

Understanding the Mann-Whitney U Test: A Tutorial with Stata Examples

Authored by
Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding the Mann-Whitney U Test: A Tutorial with Stata Examples*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13731>

Understanding the Mann-Whitney U Test

The [Mann-Whitney U Test](#), often referred to interchangeably as the **Wilcoxon rank-sum test**, serves as a crucial tool in statistical analysis for comparing differences between two independent groups. This test operates by ranking all observations across both samples and then comparing the sum of the ranks for each group. It is specifically employed when the underlying population distributions are not assumed to be [normally distributed](#), or when dealing with small sample sizes (typically when the size of either sample, n , is less than 30).

As a distribution-free method, the Mann-Whitney U test is the [nonparametric equivalent](#) to the widely used [Student's t-test](#). While the t-test relies on assumptions of normality and homogeneity of variance, the Mann-Whitney U test simply assesses whether one sample tends to have larger values than the other, making it exceptionally robust for analyzing data measured on an ordinal scale or data that violate the stringent assumptions necessary for parametric testing. This robustness is vital in many real-world research scenarios, particularly in fields like psychology, clinical trials, and environmental science, where data often defy perfect normal distributions.

This comprehensive tutorial is designed to walk you through the precise methodology required to execute a **Mann-Whitney U test** using the statistical software package, **Stata**. We will cover everything from data loading and initial inspection to the execution of the test and the formal interpretation of the output, ensuring you can confidently apply this essential nonparametric technique to your own research data.

When to Use the Mann-Whitney U Test

Choosing the appropriate statistical test is fundamental to accurate research. The decision to use the Mann-Whitney U test over its parametric counterpart, the independent samples t-test, usually hinges on two primary criteria: the nature of the data and the characteristics of the sample distribution. The core assumption that must be violated for this test to become necessary is the assumption of normality. If diagnostic checks, such as Shapiro-Wilk or Kolmogorov-Smirnov tests, suggest that your variable of interest is significantly skewed or kurtotic within one or both groups, the Mann-Whitney U test provides a reliable alternative, minimizing the risk of Type I errors associated with violating parametric assumptions.

Furthermore, the test is highly suitable when dealing with small sample sizes. When sample sizes are meager (e.g., below 30 per group), it becomes difficult to reliably assess the underlying population distribution, and the central limit theorem cannot be sufficiently relied upon to assume normality of the sampling distribution. In such cases, adopting a nonparametric approach that focuses on ranks rather than means is a safer and more statistically conservative choice. It is important to remember that the [Mann-Whitney U Test](#) does not compare means; rather, it tests whether the probability that a randomly selected observation from one population is greater than a

randomly selected observation from the second population is equal to 0.5.

In practical terms, the Mann-Whitney U test is ideal for research questions involving comparisons between two distinct, independent groups--for instance, comparing test scores between two different teaching methods, or examining income differences between two demographic categories. The test requires that the dependent variable be measured at least on an ordinal scale. When these conditions align--small samples, non-normal distributions, or ordinal data--the Mann-Whitney U test provides a statistically sound method for determining group differences without making restrictive distributional assumptions.

Case Study: Investigating Fuel Efficiency

To illustrate the application of the Mann-Whitney U test, let us consider a relevant research scenario. A team of automotive researchers is interested in determining whether a newly developed **fuel treatment** impacts the average miles per gallon (mpg) achieved by standard passenger vehicles. To conduct this investigation, they designed a controlled experiment involving 24 cars in total. These cars were randomly allocated into two independent groups: the treatment group (12 cars receiving the fuel additive) and the control group (12 cars receiving no additive).

The researchers collected the mpg data for all 24 vehicles under standardized driving conditions. Given the constraints of their experimental design--specifically the **small sample sizes** ($n = 12$ in each group)--and a justifiable suspicion that the mpg distribution might be influenced by external, unmeasured factors resulting in non-normality, they deemed the standard [t-test](#) inappropriate. Consequently, they elected to perform a Mann-Whitney U test. The central research question is: Is there a **statistically significant difference** in the median mpg between the cars that received the fuel treatment and those that did not?

The primary objective of performing this test is to either reject or fail to reject the [null hypothesis](#), which, in this context, posits that there is no difference in the distribution of mpg between the treated and untreated groups. By executing the Mann-Whitney U test in Stata, we aim to calculate a test statistic and an associated P-value to objectively assess the likelihood of observing such a difference if the null hypothesis were true. This methodical approach ensures that conclusions about the efficacy of the fuel treatment are based on robust statistical evidence appropriate for the data characteristics.

Step-by-Step Implementation in Stata: Data Preparation

The first crucial step in any statistical analysis using Stata is loading the relevant dataset. For this example, we will utilize a publicly available dataset specifically designed for demonstrating statistical procedures in Stata.

Step 1: Load the Data. To access the necessary data directly from the Stata web repository, navigate to the command prompt within the Stata interface and accurately input the following command, followed by pressing the Enter key. This action loads the dataset named `fuel2`, containing the `mpg` and treatment status variables, directly into Stata's active memory:

```
use http://www.stata-press.com/data/r13/fuel2
```



```
. use http://www.stata-press.com/data/r13/fuel2
.
```

Command ⏏

```
use http://www.stata-press.com/data/r13/fuel2
```

Step 2: View and Verify the Raw Data Structure. Before proceeding with the analysis, it is always recommended practice to visually inspect the data to ensure it loaded correctly and to understand its structure. This step helps confirm the integrity of the variables and their coding. Access the data editor by selecting **Data > Data Editor > Data Editor (Browse)** from the top menu bar.

Upon opening the data browser, you will observe two primary columns that represent the variables relevant to our test. The first column, labeled `mpg`, represents the continuous outcome variable--the miles per gallon recorded for each individual car. The second column, labeled `treat`, is the

categorical grouping variable, indicating whether or not the car received the experimental fuel treatment. Typically, this variable is binary (e.g., 0 for control, 1 for treatment), signifying the two independent groups we intend to compare using the [Mann-Whitney U Test](#).

	mpg	treat			
1	20	untreated			
2	23	untreated			
3	21	untreated			
4	25	untreated			
5	18	untreated			
6	17	untreated			
7	18	untreated			
8	24	untreated			
9	20	untreated			
10	24	untreated			
11	23	untreated			
12	19	untreated			
13	24	treated			
14	25	treated			
15	21	treated			
16	22	treated			
17	23	treated			
18	18	treated			
19	17	treated			
20	28	treated			
21	24	treated			
22	27	treated			
23	21	treated			
24	23	treated			

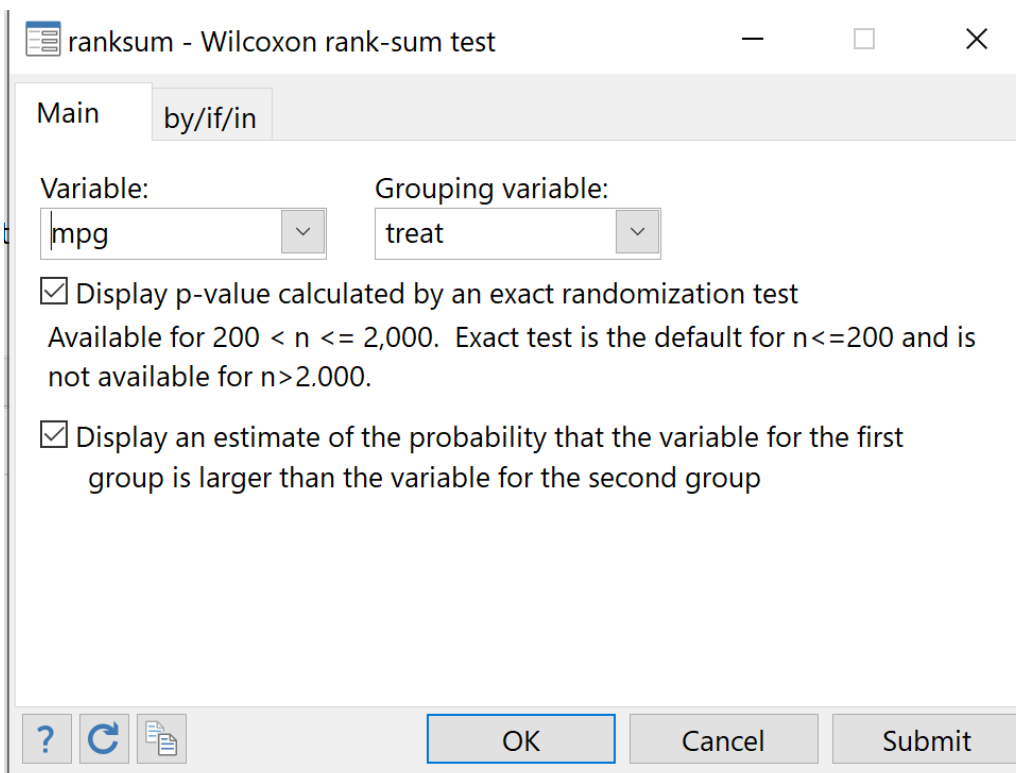
Executing the Wilcoxon Rank-Sum Test in Stata

With the data successfully loaded and verified, the next phase involves executing the statistical test

itself. In Stata, the Mann-Whitney U test is accessed via the command for the Wilcoxon rank-sum test.

Step 3: Perform the Mann-Whitney U Test (Wilcoxon Rank-Sum Test). To initiate the test using Stata's graphical user interface (GUI), follow this path: **Statistics > Summaries, tables, and tests > Nonparametric tests of hypotheses > Wilcoxon rank-sum test**. This sequence opens the dedicated dialog box for the test.

Within the dialog box, careful selection of the variables is critical for obtaining accurate results. For the field designated **Variable**, select *mpg*, as this is our measurement of interest. For the **Grouping variable**, select *treat*, which defines the two independent samples being compared. To ensure a comprehensive output, including specific probability estimates, it is advisable to select both available checkboxes under the Options section: "Display p-value for exact test" (important for small samples) and "Display estimate of the probability that the variable for the first group is larger than the variable for the second group." Once all parameters are correctly specified, click **OK** to run the analysis.



Alternatively, for users comfortable with the command line, the equivalent Stata syntax for this test is highly efficient. The command structure requires specifying the dependent variable followed by the grouping variable, using the `ranksum` command. The command to generate the same results obtained through the GUI is: `ranksum mpg, by(treat) exact`. This command executes the

Mann-Whitney U test, instructing Stata to compare the ranks of the `mpg` variable across the groups defined by the `treat` variable, and explicitly requests the exact P-value calculation.

Interpreting the Output and Drawing Conclusions

After running the command, Stata will generate a detailed output displaying the results of the Wilcoxon rank-sum test. This output is central to determining the statistical outcome of the research hypothesis.

The resulting table provides several key metrics, including the number of observations in each group, the rank sums, and the expected rank sums under the [null hypothesis](#) of no difference. However, the values of primary interest for decision-making are the test statistic (typically represented by **z**) and the associated probability value (**Prob > |z|**).

```
. ranksum mpg, by(treat) porder exact

Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

treat	obs	rank sum	expected
untreated	12	128	150
treated	12	172	150
combined	24	300	300

```

unadjusted variance      300.00
adjustment for ties      -4.04
-----
adjusted variance       295.96

Ho: mpg(treat==untreated) = mpg(treat==treated)
      z =  -1.279
      Prob > |z| =  0.2010
      Exact Prob =  0.2117

P{mpg(treat==untreated) > mpg(treat==treated)} = 0.347

```

In this specific example, the output reveals the following critical values: **z = -1.279** and **Prob > |z| = 0.2010**. The z-statistic reflects the standardized difference based on the rank sums. The P-value (0.2010) represents the probability of observing a difference in mpg distributions as extreme as, or more extreme than, the one observed, assuming the null hypothesis (no effect of the fuel

treatment) is true.

To reach a conclusion, this P-value must be compared against the predetermined **significance level** (alpha, denoted as α), which is conventionally set at 0.05. The decision rule is straightforward: if the P-value is less than α (0.05), we reject the null hypothesis, concluding that a statistically significant difference exists. Conversely, if the P-value is greater than α , we fail to reject the null hypothesis, indicating insufficient evidence to claim a significant difference. Since 0.2010 is considerably larger than 0.05, we **fail to reject the null hypothesis**. Therefore, based on the Mann-Whitney U Test, we conclude that there is not sufficient evidence to assert that the true median mpg is different between the group receiving the fuel treatment and the control group.

Formal Reporting of Nonparametric Results

The final step in the analytical process is to formally report the findings in a clear, concise, and academically appropriate manner. When reporting nonparametric tests like the Mann-Whitney U test, it is essential to include the test statistic, the degrees of freedom (or sample sizes), and the exact P-value. This allows readers to critically evaluate the conclusion based on the statistical evidence.

The following structure provides an excellent template for reporting the results of the fuel efficiency study, adhering to common statistical reporting standards:

A Mann-Whitney U test was conducted to investigate the potential impact of a novel fuel treatment on vehicle miles per gallon (mpg). The total sample comprised 24 cars, with 12 cars allocated to the treatment group and 12 cars allocated to the control group.

The results indicated that the distribution of mpg did not differ significantly between the two independent groups. The calculated test statistic was $z = -1.279$, yielding a P-value of $p = 0.2010$. Since this P-value exceeded the predetermined **significance level** of 0.05, the **null hypothesis** was retained.

In conclusion, based on the evidence provided by the **Mann-Whitney U Test**, the new fuel treatment was found not to have a statistically significant impact on the miles per gallon achieved by the cars in this study.

This reporting style clearly links the statistical result (z-statistic and p-value) back to the original research question, providing a definitive answer regarding the hypothesized difference between the groups. It emphasizes that the lack of statistical significance means the observed difference is likely due to random chance, rather than the systematic effect of the fuel treatment.