

How to Perform a Wilcoxon Signed-Rank Test in Statistics

Authored by
stats writer

January 21, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform a Wilcoxon Signed-Rank Test in Statistics*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=126950>

The **Single Sample Wilcoxon Signed-Rank Test** is a powerful non-parametric statistical method designed for situations where traditional parametric tests, like the one-sample t-test, are inappropriate. Its primary function is to assess whether the median of a single sample group significantly deviates from a pre-specified or hypothesized population value. This test is crucial in quantitative research when the assumption of data normality cannot be met, or when the sample size is critically small.

As a non-parametric approach, this test does not rely on the assumption that the data follows a bell-shaped distribution. Instead of analyzing raw data means, the Wilcoxon Signed-Rank Test operates on the differences between the sample observations and the hypothesized median, converting these differences into ranks. By comparing the sum of positive ranks to the sum of negative ranks, the test determines if the observed difference is statistically significant.

The application of this test spans numerous fields, including clinical trials, where researchers might use it to compare observed outcomes against an established standard, or in quality control, where a sample batch must be compared against a target median value. It is particularly valued for its robustness when dealing with continuous variables that exhibit severe skewness or contain outliers that would disproportionately influence parametric results.

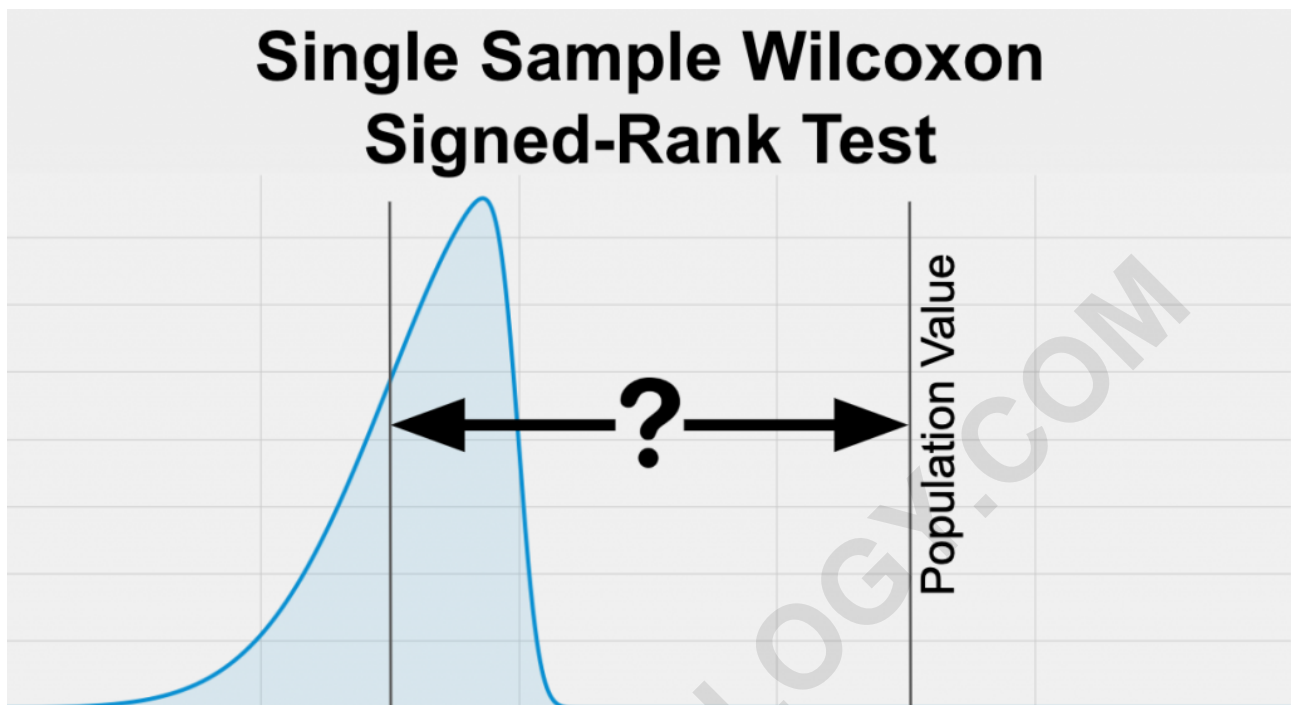
What is a Single Sample Wilcoxon Signed-Rank Test?

The **Single Sample Wilcoxon Signed-Rank Test** serves as the non-parametric alternative to the one-sample t-test. It is specifically employed when a researcher seeks to ascertain if the central tendency--represented by the median--of a single sample significantly differs from a pre-established or theoretical population median value. This statistical procedure is indispensable when the variable of interest is measured on a continuous scale but the underlying distribution is severely **skewed**, meaning the data points are clustered towards one end of the distribution rather than symmetrically grouped around the center.

Unlike parametric tests that require the calculation of means and standard deviations based on the assumption of normality, the Wilcoxon test works by ranking the absolute differences between each sample score and the hypothesized population median. By performing this ranking procedure, the test manages to bypass the stringent requirement for a normal distribution, making it highly suitable for data sets commonly encountered in fields such as psychology, economics, or environmental science, where data often violates normality assumptions.

To ensure the validity of the results derived from this test, the variable under examination must be **continuous**--that is, capable of taking on any value within a given range. Furthermore, while the test can handle small samples better than its parametric counterparts, it is generally recommended to have more than five data values, although studies aiming for high statistical power, especially when expecting small effects, may necessitate considerably larger sample sizes, often exceeding

200 observations.



The *One Sample Wilcoxon Signed-Rank Test* is also called the *One Sample Wilcoxon Test*, *Single Sample Wilcoxon Test*, *One Sample Wilcoxon Sign Test*, and the *Single Sample Wilcoxon Sign Test*.

Prerequisites and Assumptions for a Single Sample Wilcoxon Signed-Rank Test

Every inferential statistical methodology operates under specific prerequisites, known as assumptions. These assumptions represent properties that the data must satisfy for the statistical outcomes and associated probability calculations to be accurate and reliable. Violating a critical assumption can lead to misleading conclusions or inflated Type I or Type II error rates. For the Single Sample Wilcoxon Signed-Rank Test, the assumptions are fewer than those required for parametric tests, reflecting its non-parametric nature.

The fundamental requirements for successfully applying this test include:

- Continuous Measurement Scale
- Simple Random Sample Selection

It is imperative that researchers carefully evaluate their data against these requirements before proceeding with the analysis, as the validity of the resulting inference hinges on their fulfillment. We

will now explore each of these critical assumptions in detail, emphasizing their importance in generating robust statistical conclusions.

The Requirement of a Continuous Dependent Variable

The variable that is the focus of the comparison (the one you are testing for difference between your sample and the population) must be measured on a **continuous scale**. A continuous variable is characterized by its ability to assume any value within a given interval, allowing for infinite precision, although measurement devices often limit this precision in practice. This scale is vital because the Wilcoxon test relies on calculating differences and then ranking the magnitude of these differences, a process that loses meaning if the variable is purely categorical or ordinal.

Examples of variables that satisfy the continuous requirement include physical measurements such as age, weight, height, and time, as well as calculated metrics like standardized test scores or aggregate survey scores based on interval or ratio scales. These variables offer sufficient granularity for the ranking procedures inherent in the Wilcoxon Signed-Rank Test.

If the variable you are analyzing is not continuous but falls into a discrete or categorical classification, the Single Sample Wilcoxon Signed-Rank Test is unsuitable. Ensuring that the variable is truly continuous is the first step in confirming the methodological appropriateness of the test.

If the variable of interest represents a proportion (e.g., comparing 48% male vs 56% female voters) and you possess sufficient observations (typically exceeding 5 per group), the appropriate method is the One-Proportion Z-Test. Conversely, if your variable is a proportion but the sample size is small (less than 5 in a group), you should utilize the Exact Test of Goodness of Fit.

Ensuring a Simple Random Sample

The second fundamental assumption dictates that the data points comprising your sample must be derived from a **simple random sample**. This implies that every individual or data unit within the relevant population had an equal chance of being selected for your study. Random sampling is not merely a procedural step; it is a critical safeguard against selection bias, which can fundamentally undermine the generalizability and accuracy of your statistical findings.

For instance, if you are attempting to compare the average reaction time of a sample group against a known population benchmark, every participant in your sample must have been randomly selected from the target population. If selection criteria are biased--perhaps only selecting volunteers who are easily accessible--the sample group may not accurately represent the broader population, leading to flawed statistical inferences.

When the sample is not randomly determined, the results are susceptible to **bias**, a systematic tendency to produce incorrect estimates. While non-random sampling methods (like convenience sampling) are sometimes unavoidable in practice, researchers must acknowledge that such limitations severely restrict the scope of conclusions that can be drawn from the analysis. Ideally, adherence to simple random sampling ensures that the statistical relationship observed in the sample can be reliably extrapolated to the entire population.

If your data consists of paired samples (two measurements collected from the same subjects, such as pre-test/post-test scores), you must select a different approach. Use a Paired Samples T-Test if your dependent variable follows a normal distribution, or the standard Wilcoxon Signed-Rank Test (not the single-sample version) if the variable is skewed. If the goal is to compare two distinct groups instead of comparing one sample to a known population value, employ an Independent Samples T-Test for normally distributed data, or the Mann-Whitney U Test for skewed data.

Adequate Sample Size Considerations

While the Wilcoxon Signed-Rank Test is appropriate for smaller data sets compared to parametric tests, the minimum acceptable sample size is dependent upon the statistical power required for the study. A fundamental rule of thumb suggests that the sample size (N) should be greater than 5 within the group being tested. However, researchers must move beyond this minimum threshold and consider the expected **effect size**--the magnitude of the difference they anticipate finding between the sample median and the population median.

The required sample size is inversely related to the expected effect size. If you anticipate a large, readily detectable difference, you can achieve sufficient statistical power with a relatively small sample (e.g., N approx 19). Conversely, if the anticipated difference is small or subtle, a much larger sample is essential to detect the effect reliably and avoid a Type II error (failing to detect a real difference). For detecting small effects, sample sizes often need to exceed 200 observations to maintain conventional power levels (e.g., 0.80).

Researchers commonly use power analysis software, such as G*Power, to calculate the optimal sample size based on expected effect size (often referenced using Cohen's D or related metrics), the desired alpha level (e.g., 0.05), and the target power (e.g., 0.80). Proper planning for sample size ensures that the experiment is capable of providing a meaningful and statistically sound result, regardless of whether the final decision is to reject or fail to reject the null hypothesis.

Single Sample Wilcoxon Signed-Rank Test	
Effect Size	Sample Size Needed*
Small	208
Medium	35
Large	19

*sample size

calculation was conducted in G*Power with a power of 0.80, critical value (alpha) of 0.05, and 0.20, 0.50, and 0.80 used as the effect size values for small, medium, and large Cohen's D effect sizes respectively

If your sample size is greater than 30 (and you know the average and standard deviation or spread of the population values), you should run a Single Sample Z-Test, provided that your variable of interest is reasonably normally distributed.

Determining Applicability: When to use the Single Sample Wilcoxon Signed-Rank Test

Selecting the appropriate statistical test is a critical step in any analysis. The Single Sample Wilcoxon Signed-Rank Test is the methodology of choice when a precise set of research and data characteristics align. It is essential to confirm that your research question explicitly seeks a comparison of difference and that your data structure meets the specific scale and distributional criteria required by this non-parametric procedure.

You should use a Single Sample Wilcoxon Signed-Rank Test if your analysis satisfies all four of the following conditions:

You seek to determine if a single group is **statistically different** from a known or hypothesized population median.

Your variable of interest is measured on a **continuous scale**.

The analysis involves only **one sample group**.

The underlying distribution of your variable of interest is **skewed** (non-normal).

Clarifying each of these points ensures that the statistical tool used is correctly matched to the nature of the data and the research question, thereby maximizing the validity of the statistical inference.

Testing for Difference in Medians

The fundamental purpose of this test is centered on determining a **difference**. Specifically, you are looking for evidence that the median measure of your observed sample deviates significantly from a known population median or a theoretical benchmark. This type of analysis contrasts sharply with other statistical goals, such as examining the relationship between two variables (which requires correlation or regression analysis) or predicting one variable based on another (which demands predictive modeling techniques).

If your research question can be framed as: "Is the median score of our experimental group statistically different from the established population standard of X?"--then a difference test, and specifically the Single Sample Wilcoxon Signed-Rank Test, is warranted, provided the other assumptions are met.

Understanding Continuous Data Types

As previously established, the measurement scale of your dependent variable must be continuous. This includes data measured at the interval level (where differences are meaningful, but there is no true zero, e.g., temperature in Celsius) or the ratio level (where differences and ratios are meaningful due to the presence of a true zero, e.g., height, weight, or time). The ability to calculate meaningful differences is what allows the Wilcoxon procedure to rank the observations effectively.

It is important to distinguish continuous data from data types that cannot be used with this test. Excluded data types include ordered or ordinal data (e.g., ranks in a competition, Likert scales treated purely as ordinal), nominal or categorical data (e.g., gender, eye color), and binary data (e.g., presence or absence of a condition). Using these inappropriate data types will render the results of the Wilcoxon test statistically meaningless.

The Single-Sample Focus

The designation "Single Sample" explicitly defines the experimental design this test addresses. It is strictly limited to scenarios where data from only **one group** is collected and subsequently compared against an external, known population value. This population value often acts as a control benchmark or a historical standard against which the sample's performance is measured.

The test is not designed for comparing two or more distinct, independent groups against each other, nor is it suitable for comparing paired measurements taken from the same subjects (e.g.,

before and after an intervention). For analyses involving multiple groups, alternative statistical models must be chosen.

If your research design involves three or more independent groups, you should utilize a One Way Anova analysis if your dependent variable is normally distributed, or a Kruskal-Wallis One-Way ANOVA if your variable exhibits skewness. For comparisons between exactly two independent groups, employ an Independent Samples T-Test for normally distributed data, and the Mann-Whitney U Test if the data is skewed.

Handling Non-Normal (Skewed) Distributions

The most defining characteristic dictating the use of the Wilcoxon Signed-Rank Test over a t-test is the distribution of the variable of interest. If the data is **skewed**, meaning the frequency histogram shows a pronounced tail leaning significantly to the left (negative skew) or the right (positive skew), the assumption of normality required by parametric tests is violated. When this occurs, the sample mean becomes an unreliable measure of central tendency, and the resulting t-test statistics may be inaccurate.

By employing a ranking mechanism, the Single Sample Wilcoxon Signed-Rank Test mitigates the influence of extreme scores and outliers that often cause skewness. It focuses on the positional relationship of the scores relative to the hypothesized median rather than the distance of the scores from the mean, making it a robust alternative when dealing with non-normally distributed data.

If you collect measurements from the same group of students both before and after an intervention (pre-test and post-test), this constitutes paired data. In this scenario, you would need to use a Paired Samples T-Test if the difference scores are normally distributed, or the standard Wilcoxon Signed-Rank Test (the paired version) if the difference scores are skewed.

Applied Example: Utilizing the Single Sample Wilcoxon Signed-Rank Test

To illustrate the practical application of this method, consider a clinical research study investigating a new medical treatment protocol. The researchers need to determine if this new intervention significantly reduces the recovery time compared to the historically known recovery time for patients without the treatment.

Defining the Research Variables

In this clinical scenario, we are investigating the efficacy of an experimental medical treatment

designed to reduce the recovery time from a specific disease. To structure this statistical inquiry, we first define our key components:

Group 1 (Treatment Group): Individuals who have received the experimental medical treatment.

Population Value (Benchmark): The established population median recovery time, which is known to be 12 days for patients who did not receive this specific treatment.

Variable of Interest (Dependent Variable): The time required, measured in days, for the patient to fully recover from the disease. This is a continuous variable.

The core objective is to determine if the median recovery time for the treatment group is statistically less than the established population median of 12 days.

Formulating the Hypotheses and Rationale

The statistical framework requires the formulation of a null hypothesis (H_0), which posits that the experimental treatment has no significant effect. In this context, H_0 suggests that the median recovery time for Group 1 is statistically equivalent to the population benchmark (12 days). Conversely, the alternative hypothesis (H_1) argues that the treatment group's median recovery time is significantly shorter than the population median.

The choice of the Single Sample Wilcoxon Signed-Rank Test is justified because, upon inspection of the data distribution for the recovery times, we observe significant skewness. This violation of the normality assumption, which is prerequisite for the parametric Single Sample T-Test, mandates the use of this robust non-parametric alternative. The Wilcoxon test achieves its robustness by utilizing the ranks of the differences between the sample values and the hypothesized median, rather than the raw data itself.

Interpreting the Statistical Outcome (P-Value)

Once the experiment is complete and the analysis is executed, the primary output we focus on is the p-value. This value represents the probability of observing our sample results (or results more extreme than ours) if the null hypothesis were true--meaning, if the treatment truly did nothing. It is a fundamental metric in hypothesis testing.

In conventional scientific research, a p-value threshold (alpha level), typically set at 0.05, is used for decision-making. If the calculated p-value is less than or equal to 0.05, the result is considered statistically significant. This low probability suggests that the observed difference between the sample median and the population median is highly unlikely to have occurred merely by chance. Consequently, we reject the null hypothesis and conclude that the experimental treatment successfully shortened the recovery time.