

Learning the Wilcoxon Signed-Rank Test with R: A Practical Guide

Authored by
Mohammed loot

November 8, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning the Wilcoxon Signed-Rank Test with R: A Practical Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13081>

The Wilcoxon Signed-Rank Test: A Robust Non-Parametric Alternative

The [Wilcoxon Signed-Rank Test](#) stands as one of the most critical and widely adopted statistical procedures within the realm of [non-parametric statistics](#). It provides a robust and powerful alternative to the conventional [paired t-test](#), particularly when researchers are tasked with analyzing dependent samples. This test is specifically engineered to ascertain whether a statistically significant difference exists between the median scores of two related measurements. These measurements are typically collected from the same subjects or carefully matched pairs, assessing performance under two distinct conditions (e.g., before and after an intervention). Fundamentally, the Wilcoxon test is used to detect a location shift between two populations when those populations cannot be reliably assumed to adhere to a [normal distribution](#), thereby violating the core assumptions necessary for parametric statistical methods to yield reliable results.

A key distinction between this method and its parametric counterpart lies in how it processes the data. While the paired t-test requires the calculation of means and standard deviations of the differences, and critically assumes that these differences are normally distributed, the [Wilcoxon Signed-Rank Test](#) operates exclusively on the **ranks** of the absolute differences. This reliance on ranks, rather than the potentially distorted raw data values, grants the test exceptional resilience against the influence of outliers and makes it perfectly suited for data exhibiting skewed distributions or originating from non-interval scales, such as ordinal measurements. Recognizing when to transition from a paired t-test to this indispensable non-parametric methodology is absolutely crucial for maintaining the validity and integrity of statistical [hypothesis testing](#), ensuring that all conclusions derived from the data are both reliable and methodologically sound.

This comprehensive tutorial is designed to provide both the theoretical foundation and the practical steps required to master the [Wilcoxon Signed-Rank Test](#). We will meticulously detail its proper application, followed by a clear, step-by-step guide on how to execute and correctly interpret the results using the powerful statistical programming language, [R](#). We will utilize a realistic example involving paired observations and clearly demonstrate how to specify the necessary arguments within R's built-in functions to perform both the standard two-sided test and targeted one-sided analyses. For researchers working with data in biology, psychology, or the social sciences, where the assumption of normality is frequently unattainable due to limitations in sample size or inherent data characteristics, mastering this test is an essential professional competency.

Determining Suitability: When Non-Normality Demands a Shift

The decision to deploy the [Wilcoxon Signed-Rank Test](#) is primarily driven by the failure to satisfy the fundamental [normality assumption](#) associated with the [paired t-test](#). The core requirement for the parametric paired t-test is that the distribution of the differences between the paired observations must be approximately normal. When diagnostic procedures, such as visual

inspection of Q-Q plots or formal procedures like the Shapiro-Wilk test, indicate significant non-normality--especially when dealing with smaller sample sizes--the resulting p-values generated by the paired t-test can become severely unreliable. This unreliability can lead directly to inaccurate conclusions regarding the magnitude and significance of the observed effect.

The methodology of the [Wilcoxon Signed-Rank Test](#) expertly circumvents this stringent distributional requirement by converting the raw difference scores into signed ranks. The procedure begins by calculating the difference score for each pair, followed by ranking the absolute values of these differences from smallest to largest. Crucially, the original sign (positive or negative) is then reapplied to its corresponding rank. The resulting **test statistic**, conventionally denoted as **V** (or sometimes **W**), is calculated based on the sum of the positive ranks or the sum of the negative ranks. Under the [null hypothesis](#)--which postulates that there is zero difference between the two measurements--we would anticipate that the sums of the positive and negative ranks would be nearly equal. Conversely, a substantial disparity between these two sums strongly suggests a significant location shift following the intervention or treatment being studied.

Consequently, researchers must always initiate the analysis of dependent data by meticulously assessing the distribution of the difference scores. If the distribution of differences is clearly symmetric and reasonably normal, the [paired t-test](#) is generally the preferred choice due to its slightly superior statistical power. However, if the distribution is markedly skewed, exhibits multiple modes, or if the data scale is inherently ordinal, the [Wilcoxon Signed-Rank Test](#) offers a methodologically appropriate and highly robust alternative. It is an indispensable statistical tool for ensuring methodological integrity when analyzing real-world datasets that often fail to perfectly conform to idealized statistical distributions.

Setting Up the Analysis: A Practical Training Efficacy Scenario

To illustrate the application of this powerful test, let us consider a common practical scenario in sports science. Imagine a basketball coach aiming to rigorously evaluate the effectiveness of a new, intensive training program specifically engineered to enhance players' free-throw accuracy. To ensure a methodologically sound assessment, the coach implements a classical pre-test/post-test experimental design: fifteen players are recruited, and each player records their successful free throws out of twenty attempts **before** commencing the program and again **after** its completion. Because the two sets of measurements (Pre-program and Post-program scores) are collected from the exact same individuals, the data are inherently **paired** or **dependent**.

Initially, the coach considered using a standard [paired difference test](#) (the paired t-test) to evaluate whether the mean number of free throws made had significantly increased. However, a crucial diagnostic step revealed a problem: upon examining the distribution of the difference scores (calculated as Post-score minus Pre-score), the coach determined that the resulting distribution

was significantly non-normal. This clear violation of the underlying assumption required for the parametric paired t-test necessitated an immediate shift to a more conservative, [non-parametric approach](#)--the Wilcoxon Signed-Rank Test. This procedural shift is not arbitrary; it guarantees that the statistical inference remains valid and trustworthy, regardless of the sample data's distributional eccentricities.

The specific dataset collected, which meticulously details the number of successful free throws out of 20 attempts for each of the fifteen players both preceding and following the intensive training regimen, is presented visually below. This organized structure enables us to directly proceed to the statistical implementation in [R](#). We will treat the 'Before' scores and 'After' scores as two linked [vectors](#) of observations that must be analyzed jointly to assess the training program's impact.

Player	Before	After
Player #1	14	15
Player #2	17	17
Player #3	12	15
Player #4	15	15
Player #5	15	17
Player #6	9	14
Player #7	12	9
Player #8	13	14
Player #9	13	11
Player #10	15	16
Player #11	19	18
Player #12	17	20
Player #13	14	20
Player #14	14	10
Player #15	16	17

Executing the Test in R: Mastering the `wilcox.test()` Function

To correctly execute the [Wilcoxon Signed-Rank Test](#) on this dependent dataset within the [R](#) environment, we rely on the robust and highly flexible built-in function: `wilcox.test()`. This single function is designed to handle various forms of the Wilcoxon test. However, when analyzing paired or dependent data, it is absolutely paramount to explicitly include and specify the argument `paired=TRUE`. This crucial specification instructs R that the two input data [vectors](#) represent observations collected from the same units, rather than independent samples.

The general syntax required for conducting the Wilcoxon Signed-Rank Test for dependent samples is remarkably straightforward, typically requiring only three primary components: the two data

vectors containing the observations and the explicit pairing specification. This structure ensures that the ranks are computed based on the differences between the corresponding pairs, as required by the test's methodology.

```
wilcox.test(x, y, paired=TRUE)
```

The parameters utilized within this function are defined as follows, emphasizing the necessity of correct specification:

x, y: These arguments denote the two quantitative [vectors](#) containing the observed data values. In our example, these would correspond to the 'before' scores and the 'after' scores.

paired: Setting this parameter strictly to `TRUE` is non-negotiable for dependent samples. It directs R to internally calculate the differences between the paired elements of \bar{x} and \bar{y} and subsequently execute the signed-rank procedure on those difference scores, precisely adhering to the statistical requirements of a dependent samples test.

The following code block demonstrates the practical implementation of the test. First, we define the 'before' and 'after' data vectors using R's assignment operator (`<-`), and then we execute the two-sided [Wilcoxon Signed-Rank Test](#) using the specified paired syntax. This specific command requests R to test the hypothesis that there is **any** statistically significant difference--a positive or negative location shift--between the median performance across the two measurement conditions.

```
#create the two vectors of data
```

```
before <- c(14, 17, 12, 15, 15, 9, 12, 13, 13, 15, 19, 17, 14, 14, 16)
```

```
after <- c(15, 17, 15, 15, 17, 14, 9, 14, 11, 16, 18, 20, 20, 10, 17)
```

```
#perform Wilcoxon Signed-Rank Test
```

```
wilcox.test(before, after, paired=TRUE)
```

Wilcoxon signed rank test with continuity correction

data: before and after

$V = 29.5$, p-value = 0.275

alternative hypothesis: true location shift is not equal to 0

Interpreting the Two-Sided Results and Drawing Conclusions

The output generated by the `wilcox.test()` function provides two essential pieces of information required for drawing an informed statistical conclusion: the computed [test statistic](#), denoted by **V**, and its corresponding [p-value](#). In the results displayed from our basketball example, the calculated test statistic **V** is **29.5**, and the associated p-value is **0.275**. The output also explicitly reminds us of

the research hypothesis: the "alternative hypothesis: true location shift is not equal to 0," confirming that we successfully executed a two-sided test designed to detect a statistically significant difference in either the positive or negative direction.

To properly interpret these findings, we must compare the calculated [p-value](#) against the predefined significance level, or **alpha (\$alpha\$)**, which is almost universally set at 0.05 for most academic and applied research applications. The statistical decision rule is straightforward and critical for sound inference: if the p-value is less than or equal to \$alpha\$ (0.05), we possess sufficient evidence to reject the [null hypothesis](#) (H_0). Conversely, if the p-value exceeds \$alpha\$, we must conclude that we fail to reject the null hypothesis. In the context of our free-throw example, the null hypothesis posits that there is absolutely no difference in the median free throws made before versus after the intensive training program.

Given that our calculated p-value of **0.275** is substantially larger than the standard significance level of 0.05, our required course of action is to **fail to reject the null hypothesis**. This means that, based on the non-parametric statistical evidence provided by this sample, we do not have adequate statistical support to conclude that the training program caused a significant or systematic change--either an increase or a decrease--in the number of free throws successfully made by the players. Despite some apparent visual improvements in individual pairs, the overall statistical assessment indicates that any observed variability is most likely attributable to random chance or measurement noise, rather than a systematic effect induced by the training intervention.

Utilizing Directional Hypotheses: Performing One-Tailed Tests

While the two-sided test is appropriate for exploratory analyses or when the direction of the effect is uncertain, researchers often opt for one-tailed (or directional) tests when they possess a strong, theoretically derived expectation concerning the specific direction of the outcome. For instance, the basketball coach might have hypothesized exclusively that the rigorous training program could only lead to **increased** performance, explicitly ruling out the possibility of a decrease. In [R](#), specifying a directional test is simple using the `alternative` argument within the `wilcox.test()` function.

There are two primary options for specifying the directional test: `alternative="less"` and `alternative="greater"`. When we specify `alternative="less"`, R performs a left-tailed test, which investigates whether the true location shift is significantly less than zero (i.e., whether the 'after' scores are significantly lower than the 'before' scores). Conversely, specifying `alternative="greater"` executes a right-tailed test, checking if the true location shift is statistically greater than zero (i.e., if the 'after' scores are significantly higher than the 'before' scores, indicating improvement).

It is important for interpretation that the core [test statistic V](#) (which remains 29.5 in our example) is constant across the two-sided, left-tailed, and right-tailed tests, as V is derived exclusively from

the ranks of the absolute differences, independent of the hypothesized direction. However, the calculated [p-value](#) changes substantially because the probability calculation is restricted to assessing the likelihood of the result occurring in only one tail of the sampling distribution. The code below illustrates the execution of both the left-tailed and right-tailed tests on our training dataset, highlighting the resulting differences in the p-values based on the direction of the hypothesized shift.

#perform left-tailed Wilcoxon Signed-Rank Test

```
wilcox.test(before, after, paired=TRUE, alternative="less")
```

Wilcoxon signed rank test with continuity correction

data: before and after

V = 29.5, p-value = 0.1375

alternative hypothesis: true location shift is less than 0

#perform right-tailed Wilcoxon Signed-Rank Test

```
wilcox.test(before, after, paired=TRUE, alternative="greater")
```

Wilcoxon signed rank test with continuity correction

data: before and after

V = 29.5, p-value = 0.8774

alternative hypothesis: true location shift is greater than 0

Analyzing the directional outputs, we see that in the left-tailed test (hypothesizing a decrease, `alternative="less"`), the p-value is 0.1375. Since this value remains above the 0.05 threshold, we find no statistical evidence to support the conclusion that the free-throw scores significantly decreased after the program. Similarly, the right-tailed test (hypothesizing an increase, `alternative="greater"`) yields a p-value of 0.8774, which is also far from statistical significance. This comprehensive analysis confirms the initial two-sided finding: regardless of the specific directional hypothesis tested, the collected data do not provide sufficient evidence to support a statistically significant effect of the training program on the players' free-throw performance when applying the robust [non-parametric](#) Wilcoxon Signed-Rank Test.

Conclusion and Resources for Further Study

The [Wilcoxon Signed-Rank Test](#) represents an essential statistical methodology for the rigorous analysis of paired data, particularly in scenarios where the restrictive assumptions of parametric tests--most critically, the [normality of difference scores](#)--cannot be reasonably met. By focusing its calculations on the ranks and signs of the differences, this test successfully evaluates whether a

given intervention has caused a genuine and significant location shift in the distribution of paired observations.

We have successfully demonstrated the complete implementation pipeline for this crucial test using the powerful `wilcox.test()` function within R, covering the necessary syntax for handling dependent samples and illustrating how to specify both the general two-sided hypothesis and specific directional hypotheses. Correctly understanding the output, particularly the V [test statistic](#) and the corresponding [p-value](#), empowers researchers to make methodologically sound decisions regarding their research questions.

In our basketball efficacy example, the consistent failure to reject the [null hypothesis](#) across all tests strongly suggests that the intensive training program, while perhaps conceptually sound, did not produce a statistically significant improvement in free-throw accuracy across the sampled population under the experimental conditions. This outcome critically underscores the necessity of applying statistical rigor when evaluating real-world interventions, ensuring that conclusions are based on statistically defensible evidence.

Additional Resources

For readers interested in expanding their expertise in [non-parametric statistics](#) or exploring advanced options within the R statistical environment, the following resources are highly recommended for providing excellent additional context and methodological detail.

Statistical Methods for the Social Sciences (Reference for non-parametric methods).

Official R Documentation for the `wilcox.test()` function.

A comprehensive guide to Hypothesis Testing in R.