

# Mann-Whitney U Test in R: A Step-by-Step Tutorial for Beginners

Authored by  
**Mohammed loot**

November 8, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Mann-Whitney U Test in R: A Step-by-Step Tutorial for Beginners*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13780>

## Understanding the Mann-Whitney U Test

The [Mann-Whitney U test](#), often known interchangeably as the **Wilcoxon rank-sum test**, stands as a cornerstone of statistical analysis when comparing two independent groups. Unlike its parametric counterparts, this powerful tool does not assume that the data follows a specific distribution, making it exceptionally versatile. Its primary purpose is to determine whether the distributions of scores for two independent populations differ significantly. This test becomes indispensable when researchers encounter data that fails to meet the strict assumptions of [parametric tests](#), such as when data is measured on an ordinal scale or when the underlying population distributions are demonstrably not **normally distributed**.

Statistically, the Mann-Whitney U test functions as the gold-standard [nonparametric](#) alternative to the independent samples **Student's t-test**. While the t-test focuses on comparing the arithmetic means, the U test shifts focus to comparing the medians or, more precisely, comparing the ranks of the observations. It calculates the probability that a randomly chosen observation from the first group will be larger than a randomly chosen observation from the second group. This fundamental reliance on the rank ordering of data, rather than the raw numerical values themselves, provides significant robustness against the influence of **outliers** and deviations from ideal distributional characteristics.

To confidently apply this methodology, a few fundamental prerequisites must be satisfied by the experimental design. Firstly, the study must involve two **independent samples** (or groups). Secondly, the dependent variable must be measured either on an ordinal scale or on a continuous scale where the assumption of normality has been violated. The application of this rigorous test is greatly facilitated by the statistical environment of [R](#), which provides dedicated and efficient functions to execute the analysis and generate clear, statistically interpretable results for hypothesis testing.

## Setting Up the Case Study Data

To provide a practical and reproducible demonstration of the [Mann-Whitney U test](#), we will analyze a hypothetical research scenario common in clinical and health sciences. Imagine a team of pharmaceutical researchers interested in assessing the effectiveness of a new medication designed to substantially reduce the incidence of panic attacks in patients suffering from panic disorder. This investigation utilizes a rigorous **randomized controlled trial** (RCT) design, which is the optimal methodological approach for establishing causal links and comparing group differences.

In this study, a total cohort of 12 patients was randomly and equally allocated into one of two distinct treatment arms: the **New Drug Group** or the **Placebo Group**, resulting in 6 participants

per group. Over a standardized period of one month, each participant meticulously recorded the total number of panic attacks they experienced. The analytical objective is straightforward: to statistically contrast the performance metrics of the active pharmaceutical treatment against the inert placebo. For this analysis, we will adhere to the standard convention of setting the statistical significance level, or [alpha](#) ( $\alpha$ ), at 0.05.

The raw data collected from the 12 participants detailing their monthly attack frequency is presented in the table below. Before initiating the test, this raw data must be properly organized within the [R environment](#). R allows for flexibility in data input, either by defining separate **numerical vectors** for each group or, preferably for larger datasets, structuring the entire sample into a unified data frame that includes a grouping variable.

NEW DRUG (Attacks per month)	PLACEBO (Attacks per month)
3	4
5	8
1	6
4	2
3	1
5	9

The ultimate goal is the rigorous statistical determination of whether the number of panic attacks differs significantly between the two groups. In the [R](#) statistical environment, the execution of the Mann-Whitney U test is exclusively managed by the powerful and flexible built-in function, `wilcox.test()`. We will now proceed to demonstrate the two most common methods for implementing this function.

## Executing the Mann-Whitney U Test in R

The `wilcox.test()` function is central to performing the Wilcoxon rank-sum test in [R](#). Its design incorporates robust functionality that accommodates various data input structures, all while guaranteeing identical statistical results. This versatility ensures that analysts can seamlessly integrate the test into diverse analytical pipelines, regardless of whether they are working with simple vectors or complex data frames. We will explore the two primary methodologies for feeding data into the function, starting with the simplest approach suitable for smaller datasets.

### Option 1: Entering Data as Two Separate Vectors

When dealing with smaller samples, or when the data is already segmented into distinct lists

corresponding to the independent groups, the most intuitive approach is to define the data as two separate **numerical vectors**. This method minimizes the setup required and immediately allows the analyst to pass the groups directly as primary arguments to the `wilcox.test()` function. This process is exceptionally clear and easy to audit, making it popular for quick exploratory analyses or teaching examples.

Below is the R code demonstrating how we define the vectors for the "new drug" and "placebo" groups based on our case study data, and subsequently execute the two-sided Mann-Whitney U test. The default setting of the function performs a two-sided test, which assesses the general hypothesis that the two distributions are simply "different" from one another (i.e., the location shift is not zero).

```
# Create a vector for the 'new drug' group
```

```
new <- c(3, 5, 1, 4, 3, 5)
```

```
placebo <- c(4, 8, 6, 2, 1, 9)
```

```
# Perform the two-sided Mann Whitney U test
```

```
wilcox.test(new, placebo)
```

```
# Output interpretation
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: new and placebo
```

```
W = 13, p-value = 0.468
```

```
alternative hypothesis: true location shift is not equal to 0
```

The resulting output provides two critical pieces of information: the calculated test statistic, which is **W = 13**, and the corresponding two-sided [p-value](#) of **0.468**. We use these values in the subsequent stage of hypothesis testing to draw our final conclusions about the drug's efficacy.

## Option 2: Entering Data into One Data Frame using Formula Notation

For analyses involving larger datasets, managing multiple variables, or adhering to best practices in data science, consolidating all information into a single [data frame](#) is highly recommended. This structure requires the dependent variable scores (the number of panic attacks) to be listed in one column, and a separate **factor variable** to define the group membership (e.g., "new" or "placebo"). This setup mirrors the typical structure required for most regression or ANOVA modeling in R.

Once the data frame is properly constructed, the Mann-Whitney U test is executed using R's powerful and highly readable formula notation: `dependent_variable ~ grouping_variable`. This format is intuitive and simplifies the process of specifying which variable contains the

measurements and which variable defines the groups to be compared.

### # Create a data frame with scores and group identifiers

```
drug_data <- data.frame(attacks = c(3, 5, 1, 4, 3, 5, 4, 8, 6, 2, 1, 9),  
drug_group = c(rep("new", 6), rep("placebo", 6)))
```

```
# Perform the Mann Whitney U test using the formula notation
```

```
wilcox.test(attacks~drug_group, data = drug_data)
```

```
# Output
```

```
data: attacks by drug_group
```

```
W = 13, p-value = 0.468
```

```
alternative hypothesis: true location shift is not equal to 0
```

Crucially, both the vector method and the formula method yield the identical test statistic ( $W = 13$ ) and the same [p-value \(0.468\)](#). This consistency confirms the accuracy and reliability of the `wilcox.test()` function regardless of the input organizational strategy. This ability to handle different data formats flexibly is a defining characteristic of robust statistical programming in R.

## Interpretation of Results and Hypothesis Testing

The conclusive step in any statistical analysis is the formal interpretation of the results, which hinges on comparing the calculated [p-value](#) against the predefined significance level, alpha ( $\alpha = 0.05$ ). The underlying premise of the [Mann-Whitney U test](#) is framed by two competing hypotheses. The [null hypothesis](#) ( $H_0$ ) posits that there is no difference in the distribution of panic attack scores between the new drug group and the placebo group. Conversely, the alternative hypothesis ( $H_1$ ) suggests that a genuine, statistically significant difference exists.

In our specific case study, the calculated p-value derived from the `wilcox.test()` function is **0.468**. Applying the decision rule of hypothesis testing, we compare this value to our threshold: since 0.468 is substantially greater than the stipulated significance level of 0.05, the standard procedure dictates that we must **fail to reject the null hypothesis**. This outcome is statistically profound, meaning that the data does not provide sufficient evidence to conclude that the new medication had a meaningful effect.

Translating this statistical finding back into practical, clinical terms means that the researchers cannot confidently assert that the new drug is superior to the placebo in mitigating the frequency of panic attacks. Although the raw data might show a slight numerical difference in the median attack rate between the two groups, the statistical test confirms that this observed variation is highly likely attributable to standard **random sampling fluctuation** inherent in the small sample size, rather than a true treatment effect caused by the new drug.

## Advanced Usage Notes on the `wilcox.test()` Function

By default, the standard implementation of the [Mann-Whitney U test](#) in R performs a **two-sided test**, which is non-directional and merely seeks to establish if the two distributions are different in any direction. However, in rigorous scientific research, hypotheses are often directional. For example, researchers might specifically hypothesize that the new drug will result in *fewer* panic attacks compared to the placebo, requiring a more focused statistical approach.

To accommodate such directional predictions, the `wilcox.test()` function includes the essential `alternative` argument, which enables the specification of a **one-sided (or one-tailed) test**. If the hypothesis dictates that the first group (`new`) is expected to be statistically smaller than the second group (`placebo`), the argument is set to `alternative="less"`. This crucial adjustment channels the statistical power to detect a difference exclusively in the hypothesized direction, offering increased sensitivity for that specific outcome.

### # Create a vector for each group

```
new <- c(3, 5, 1, 4, 3, 5)
```

```
placebo <- c(4, 8, 6, 2, 1, 9)
```

```
# Perform the Mann Whitney U test, specifying alternative="less"
```

```
wilcox.test(new, placebo, alternative="less")
```

### # Output

```
Wilcoxon rank sum test with continuity correction
```

```
data: new and placebo
```

```
W = 13, p-value = 0.234
```

```
alternative hypothesis: true location shift is less than 0
```

Reviewing the one-sided output, we observe that the test statistic **W remains 13**, which is expected since the ranking calculation is independent of the hypothesis direction. However, the [p-value](#) has decreased significantly to **0.234** (exactly half of the two-sided p-value:  $0.468 / 2$ ). Despite this increase in statistical sensitivity, 0.234 is still greater than our 0.05 threshold. Consequently, even with a directional hypothesis, we still fail to reject the [null hypothesis](#), confirming that the evidence does not support the conclusion that the new drug significantly reduced panic attacks compared to the placebo.

## Further Statistical Resources

For analysts dedicated to mastering nonparametric statistical methods within the [R](#) environment, expanding knowledge beyond the basic execution of the Mann-Whitney U test is essential.

Competence in the following related areas ensures a comprehensive ability to handle data sets that defy traditional parametric assumptions.

**Effect Size Calculation:** While hypothesis tests determine statistical significance, they do not quantify the practical importance or magnitude of the difference. It is crucial to learn how to calculate and interpret the appropriate [effect size](#) for the Mann-Whitney U test, typically the **rank-biserial correlation**, which provides vital contextual information alongside the p-value.

**Assumption Testing:** A fundamental skill is the ability to formally test data assumptions. Familiarize yourself with formal tests for normality (e.g., the **Shapiro-Wilk test**) and homogeneity of variances (e.g., **Levene's test**). These tests provide the necessary evidence to confidently choose between parametric and [nonparametric](#) approaches.

**Related Nonparametric Tests:** Broaden your repertoire by exploring other common nonparametric procedures. Key tests include the **Wilcoxon Signed-Rank Test** (the nonparametric equivalent for paired or dependent samples) and the **Kruskal-Wallis H Test** (used when comparing three or more independent groups, serving as the nonparametric alternative to one-way ANOVA).