determining the t-statistic and associated p-value. If data points are related across pairs--for example, if participants were sampled from the same family unit and their scores influenced each other--the true degrees of freedom are inflated, making the test overly optimistic about its findings.

How to Check the Independence Assumption

Unlike assumptions related to distribution shape (like Normality), the assumption of independence is typically assessed through careful examination of the study design and data collection process, rather than through numerical data analysis. The easiest and most reliable way to check this assumption is to verify that each observation pair was collected using a rigorous methodology, such as random sampling.

If a robust random sampling method was utilized--such as simple random sampling or stratified random sampling applied to the experimental units--then it is generally safe to assume that each pair is independent of every other pair. Researchers should document the sampling procedure clearly to demonstrate this independence. If the data collection was non-random or involved clustering (e.g., measuring all students in three specific classrooms), dependence is a serious concern.

Addressing Violations of Independence

If the assumption of independence is violated, the results of the paired samples t-test are considered completely **invalid**. Since independence relates directly to the fundamental structure of the data collection, statistical transformation or non-parametric alternatives cannot typically remedy the issue. The bias introduced by non-independence cannot be corrected mathematically in the context of the t-test.

In this difficult scenario, the most statistically sound course of action is to suspend the analysis and, if possible, collect an entirely new dataset using a proper random sampling method. This ensures that the relationship between subjects is minimized, confirming that each observation pair

is independent and that the subsequent statistical inferences will be trustworthy. If new data collection is infeasible, researchers must acknowledge the limitation and may need to resort to more complex statistical models, such as mixed-effects models, designed to account for nested or dependent data structures.

Assumption 2: Normality of the Difference Scores

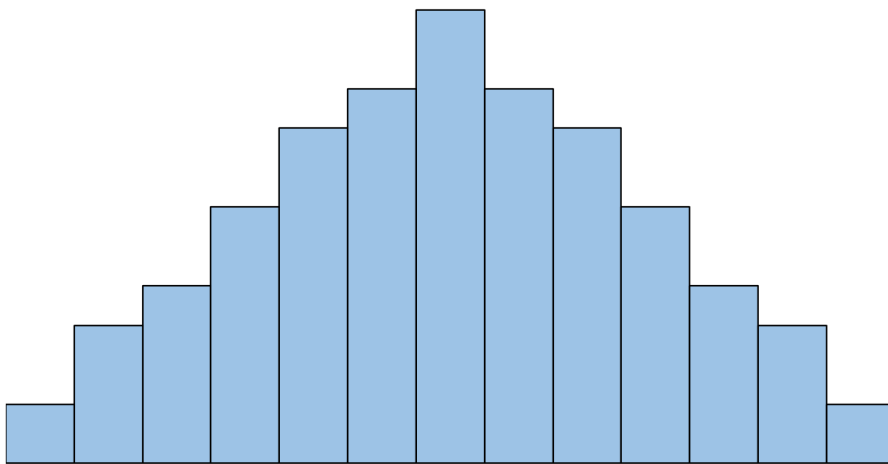
The paired samples t-test does not assume that the raw scores (e.g., the Pre-test scores themselves) are normally distributed. Instead, it assumes that the population distribution of the **difference scores** (Post minus Pre) should be approximately normally distributed. This is a crucial distinction. The t-test relies on the Central Limit Theorem (CLT); however, for the t-statistic itself to follow the appropriate distribution, the underlying differences must be reasonably bell-shaped, especially when dealing with small sample sizes (typically $N < 30$).

If the sample size is large ($N > 30$), the t-test is relatively robust against minor to moderate violations of normality due to the CLT, meaning the sampling distribution of the mean difference will still approach normality regardless of the population distribution shape. However, in smaller samples, a substantial deviation from normality--such as high skewness or kurtosis--can significantly alter the Type I error rate, making the results unreliable. Therefore, checking this assumption is mandatory.

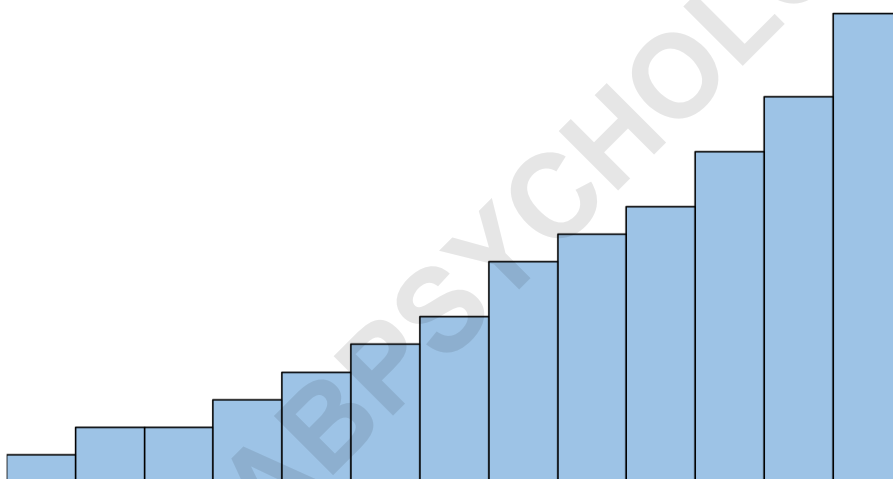
Visualizing Normality via Histograms

The most straightforward and accessible method for checking the normality assumption is to calculate the difference scores for all pairs and then generate a histogram of these paired differences. We then visually inspect the histogram to determine whether or not it exhibits the characteristic symmetric bell shape of a normal distribution.

For example, if the histogram of the difference scores is roughly symmetric, unimodal, and centered, exhibiting a bell shape, then we would typically conclude that the normality assumption is reasonably met:



Conversely, if the histogram shows severe skewness (a long tail on one side) or multiple distinct peaks (bimodal distribution), the normality assumption is likely violated. For example, a histogram indicating strong positive skew would look something like this, suggesting a violation:



Handling Normality Violations

If the normality assumption is clearly violated and the sample size is small, relying on the parametric paired t-test is risky. In such cases, one powerful alternative is to perform a non-parametric test. The appropriate non-parametric equivalent to the paired samples t-test is the Wilcoxon Signed-Rank Test.

The Wilcoxon Signed-Rank Test works by analyzing the ranks of the differences, rather than the raw difference scores, meaning it does not make the assumption that the paired differences are normally distributed. It is particularly useful when dealing with ordinal data or when the continuous data exhibit severe non-normality. Alternatively, if the violation is minor, data transformation (such

as a logarithm or square root transformation) might normalize the distribution, allowing the researcher to proceed with the paired t-test on the transformed data.

Assumption 3: Absence of Extreme Outliers

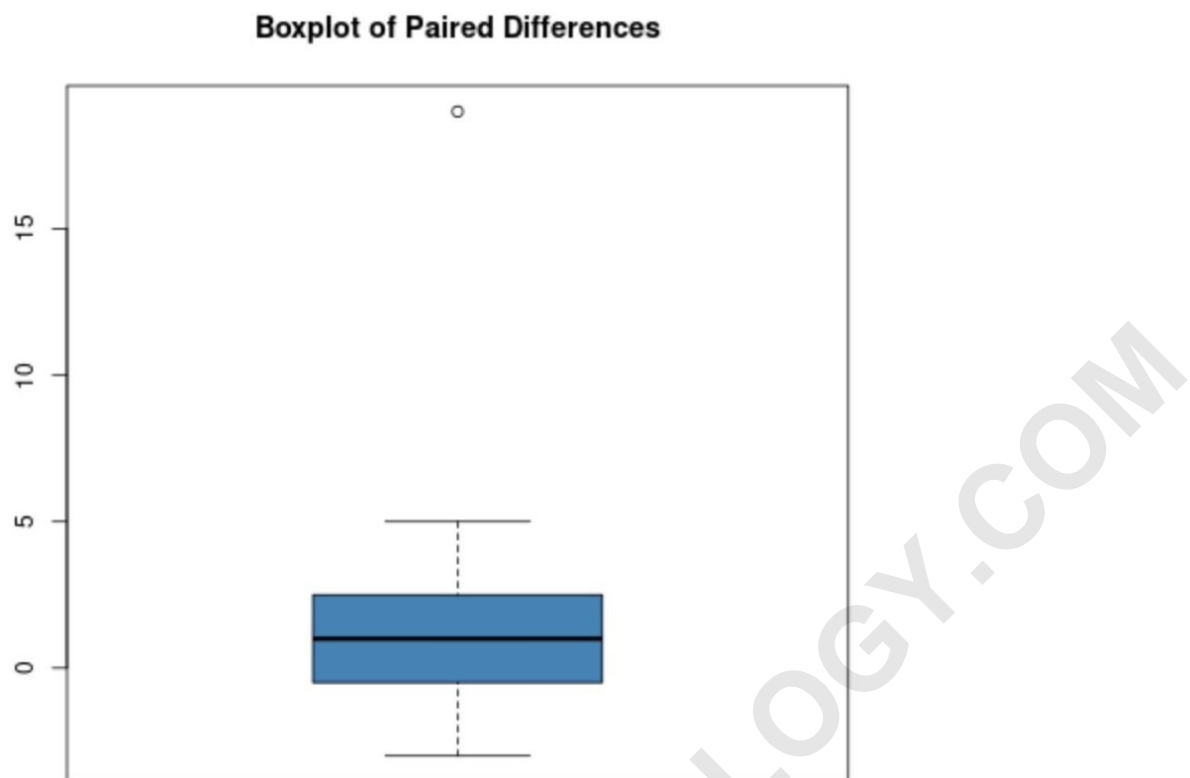
The third essential assumption for a valid paired t-test is that the data, specifically the difference scores, should not contain any extreme outliers. An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In the context of the paired t-test, an outlier represents a pair whose difference score is disproportionately large compared to the rest of the sample.

The presence of outliers is problematic because the t-test is based on the mean and standard deviation, both of which are highly sensitive to extreme values. A single, large outlier can drastically inflate the variance (standard deviation) and skew the mean difference, leading to an artificially larger standard error. This inflated standard error may result in a non-significant finding even when a true effect exists, or conversely, a Type I error if the outlier falls just outside the expected range and pushes the mean significantly.

Identifying Outliers Using Boxplots

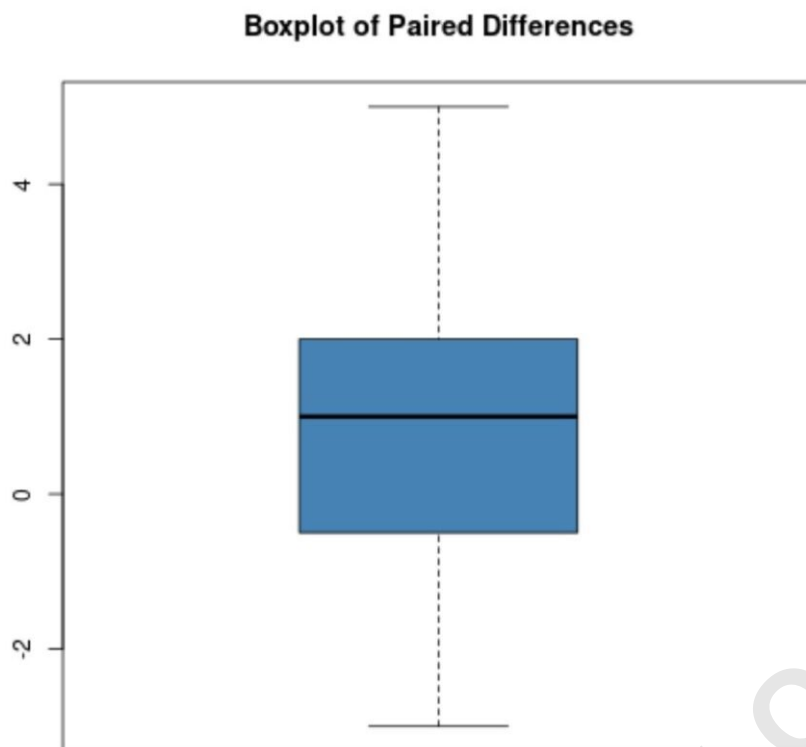
The most effective and commonly used graphical method for checking for outliers in the difference scores is the creation of a boxplot. A boxplot visualizes the quartile range (Interquartile Range, or IQR) of the data, and observations falling substantially outside 1.5 times the IQR are flagged as potential outliers.

For instance, suppose we generate a boxplot of the paired differences that appears as follows:



In this example, the bulk of the paired differences cluster around the mean (zero), but there is one paired difference marked by a circle, located at a value of approximately 19. This point lies far beyond the whisker boundaries and is a clear outlier. A circle is typically used in a boxplot to indicate an outlier value.

However, suppose the boxplot of paired differences looked like this instead:



There are no clear outliers in this boxplot so we would assume that there are no extreme outliers in the data.

Strategies for Managing Outliers

If this assumption is violated, the results of the paired samples t-test could be unusually affected by the presence of the extreme value. Dealing with outliers requires careful consideration and transparency in reporting.

One approach is to investigate the source of the outlier. If you suspect that the extreme value represents a faulty data point, perhaps due to measurement error, transcription error, or equipment malfunction, it may be justifiable to remove the outlier from the analysis. However, removal must always be documented and justified transparently in the final report.

Alternatively, if the outlier is deemed genuine--a true but unusual observation--you can keep the outlier in the dataset but perform both the standard paired t-test and the non-parametric Wilcoxon Signed-Rank Test. By comparing the results, you can assess the robustness of your findings. If both tests yield similar conclusions, the outlier may not have critically biased the results. If the results differ significantly, it is best practice to report the findings of the non-parametric test, as it is less sensitive to extreme values.

Further Reading on Statistical Tests

Understanding these three assumptions--Independence, Normality of Differences, and Absence of Extreme Outliers--is vital for ensuring the reliability of any research utilizing the paired t-test. When these foundations are solid, researchers can draw strong, defensible conclusions about the effect or difference observed between two related measurements.

The following tutorials explain the assumptions made in other statistical tests:

ARABPSYCHOLOGY.COM