

How to Easily Perform a Wilcoxon Signed Rank Test in SAS

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Wilcoxon Signed Rank Test in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103452>

The Wilcoxon Signed-Rank Test (WSRT) is a robust nonparametric statistical test designed to determine if there is a statistically significant difference between the median ranks of two related samples. This makes it the non-parametric alternative to the paired samples t-test. The WSRT is indispensable when analyzing paired data--such as before-and-after measurements or matched subjects--where the distribution of the differences between these pairs cannot be assumed to follow a normal distribution, thereby violating a critical assumption of parametric testing. In the context of SAS programming, performing this test involves calculating the difference scores between the paired observations and then utilizing the powerful PROC UNIVARIATE procedure to compute the necessary test statistics and associated p-value. This detailed guide walks through the theory, implementation, and interpretation of the WSRT using SAS.

The Wilcoxon Signed-Rank Test is the non-parametric version of the paired samples t-test. It is used to test whether or not there is a significant difference between two population medians when the distribution of the differences between the two related samples cannot be assumed to be normal.

This test is used to test whether or not there is a significant difference between two population medians when the distribution of the differences between the two samples cannot be assumed to be normal. The following example shows how to perform a Wilcoxon Signed-Rank Test efficiently in SAS using the standard method applied to difference scores.

Understanding the Wilcoxon Signed-Rank Test

The WSRT stands as a foundational method within nonparametric statistics, providing a powerful means to compare two related samples or repeated measurements taken from the same individuals. It serves as the primary alternative when the assumptions required for its parametric counterpart, the paired samples t-test, are violated. Specifically, the WSRT assesses whether the median difference between the paired observations is zero. Unlike the t-test, which focuses on means and requires the differences to be normally distributed, the WSRT relies on the ranks of the differences, making it less sensitive to outliers and distributional shape.

The fundamental mechanism of the WSRT involves three key stages. First, the difference between each pair of observations is calculated. Second, the absolute values of these differences are ranked. Finally, the ranks are assigned the sign of the original difference (positive or negative). The test statistic, often denoted as W , is derived by summing the ranks of the positive differences (or sometimes the negative ones). By comparing this sum to a critical value or calculating the associated p-value, researchers can determine whether the observed differences are sufficiently large to reject the hypothesis of no effect.

This test is particularly valuable in fields like clinical trials, engineering, and psychology, where data

often originates from small samples or exhibits severe skewness. For instance, if researchers track the response of patients before and after a treatment, and the effectiveness measure is highly non-normal, the WSRT provides a statistically sound method for concluding whether the treatment introduced a significant change. Its reliance on ordinal information (ranks) rather than raw scores provides statistical resilience, maintaining adequate power even when distributional assumptions are not met.

When to Use Nonparametric Methods

Choosing the correct statistical procedure hinges on understanding the nature of the data and the assumptions inherent in the test. The WSRT is required when working with paired data that fails the assumption of normality. When performing a parametric test like the paired t-test, we assume that the population of differences follows a normal distribution. If diagnostic checks, such as the Shapiro-Wilk test or visual inspection of a Q-Q plot of the differences, reveal significant departure from normality, using the WSRT prevents inaccurate inferences and unreliable conclusions.

Furthermore, the WSRT is the test of choice for data measured on an ordinal scale, even if the sample size is large. While the Central Limit Theorem might justify the use of parametric tests on non-normal interval data when the sample size is large enough, this principle does not apply if the underlying measurement scale is strictly ordinal. Since the WSRT operates strictly on the relative magnitudes (ranks) of the differences, it aligns perfectly with the properties of ordinal data, preserving the integrity of the analysis and ensuring the validity of the results.

It is important to remember that while nonparametric tests offer robustness, they generally possess less statistical power than their parametric counterparts if the parametric assumptions are perfectly met. Therefore, analysts must execute due diligence: if the normality assumption is met, the paired t-test is usually preferred. However, in ambiguous cases or when working with small, potentially skewed datasets, the default safety net provided by the Wilcoxon Signed-Rank Test in SAS ensures reliable statistical inference, focusing on the consistency and direction of change rather than absolute magnitude.

Example: Setting Up the Engineering Experiment Data

Suppose an engineer wants to know if a new fuel treatment leads to a change in the average miles per gallon (mpg) of a certain car model. To test this hypothesis, the engineer measures the mpg of 12 identical cars first without the fuel treatment and then again with the fuel treatment applied. This repeated measures design generates paired observations, making the Wilcoxon Signed-Rank Test highly appropriate if the difference scores are non-normal.

The results of the mileage test are shown in the table below, structured with paired observations for each car:

With Treatment	Without Treatment
20	24
23	25
21	21
25	22
18	23
17	18
18	17
24	28
20	24
24	27
23	21
19	23

To use these data in SAS for the WSRT, we must first define the paired variables and then calculate the critical difference score. This score represents the change in performance for each vehicle, which will be the input variable for our subsequent statistical procedure.

Step-by-Step SAS Implementation for WSRT

We can use the following code structure to perform the Wilcoxon Signed-Rank Test in SAS. This process involves two distinct data steps and one procedure step. First, we create the initial dataset; second, we calculate the difference variable; and third, we run the PROC UNIVARIATE statement.

/*create dataset: Inputting the raw paired data for 12 cars*/

```
data my_data;  
input car with_fuel without_fuel;  
datalines;  
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
```

```
;
```

```
run;
```

```
/*create new dataset with difference between two fuel treatments: Calculating the paired difference scores*/
```

```
data my_data2;
```

```
set my_data;
```

```
diff=with_fuel-without_fuel; /* A negative value means 'without_fuel' had higher mpg */
```

```
run;
```

```
/*perform Wilcoxon Signed Rank Test: Running PROC UNIVARIATE on the difference variable*/
```

```
proc univariate data=my_data2;
```

```
var diff; /* Specify the variable representing the differences */
```

```
run;
```

The use of PROC UNIVARIATE is highly efficient because, when applied to a single variable, it automatically includes the calculation of the Wilcoxon Signed-Rank Test statistic (S or W) within its output, testing the difference variable against a hypothesized median of zero. If the median of the difference scores is zero, it implies that the treatment had no overall effect, supporting the null hypothesis.

Reviewing the SAS Output Tables

Executing the code generates a detailed output, providing both descriptive statistics and the results of various location tests. It is essential to navigate this output correctly to locate the specific WSRT results.

The UNIVARIATE Procedure
Variable: diff

Moments			
N	12	Sum Weights	12
Mean	-1.75	Sum Observations	-21
Std Deviation	2.70100991	Variance	7.29545455
Skewness	0.5792226	Kurtosis	-1.101612
Uncorrected SS	117	Corrected SS	80.25
Coeff Variation	-154.34342	Std Error Mean	0.7797144

Basic Statistical Measures			
Location		Variability	
Mean	-1.75000	Std Deviation	2.70101
Median	-2.50000	Variance	7.29545
Mode	-4.00000	Range	8.00000
		Interquartile Range	4.50000

Tests for Location: Mu0=0				
Test		Statistic		p Value
Student's t	t	-2.24441	Pr > t	0.0463
Sign	M	-2.5	Pr >= M	0.2266
Signed Rank	S	-22.5	Pr >= S	0.0469

From the descriptive statistics section, we can immediately ascertain the central tendency of our difference scores. We see that the mean difference in mpg between the cars that received the treatment and those that didn't is **-1.75**. This negative value indicates that the average mpg dropped by 1.75 miles when the fuel treatment was used. However, we need the statistical test result to confirm if this observed drop is significant or merely due to random chance.

The definitive result is found in the table titled **Tests for Location**. This table provides the results for three critical tests, allowing for comparison between parametric and nonparametric methods. Focusing on the Wilcoxon row, we extract the following vital statistics:

The Wilcoxon Signed-Rank Test statistic (S): **-22.5**. This reflects the standardized sum of the ranks of the differences.

The corresponding probability value, or p-value (Prob > |S|): **0.0469**. This value determines our statistical decision.

Interpreting the Results and Drawing Conclusions

The statistical interpretation hinges entirely on comparing the calculated p-value against the

predefined significance level, alpha (α). We typically set α at 0.05, representing a 5% risk of incorrectly rejecting a true null hypothesis (Type I error).

Recall that the Wilcoxon Signed-Rank Test uses the following null hypothesis and alternative hypotheses:

H₀: The median difference in mpg is zero. (The treatment has no effect.)

H_A: The median difference in mpg is not zero. (The treatment causes a significant change.)

Since the p-value of the test (0.0469) is less than the chosen significance level of 0.05, we are compelled to reject the null hypothesis (H_0).

This means we have sufficient statistical evidence to conclude that the fuel treatment significantly alters the miles per gallon achieved by the cars. Given that the sample mean difference was negative (-1.75), the specific conclusion is that the new fuel treatment caused a statistically significant reduction in fuel efficiency. The engineer should consequently reject this new formula based on the findings of the Wilcoxon Signed-Rank Test.