Perform a Mann-Whitney U Test in SAS

Authored by

Mohammed looti

November 1, 2025

The Mann-Whitney U Test, often also known as the Wilcoxon Rank-Sum Test, stands as a cornerstone of modern nonparametric statistics. This robust method is indispensable for researchers and analysts tasked with comparing the distributions of two independent samples when the stringent assumptions of parametric methods cannot be satisfied. Specifically, it is the preferred choice when data are known to be non-interval, such as ordinal data, or when the underlying population distributions deviate significantly from the ideal normal distribution.

Unlike the independent samples t-test, which relies on comparing means and assumes homogeneity of variances, the Mann-Whitney U Test operates by ranking all observations across both groups. It then determines whether the sum of the ranks in one sample is significantly different from the sum of the ranks in the other. This powerful approach provides a statistically sound foundation for drawing conclusions, even with relatively small sample sizes (typically when $N < 30$) or when dealing with highly skewed data, thereby offering a reliable nonparametric alternative to its parametric counterpart.

This comprehensive guide is designed to walk you through the precise steps required to execute the Mann-Whitney U test effectively within the SAS statistical software environment. We will cover everything from data preparation and syntax configuration using the appropriate procedures to the critical interpretation of the output, ensuring you can confidently apply this essential test to your own research challenges.

Foundational Principles of the Mann-Whitney U Test

The decision to employ a nonparametric test is often dictated by the characteristics of the dataset. When analyzing real-world phenomena, data often fail to meet the requirements for parametric tests--namely, that the dependent variable must be measured on an interval or ratio scale, and that the residuals must be normally distributed within each group. Ignoring these violations can lead to inflated Type I error rates or reduced statistical power, rendering the results unreliable. The Mann-Whitney U test strategically bypasses these assumptions by shifting the focus from the raw scores (and their means) to the relative ordering of the observations (ranks).

The core philosophy of this test centers on assessing the probability that a randomly selected observation from one population (A) is greater than a randomly selected observation from the second population (B). The test statistic, U, quantifies the degree of overlap between the two samples' distributions. A key assumption underpinning the Mann-Whitney U test is that the distributions of the two populations, if they differ, differ only in location (median), but have the same shape. If this assumption holds, a statistically significant result indicates a difference in the central tendency or median between the two groups, providing a clear answer to the research question regarding group differences.

Mastering the application of this particular test is crucial for data analysts across diverse sectors,

including clinical trials, public health reporting, and quality control engineering. When datasets are constrained by small sample sizes or when the data intrinsically represent rankings (such as consumer preference scores or pain levels), the use of the Mann-Whitney U test ensures that any conclusions drawn regarding population differences are statistically defensible and robust against distributional anomalies.

Case Study: Evaluating Fuel Treatment Efficacy in SAS

To demonstrate the practical utility of the Mann-Whitney U test, let us examine a detailed scenario involving automotive research. A team of engineers has developed a novel fuel treatment and seeks to determine if its application leads to a statistically significant improvement in vehicle fuel efficiency, measured in miles per gallon (mpg). Since the cost of testing is high, the experiment involves a limited scale, which immediately signals the potential need for a nonparametric approach.

The experimental design involved collecting data from a total of 24 independent vehicles. This sample was equally divided: 12 vehicles were assigned to the treated group, receiving the special fuel additive, while the remaining 12 vehicles formed the control group (untreated). The measured mpg values for all 24 cars were recorded. Due to the small sample size (N=12 per group) and the inherent variability often associated with fuel efficiency measurements--which frequently result in skewed data rather than a perfect normal distribution--the researchers correctly identified the Mann-Whitney U Test as the most appropriate statistical tool for comparing the median mpg values between the two groups.

The data collected from the 24 vehicles, representing the two independent samples, are essential for the subsequent SAS analysis. This setup allows us to test the fundamental hypothesis: Does the fuel treatment shift the median fuel efficiency distribution?

Treated	Untreated
24	20
25	23
21	21
22	25
23	18
18	17
17	18
28	24
24	20
27	24
21	23
23	19

Step 1: Structuring and Inputting Data into SAS

The initial and most critical phase of any statistical analysis within the SAS environment involves meticulously preparing and structuring the raw data. For this two-sample comparison, we require two fundamental variables: the numerical dependent variable, `mpg` (miles per gallon), and the categorical independent variable, `group`, which serves to classify each observation as either 'treated' or 'untreated'. Defining these variables accurately ensures that the subsequent statistical procedures can correctly identify and compare the two independent populations.

We utilize a standard SAS DATA step to input the 24 observations directly into a new dataset, which we name `mpg_data`. The use of the dollar sign ($) after the `group` variable in the `input` statement indicates that `group` is a character variable, which is necessary for defining the categorical levels. The following code segment is essential for transforming the raw measurements into a structured, analytical format suitable for nonparametric testing.

Attention to detail during this input phase is paramount, as any errors in assigning observations to the correct treatment level will invariably compromise the integrity and validity of the final statistical conclusions. We must verify the data input accuracy before proceeding to the analytical procedure.

```
/* Preparation: Creating the mpg_data dataset for analysis */
data mpg_data;
input group $ mpg;
datalines;
treated 24
```

treated 25
treated 21
treated 22
treated 23
treated 18
treated 17
treated 28
treated 24
treated 27
treated 21
treated 23
untreated 20
untreated 23
untreated 21
untreated 25
untreated 18
untreated 17
untreated 18
untreated 24
untreated 20
untreated 24
untreated 23
untreated 19
;
run;

Step 2: Executing the Mann-Whitney U Test using PROC NPAR1WAY

Once the data is correctly structured in the `mpg_data` dataset, the execution of the Mann-Whitney U Test is managed through the specialized PROC NPAR1WAY procedure in SAS. This procedure is specifically engineered for performing nonparametric analyses involving a single classification variable, making it the perfect choice for our independent two-sample comparison. The key instruction within this procedure is the `WILCOXON` option. Although the procedure is named for the Wilcoxon method, the resulting output provides the statistics required for both the Wilcoxon Rank-Sum Test and the Mann-Whitney U Test, as they are mathematically equivalent methods for this scenario.

The syntax requires precise definition of the roles of the variables. The `CLASS` statement is mandatory; it designates the categorical variable (`group`) that defines the two independent

samples we wish to compare. The `VAR` statement, conversely, identifies the continuous or measured variable (`mpg`) upon which the comparison will be based. By coupling `PROC NPAR1WAY` with the `WILCOXON` option, we direct SAS to calculate the necessary rank sums and the corresponding test statistics, thereby providing the crucial p-value needed for hypothesis testing.

The following concise code snippet represents the entire analytical procedure required for generating the statistical output, demonstrating the efficiency and power of the PROC NPAR1WAY routine in handling nonparametric comparisons.

```
/* Execution: Performing the Mann-Whitney U test using NPAR1WAY */
proc npar1way data=mpg_data wilcoxon;
class group;
var mpg;
run;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable mpg Classified by Variable group

group	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
treated	12	172.0	150.0	17.203387	14.333333
untreate	12	128.0	150.0	17.203387	10.666667

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic	Z	Pr > Z	Pr > |Z|	t Approximation Pr > Z	t Approximation Pr > |Z|
172.0000	1.2498	0.1057	0.2114	0.1120	0.2240

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
1.6354	1	0.2010

Interpreting the Statistical Output and Drawing Conclusions

The statistical output generated by SAS, specifically the section labeled "Wilcoxon Two-Sample Test," holds the key to answering our research question. The primary value we must extract and

scrutinize is the two-sided asymptotic p-value. In this specific analysis concerning the fuel treatment, the calculated p-value is determined to be **0.2114**. This value represents the probability of observing a difference in the rank sums as extreme as, or more extreme than, the one observed in our sample, assuming that the null hypothesis is true.

To formalize the decision-making process, we must explicitly state the hypotheses being tested:

H0 (The Null Hypothesis): The two population distributions of mpg are identical; specifically, the median mpg for the treated group is the same as the median mpg for the untreated group. The fuel treatment has no measurable effect.
HA (The Alternative Hypothesis): The two population distributions are not identical, meaning the median mpg for the treated group is significantly different from the median mpg for the untreated group. The fuel treatment affects fuel efficiency.

We compare the derived p-value (0.2114) against the predetermined significance level, or alpha ($alpha$), which is conventionally set at 0.05 (representing a 95% confidence level). The decision rule dictates that if the p-value is less than $alpha$, we reject H$_{0}$. Conversely, if the p-value is greater than $alpha$, we fail to reject H$_{0}$. Since 0.2114 is substantially greater than 0.05, we must firmly **fail to reject the null hypothesis**. Statistically, this means that the observed difference in fuel efficiency between the treated and untreated cars is not significant enough to confidently attribute it to the fuel additive itself. The variations observed could plausibly be due to random sampling fluctuation.

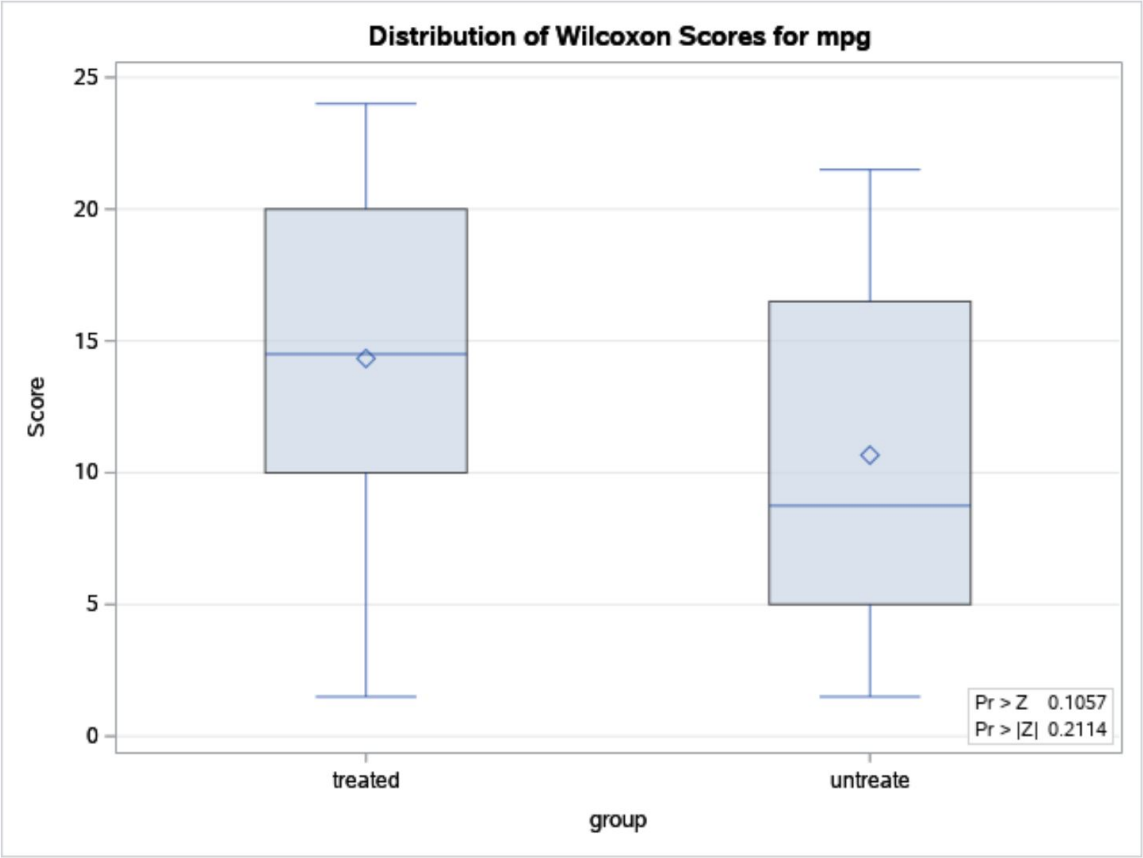
In practical terms for the automotive researchers, the conclusion is clear: based on this sample and the robust Mann-Whitney U test, there is insufficient statistical evidence to assert that the new fuel treatment causes a significant change in the median miles per gallon achieved by the vehicles at the standard 95% confidence threshold. Further research with a larger sample size or a different experimental design would be necessary to try and detect a smaller effect, if one truly exists.

Visual Confirmation Through Boxplots

Statistical software like SAS often complements numerical tests with graphic visualizations, providing an intuitive way to understand the data structure and reinforce the statistical conclusion. The `PROC NPAR1WAY` procedure automatically generates comparative boxplots, which are excellent tools for visualizing the central tendency, spread, and skewness of the measured variable (mpg) within each independent group (treated vs. untreated).

The boxplot visualization clearly displays the distribution of mpg values. While the visual evidence may suggest that the treated group's box is positioned slightly higher than the untreated group's box--indicating a slightly higher median and generally better performance--the crucial observation is the significant overlap between the two boxes. The substantial overlap in the interquartile ranges

(IQR) and the overall range of data points visually confirms the finding of the Mann-Whitney U Test: the difference in location (median) between the two distributions is not large enough, relative to the variability within each group, to be deemed statistically significant.



Therefore, the boxplots serve as a powerful interpretive aid, showing why the resulting p-value was high (0.2114). Had the treatment been highly effective, we would expect to see the two boxes clearly separated with minimal or no overlap, reflecting a strong shift in the distribution of the treated group.

Expanding Your Nonparametric Skillset in SAS

Successfully executing and interpreting the Mann-Whitney U Test in SAS is a fundamental requirement for comprehensive statistical computing. This test represents just one facet of the robust capabilities offered by the PROC NPAR1WAY procedure, which can be adapted to perform several other one-way nonparametric analyses, such as the Kruskal-Wallis test when comparing three or more independent groups.

For analysts seeking to deepen their methodological toolkit, it is highly recommended to explore additional tutorials focused on advanced SAS procedures. Understanding how to handle data that

violate parametric assumptions--whether through transformation, bootstrapping, or the use of nonparametric equivalents--is essential for ensuring that statistical models accurately reflect the underlying data structure and lead to reliable conclusions across all research fields.

These skills solidify the foundation of rigorous data analysis, enabling researchers to tackle complex challenges involving skewed, ordinal, or small-sample datasets with confidence and statistical precision.