

# How to Perform a Wilcoxon Signed-Rank Test in R to Compare Paired Samples

Authored by  
**stats writer**

March 13, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Perform a Wilcoxon Signed-Rank Test in R to Compare Paired Samples*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135558>

## Understanding the Wilcoxon Signed-Rank Test in Statistical Analysis

The **Wilcoxon Signed-Rank Test** is a sophisticated **non-parametric** statistical procedure used to determine whether there is a significant difference between two related or paired samples. This test serves as a robust alternative to the **paired Student's t-test**, specifically when the underlying data does not satisfy the stringent assumptions of normality. In many real-world scenarios, researchers encounter data sets that are skewed or contain outliers, making parametric methods unreliable. The Wilcoxon Signed-Rank Test bypasses these issues by focusing on the **ranks** of the differences between pairs rather than the raw values themselves, thereby providing a more resilient measure of central tendency for **ordinal** or interval data.

Within the **R programming language**, this test is implemented via a versatile function that allows for quick and accurate computation. The primary objective is to evaluate the **null hypothesis**, which posits that the median difference between the pairs is zero. By calculating the absolute differences between observations, ranking them, and then re-applying the original signs, the test generates a statistic that reflects the magnitude and direction of the shifts within the sample. This methodology is particularly valuable in **clinical trials**, psychological assessments, and behavioral studies where "before and after" measurements are standard, but the sample size may be too small to justify a **normal distribution** assumption.

When performing the **Wilcoxon Signed-Rank Test** in R, the output provides critical metrics including the test statistic (often denoted as V), the **p-value**, and optional **confidence intervals**. These metrics allow the analyst to make informed decisions regarding the **statistical significance** of their findings. If the resulting p-value is lower than a predefined threshold, typically 0.05, the null hypothesis is rejected in favor of the **alternative hypothesis**, suggesting that the treatment or intervention has a measurable effect. This detailed exploration of the test in R will provide you with the necessary tools to handle non-normal paired data with confidence and precision.

## Theoretical Foundations: Non-Parametric vs. Parametric Tests

To fully appreciate the utility of the **Wilcoxon Signed-Rank Test**, one must understand the distinction between parametric and **non-parametric** statistics. Parametric tests, such as the **paired t-test**, assume that the data follows a specific probability distribution, usually the **normal distribution** (Gaussian distribution). They also assume **homoscedasticity**, or equal variance among groups. However, when these assumptions are violated--such as when the distribution of differences between pairs is heavily skewed or has heavy tails--parametric tests can yield misleading results, potentially leading to **Type I or Type II errors**. The Wilcoxon test provides a safer route because it does not require the data to be normally distributed.

The **Wilcoxon Signed-Rank Test** is specifically designed for dependent or paired samples. This

means that each observation in one group is uniquely linked to an observation in the second group. Common examples include measurements taken on the same individual at two different time points or measurements from matched pairs (e.g., twins). While the t-test compares the means of these differences, the Wilcoxon test compares the **median** of the differences. By using ranks, the test effectively limits the influence of **outliers**, which could otherwise drastically shift the mean and distort the conclusions of a parametric analysis. This makes it an essential tool for data scientists working with real-world datasets that are often "messy" or limited in scope.

Furthermore, the **Wilcoxon Signed-Rank Test** is often preferred when the data is **ordinal**. Ordinal data represents categories with a logical order but without a consistent measurable distance between them, such as Likert scale responses (e.g., "strongly disagree" to "strongly agree"). Since means and standard deviations are not mathematically meaningful for ordinal data, the ranking method employed by the Wilcoxon test is the most appropriate way to determine if a shift in sentiment or performance has occurred. By mastering this test in **R**, you ensure that your statistical toolkit is equipped to handle a diverse range of data types and experimental designs.

## Implementing the `wilcox.test()` Function in R

In the **R** environment, the primary tool for executing this analysis is the **wilcox.test()** function. This function is highly flexible, capable of performing both the **Mann-Whitney U test** (for independent samples) and the **Wilcoxon Signed-Rank Test** (for paired samples). To specify that the data is paired, the user must set the **paired** argument to **TRUE**. Without this specification, R will treat the two vectors as independent, which would be a fundamental error in experimental design and lead to incorrect **p-values**. The syntax is straightforward, typically requiring two numeric vectors representing the paired observations.

The basic syntax for a paired test is **wilcox.test(x, y, paired = TRUE)**. Here, **x** and **y** are the numeric vectors containing the data points for the two conditions. Beyond these primary arguments, the function offers several parameters to refine the analysis. The **alternative** argument allows the user to choose between a two-sided test (default) or a one-sided test ("greater" or "less"). Additionally, the **conf.int** argument can be set to **TRUE** to generate a confidence interval for the **median** difference, and the **conf.level** parameter allows for the adjustment of the confidence threshold, usually set at 0.95.

Another important aspect of the **wilcox.test()** function is its handling of ties and zeroes. When two pairs have the same difference, or when the difference between pairs is zero, the function must apply specific rules to assign ranks. By default, **R** uses a **continuity correction** to account for the fact that the discrete rank distribution is being used to approximate a continuous distribution. Understanding these technical nuances is vital for advanced users who require exact calculations for small sample sizes. The function's ability to provide a "V" statistic--the sum of the ranks for the

positive differences--serves as the foundation for determining whether the **null hypothesis** should be rejected.

## A Practical Case Study: Basketball Performance Analysis

To illustrate the application of the **Wilcoxon Signed-Rank Test**, let us consider a practical scenario involving a basketball coach. The coach implements a new training program aimed at improving the free-throw accuracy of 15 players. To measure the effectiveness of this program, the coach records the number of successful free throws out of 20 attempts for each player both before and after the training period. This "pre-test/post-test" design creates **paired data**, as each "after" score is directly linked to a specific player's "before" score.

Initially, the coach might consider a **paired t-test**. However, upon examining the differences between the before and after scores, the coach discovers that the distribution of these differences is significantly skewed and does not follow a **normal distribution**. Given this violation of parametric assumptions, the coach shifts to the **Wilcoxon Signed-Rank Test**. This ensures that the analysis remains valid despite the non-normality of the data. The goal is to see if the training program led to a **statistically significant** shift in the median number of free throws made.

The following table presents the raw data collected from the 15 players. Each row represents a single player, showing their performance at the two different time intervals:

| 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    |

Using **R** to analyze this data allows the coach to handle the computation of ranks and signs

efficiently. The process involves creating two vectors in the R environment, one for the "before" scores and one for the "after" scores, and then passing them into the `wilcox.test()` function with the `paired=TRUE` parameter. This approach provides a clear, mathematical answer to whether the training program was effective or if the observed changes were merely due to random chance.

## Step-by-Step Data Implementation in R

With our data identified, we can now proceed to the actual coding process in **R**. First, we define the data as two distinct vectors. This is done using the `c()` function, which combines individual values into a single data structure. Once the vectors are created, we execute the `wilcox.test()` function. This function will calculate the differences, rank them based on their absolute values, and then sum the ranks associated with positive differences to find the V statistic. This entire process is automated, ensuring accuracy and consistency in the results.

The code below demonstrates how to initialize the variables and run the **Wilcoxon Signed-Rank Test**. Note how the `paired=TRUE` argument is essential for informing the function that the data points in the "before" vector correspond exactly to the data points in the "after" vector. The output will then display the results of the test, including the p-value, which is the primary metric used for hypothesis testing.

```
#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
```

As we examine the output, we see that the calculated test statistic (V) is 29.5, and the associated p-value is 0.275. In the context of **statistical significance**, we compare this p-value to our **alpha level** (usually 0.05). Since 0.275 is significantly higher than 0.05, we do not have sufficient evidence to reject the **null hypothesis**. Consequently, the coach must conclude that the training program did not produce a statistically significant change in the players' free-throw performance.

## Interpreting the Results and Continuity Correction

The output of the `wilcox.test()` function in **R** includes a note about "continuity correction." This is a technical adjustment made when a discrete **probability distribution** is used to approximate a continuous one. In the case of the **Wilcoxon Signed-Rank Test**, the distribution of the test statistic  $V$  is discrete. However, for larger sample sizes, **R** often uses a **normal approximation** to calculate the p-value. The **continuity correction** helps to make this approximation more accurate by adjusting the test statistic slightly toward the mean of the distribution.

The "V" value reported in the results represents the sum of the ranks for the pairs where the second value was greater than the first. If the two samples were identical, we would expect the sum of the positive ranks to be roughly equal to the sum of the negative ranks. A  $V$  value that is very large or very small relative to the total possible sum of ranks would indicate a significant **location shift**. In our basketball example, the  $V$  value of 29.5, paired with the high p-value, suggests that the ranks were distributed in a way that is highly likely under the null hypothesis of "no change."

It is also important to note the phrase "true location shift is not equal to 0" in the output. This is **R**'s way of describing the **alternative hypothesis** for a two-sided test. A location shift refers to the displacement of the **median** of the differences. By stating that the shift is not equal to zero, we are testing for any difference, regardless of whether it is an improvement or a decline. This comprehensive interpretation is vital for any researcher looking to communicate their findings clearly and accurately to a broader audience.

## Directional Hypotheses: Left-Tailed and Right-Tailed Tests

While the default two-sided test is useful for general exploration, researchers often have a specific direction in mind for their **hypothesis**. For instance, the basketball coach likely expects the training program to *\*increase\** the number of free throws made, rather than just changing them in any direction. In **R**, this is handled by the **alternative** argument within the `wilcox.test()` function. By specifying "less" or "greater," you can perform a one-sided **Wilcoxon Signed-Rank Test**.

A "less" alternative hypothesis tests whether the median of the differences is less than zero, while a "greater" alternative tests whether it is greater than zero. These are known as directional tests. Performing a directional test can increase the **statistical power** of your analysis if the direction of the effect is correctly predicted. However, it should be used with caution and ideally be decided upon before the data is collected to avoid **p-hacking** or bias in the reporting of results.

The code snippet below illustrates how to implement these directional tests in **R** using our existing basketball dataset. By comparing the p-values from these tests to the original two-sided test, you can see how the **statistical significance** changes based on the nature of the alternative

hypothesis:

**#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

## Core Assumptions and Best Practices

Although the **Wilcoxon Signed-Rank Test** is a **non-parametric** test, it is not entirely free of assumptions. To ensure the validity of your results in **R**, several criteria must be met. First, the data must be paired or dependent. Second, the differences between the pairs should be independent of one another. For example, the performance of one basketball player should not influence the performance of another player in the study. Third, the differences should come from a **symmetric distribution**. While the distribution does not need to be normal, it should be roughly symmetrical around its median for the test of the median to be fully valid.

Another best practice is to always visualize your data before running any statistical test. In **R**, you can use **box plots** or **histograms** of the differences to check for symmetry and identify potential **outliers**. If the distribution of differences is extremely asymmetrical, even the Wilcoxon test might struggle, and other methods or data transformations might be necessary. Furthermore, always report the **effect size** alongside your p-value. While the p-value tells you if a difference exists, the effect size tells you how large or meaningful that difference is in a practical context.

In conclusion, the **Wilcoxon Signed-Rank Test** is an indispensable tool for researchers dealing with paired data that fails to meet the assumptions of normality. By using the **wilcox.test()** function in **R**, you can perform this analysis with a high degree of precision, whether you are conducting a simple two-sided test or a more specific directional study. By following the steps and interpretations

outlined in this guide, you can ensure that your statistical conclusions are robust, transparent, and grounded in sound mathematical principles. Always remember to consider the context of your data and the underlying assumptions of the test to achieve the most reliable insights from your research.

ARABPSYCHOLOGY.COM