

# How to Perform a Wilcoxon Signed Rank Test in Stata: A Step-by-Step Guide

Authored by  
**stats writer**

March 8, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Perform a Wilcoxon Signed Rank Test in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134664>

## Perform a Wilcoxon Signed Rank Test in Stata

The **Wilcoxon Signed-Rank Test** represents a foundational **non-parametric** method utilized within the field of **inferential statistics** to compare two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ. It is widely recognized as the non-parametric alternative to the **paired Student's t-test**, specifically designed for scenarios where the data does not meet the stringent assumptions of normality. When researchers encounter **ordinal data** or continuous data that exhibit significant **skewness**, the Wilcoxon Signed-Rank Test offers a robust framework for testing the **null hypothesis** that the median difference between pairs is zero.

In the context of modern data analysis, **Stata** provides a comprehensive suite of commands to execute this test efficiently. This **statistical software** package is highly favored in social sciences, epidemiology, and economics due to its precision and the clarity of its output. By utilizing the **Wilcoxon Signed-Rank Test**, analysts can determine if there is a statistically significant change or difference across two distinct conditions without being hindered by **outliers** that might otherwise bias a parametric test. This tutorial provides an in-depth exploration of how to implement this procedure, interpret the resulting **p-values**, and report the findings in a professional manner.

The underlying logic of this test involves calculating the differences between paired observations, ranking these differences based on their absolute magnitude, and then assigning the original signs to these ranks. By summing the ranks of the positive and negative differences, **Stata** generates a **test statistic** that helps determine the likelihood that the observed differences occurred by chance. This approach ensures that the magnitude of the change is accounted for, rather than just the direction of the change, which distinguishes it from the simpler **Sign Test**.

### Theoretical Foundations of Non-Parametric Paired Tests

Understanding the theoretical framework of the **Wilcoxon Signed-Rank Test** is essential for any researcher aiming to produce valid results. This test is specifically categorized as a **non-parametric statistics** tool, meaning it does not rely on data belonging to any particular **probability distribution**. In many real-world datasets, the assumption of a **normal distribution** is frequently violated, particularly in small sample sizes where **normality tests** such as the **Shapiro-Wilk test** may lack sufficient power to provide definitive guidance. Consequently, opting for a rank-based approach provides a safer and more reliable estimation of differences between paired groups.

The test operates under the primary assumption that the distribution of the differences between the two pairs is symmetric about the median. While it does not require normality, this symmetry assumption is crucial for the interpretation of the test as a comparison of medians. If the distribution is not symmetric, the test is strictly a comparison of **mean ranks**. Researchers must be

diligent in verifying these conditions, as the misuse of **statistical tests** can lead to **Type I errors** or **Type II errors**, potentially undermining the credibility of the entire study. Within **Stata**, these nuances are handled through robust algorithms that calculate the exact distribution of the test statistic whenever possible.

Furthermore, the **paired difference test** design is inherently powerful because it controls for subject-level variability. By comparing the same individual or item under two different conditions, the researcher eliminates the "noise" generated by differences between individuals. This increase in **statistical power** makes the **Wilcoxon Signed-Rank Test** an attractive option for clinical trials, educational assessments, and industrial quality control where subjects are measured before and after an intervention. The methodology remains a cornerstone of **quantitative research**, providing a pathway to **statistical significance** when data characteristics are less than ideal.

## When to Utilize the Wilcoxon Signed-Rank Test Over Parametric Alternatives

Choosing between a **paired t-test** and the **Wilcoxon Signed-Rank Test** is a critical decision in the **data analysis** workflow. The primary driver for selecting the Wilcoxon approach is the violation of the normality assumption in the differences between pairs. If a **histogram** or a **Q-Q plot** reveals that the differences are heavily skewed or contain influential **outliers**, the t-test may provide a misleading **p-value**. In such instances, the Wilcoxon test remains valid because it relies on the ranks of the data rather than the raw values, effectively neutralizing the impact of extreme observations that do not represent the majority of the sample.

Another compelling reason to use this test is when the **level of measurement** for the dependent variable is **ordinal**. Since ordinal scales (such as Likert scales) do not have consistent intervals between points, calculating a mean is technically inappropriate. The **Wilcoxon Signed-Rank Test** handles ordinal data gracefully by focusing on the relative ordering of the observations. This makes it the preferred choice for psychological surveys and consumer preference studies where "greater than" or "less than" relationships are more meaningful than the actual numerical values assigned to responses.

Finally, sample size considerations often dictate the use of non-parametric methods. With very small samples (e.g.,  $n < 15$ ), it is notoriously difficult to assess the distribution of the population. While the t-test is somewhat robust to minor deviations from normality, the **Wilcoxon Signed-Rank Test** provides a more conservative and justifiable approach for small-scale pilot studies. By prioritizing the **median** and rank-order, researchers ensure that their conclusions are derived from the core tendencies of the data rather than being skewed by the mathematical properties of the **arithmetic mean**.

## Step 1: Loading and Inspecting the Dataset in Stata

To perform a **Wilcoxon Signed-Rank Test** in **Stata**, we begin by importing the relevant data. For this demonstration, we will utilize the **fuel** dataset, which is a classic example provided by StataCorp for illustrating paired comparisons. This dataset records the miles per gallon (mpg) of 12 different automobiles before and after receiving a specific fuel treatment. To load this dataset into your active memory, execute the following **command-line** instruction in the Stata Command window:

```
use http://www.stata-press.com/data/r13/fuel
```

Once the data is successfully loaded, it is a **best practice** to perform an initial inspection to understand the structure of the variables. This step ensures that there are no missing values or obvious entry errors that could invalidate the **statistical inference**. Use the **list** command to view the observations directly in the Results window:

**list**

```
. use http://www.stata-press.com/data/r13/fuel
```

```
. list
```

	mpg1	mpg2
1.	20	24
2.	23	25
3.	21	21
4.	25	22
5.	18	23
6.	17	18
7.	18	17
8.	24	28
9.	20	24
10.	24	27
11.	23	21
12.	19	23

In the resulting table, the first column, **mpg1**, represents the performance of the vehicles prior to the intervention, while the second column, **mpg2**, captures the performance after the treatment. Each row corresponds to a single vehicle, maintaining the **paired samples** structure necessary for the test. Observing the raw data allows the researcher to notice any immediate trends, such as

whether most vehicles showed an increase in efficiency, which will later be formally tested for **statistical significance** through the **signrank** command.

## Step 2: Executing the Wilcoxon Signed-Rank Test Command

The core of the analysis involves the **signrank** command, which is specifically designed to perform the **Wilcoxon Signed-Rank Test** on two variables representing paired observations. The syntax in **Stata** is straightforward: you specify the first variable, followed by an equals sign, and then the second variable. This tells the software to calculate the difference (mpg1 - mpg2) for each observation and proceed with the ranking process. Enter the following command to run the analysis:

```
signrank mpg1 = mpg2
```

Upon execution, Stata processes the 12 pairs of data. It identifies how many pairs resulted in a positive difference, how many were negative, and how many resulted in a tie (where the mpg remained unchanged). It then calculates the **sum of ranks** for the positive and negative differences. This detailed breakdown is essential for understanding where the variation in the dataset lies and provides the necessary components for the calculation of the **Z-statistic**.

```
. signrank mpg1=mpg2
```

```
Wilcoxon signed-rank test
```

sign	obs	sum ranks	expected
positive	<b>3</b>	<b>13.5</b>	<b>38.5</b>
negative	<b>8</b>	<b>63.5</b>	<b>38.5</b>
zero	<b>1</b>	<b>1</b>	<b>1</b>
all	<b>12</b>	<b>78</b>	<b>78</b>

```
unadjusted variance    162.50
adjustment for ties    -1.63
adjustment for zeros   -0.25
-----
adjusted variance      160.63
```

```
Ho: mpg1 = mpg2
```

```
z = -1.973
Prob > |z| = 0.0485
Exact Prob = 0.0479
```

The output table produced by Stata is highly informative. It displays the number of observations,

the sum of positive ranks, the sum of negative ranks, and the expected value under the **null hypothesis**. This transparency allows researchers to verify the calculations manually if necessary, although **Stata**'s automated **algorithms** are highly reliable and account for ties by assigning average ranks, ensuring the most accurate **test statistic** possible.

### Step 3: Interpreting the Statistical Output and Hypothesis Testing

Interpreting the output of the **Wilcoxon Signed-Rank Test** requires focusing on the **test statistic** and the associated **p-value**. In the example provided, the output indicates a **z-test** statistic of -1.973. This value represents the number of **standard deviations** the observed sum of ranks deviates from the expected sum under the null hypothesis. A negative Z-value suggests that the ranks associated with one group are systematically lower than the other, providing evidence of a directional shift in the data.

The most critical component for decision-making is the **p-value**, which in this case is 0.0485. In most scientific research, an **alpha level** of 0.05 is used as the threshold for **statistical significance**. Because our calculated p-value (0.0485) is less than 0.05, we possess sufficient evidence to **reject the null hypothesis**. The null hypothesis ( $H_0$ : mpg1 = mpg2) posits that there is no difference between the two groups; by rejecting it, we conclude that the fuel treatment had a measurable and statistically significant impact on the car's fuel efficiency.

It is also important to consider the **confidence interval** and the context of the **two-tailed test** versus a **one-tailed test**. Stata defaults to a two-tailed probability, which tests for any difference regardless of direction. If a researcher had a clear **directional hypothesis** (e.g., that the treatment would only increase mpg), a one-tailed test might be considered, though this must be justified a priori. In this instance, the result is significant at the 5% level, suggesting the findings are robust and unlikely to be the result of random **sampling error**.

### Step 4: Formulating a Professional Report of the Results

The final stage of the **data analysis** process is the clear and concise reporting of the findings. A professional report should include the name of the test performed, the **sample size**, the test statistic, and the **p-value**. It should also provide a plain-English interpretation of what these numbers mean in the context of the original research question. This ensures that the results are accessible to stakeholders who may not have an advanced background in **statistics**.

Based on our analysis in **Stata**, the report could be structured as follows:

**Methodology:** A **Wilcoxon Signed-Rank Test** was conducted to evaluate whether a specific fuel treatment significantly altered the mean miles per gallon (mpg) of vehicles.

**Sample:** The study utilized a **paired sample** of 12 automobiles, measuring performance both

before and after the application of the treatment.

**Findings:** The analysis revealed a statistically significant difference in fuel efficiency ( $z = -1.973$ ,  $p = 0.0485$ ).

**Conclusion:** Because the **p-value** fell below the 0.05 threshold, the **null hypothesis** was rejected. These results provide evidence that the fuel treatment effectively changed the mpg characteristics of the vehicles.

By following this structured approach, you ensure that your **quantitative analysis** is rigorous and your conclusions are backed by solid evidence. The **Wilcoxon Signed-Rank Test** remains an indispensable tool in **Stata** for any researcher dealing with paired data that does not conform to the requirements of parametric testing. Its ability to provide valid inferences in the presence of non-normal distributions makes it a versatile and powerful component of the **statistical** toolkit.

## Advanced Considerations and Dealing with Ties

While the basic execution of the **signrank** command is simple, advanced users should be aware of how **Stata** handles "ties" in the data. A tie occurs when the "before" and "after" measurements are identical, resulting in a difference of zero. In the **Wilcoxon Signed-Rank Test**, these observations are typically excluded from the ranking process, which can effectively reduce the **sample size** and, consequently, the **statistical power** of the test. If your dataset contains a large number of ties, it may be necessary to investigate why the measurements are not showing variation or to consider alternative **non-parametric** methods.

Another consideration is the **effect size**. While the **p-value** tells you if a difference exists, it does not tell you how large or important that difference is in a practical sense. For the Wilcoxon test, researchers often calculate a **rank-biserial correlation** or use the Z-statistic to derive an r-value effect size. Reporting the effect size alongside the **p-value** provides a more complete picture of the research findings, helping readers understand the **clinical significance** or practical utility of the intervention being studied.

Finally, always ensure that your **data cleaning** process is documented. In **Stata**, using **do-files** allows you to keep a record of every command used, from the initial **use** command to the final **signrank** execution. This promotes **reproducibility** in research, allowing others to verify your results by running the same code on the same dataset. As **data science** continues to evolve, the emphasis on transparent and reproducible **statistical analysis** becomes increasingly paramount.