A Comprehensive Guide to Performing the Mann-Whitney U Test in Excel

Authored by

Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *A Comprehensive Guide to Performing the Mann-Whitney U Test in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from https://statistics.arabpsychology.com/?p=13567

The Core Principles of the Mann-Whitney U Test

The Mann-Whitney U test, frequently recognized by its alternative name, the **Wilcoxon rank-sum test**, represents a crucial tool in the field of statistical inference. Its primary function is to rigorously compare whether two independent populations are likely to have originated from the same distribution, effectively determining if there is a statistical difference between their medians. This powerful method is universally categorized as a nonparametric procedure because, unlike its parametric counterparts, it operates without making restrictive assumptions about the specific shape or parameters of the underlying population distribution.

A key advantage of utilizing this procedure is its applicability when dealing with datasets that violate the strict assumptions necessary for parametric tests. Specifically, the **Mann-Whitney U test** is the preferred alternative when the data cannot be assumed to follow a normal distribution, or when the sample sizes are notably small (a common guideline suggests $n lt 30$). Its reliance on the distribution of **ranks**, rather than the raw data values themselves, makes it exceptionally resilient to outliers and variances in data scale, providing reliable results across diverse fields, including medical research and behavioral sciences.

Functionally, this test serves as the essential nonparametric equivalent to the independent samples T-test. While the standard **T-test** necessitates that the user confirms assumptions of population normality and homogeneity of variance, the **Mann-Whitney U test** sidesteps these requirements entirely. By transforming the raw scores into ordered **ranks**, it focuses on comparing the median positions of the two groups, offering a robust and statistically sound method for comparing two independent groups when foundational assumptions are difficult or impossible to verify.

This extensive guide provides a detailed walkthrough of how to meticulously perform a **Mann-Whitney U test** using **Microsoft Excel**. While professional statistical software packages are generally recommended for highly complex analyses, **Excel** offers readily available and accessible functionality for executing this specific rank-based calculation. By following these precise, sequential steps, users can ensure high accuracy in deriving the necessary test statistics.

Establishing the Case Study and Rationale

To illustrate the practical application of the **Mann-Whitney U test**, consider a standard research scenario investigating the efficacy of an experimental fuel treatment designed to boost automotive efficiency. The core statistical question centers on whether this new treatment results in a statistically meaningful alteration in the miles per gallon (**mpg**) performance of vehicles. To address this, a controlled experiment was implemented, involving two distinct and independent groups of automobiles.

The researchers meticulously collected **mpg** data from 12 cars that received the new fuel treatment (the treated group) and 12 cars that did not (the control group). Due to the modest size of the samples ($n=12$ for each group) and the inherent uncertainty regarding whether fuel efficiency measurements perfectly conform to a normal distribution, the researchers prudently opted for the Mann-Whitney U test. This strategic choice guarantees that the analysis remains statistically valid even if the stringent distributional requirements of parametric tests, such as the **T-test**, are violated by the data structure.

The primary analytical objective is to ascertain if a statistically significant difference in **mpg** performance exists between the two groups, basing this determination on the comparison of their median **ranks** rather than their arithmetic means. The subsequent methodology outlines the precise calculations required within **Excel** to determine the crucial components of the test, ultimately leading to the derivation of the standardized **z test statistic** and the critical associated P-value.

Step 1: Data Structuring and Preparation in Excel

The foundation of any robust statistical investigation is the accurate entry and systematic organization of the raw data. When preparing to conduct the **Mann-Whitney U test**, it is imperative that the data points for the two independent groups are entered into distinct, separate columns within the **Excel** spreadsheet. This fundamental organizational structure is not merely for clarity; it is essential for facilitating the subsequent, complex calculations required for accurate ranking.

Following our fuel treatment example, the measured **mpg** data for the 'Treated' group should occupy one column (e.g., Column A), and the 'Control' group data should occupy the adjacent column (e.g., Column B). Typically, the data entry begins in cells A2 and B2. Maintaining this side-by-side, consistent structure from the very beginning is vital, as it significantly reduces the likelihood of errors during the subsequent ranking process, which demands precise referencing of data ranges.

This preparatory stage ensures that all data points are correctly aligned and identifiable before advancing to the core ranking procedure, which is the most complex computational step of the entire analysis. The visual representation below demonstrates the required two-column data layout necessary to proceed with the calculation:

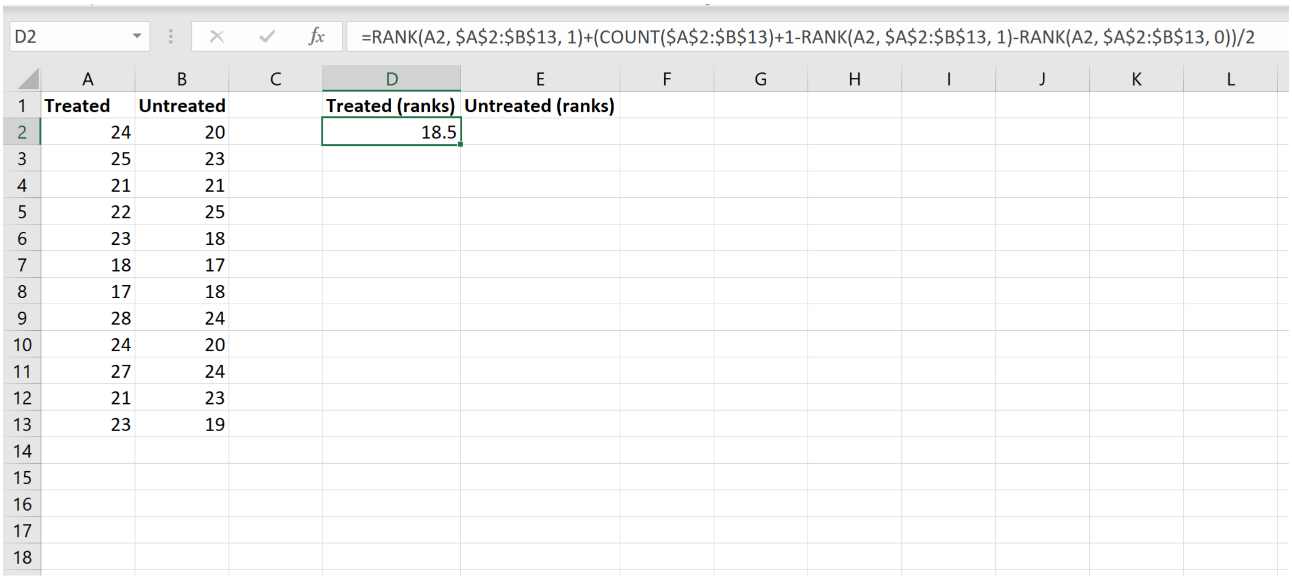
	A	B	C	D	E
1	**Treated**	**Untreated**			
2	24	20			
3	25	23			
4	21	21			
5	22	25			
6	23	18			
7	18	17			
8	17	18			
9	28	24			
10	24	20			
11	27	24			
12	21	23			
13	23	19			
14					
15					
16					
17					
18					

Step 2: The Critical Process of Combined Ranking

The statistical power of the Mann-Whitney U test derives from its ranking procedure. This step involves pooling the data from both the Treated and Control groups into a single combined dataset. In our scenario, with 12 observations per group, the total sample size is $N=24$. We must then assign a numerical **rank** to every observation across this combined pool, starting with a rank of 1 for the smallest value, 2 for the second smallest, and continuing up to the largest observation. It is absolutely crucial that any occurrences of tied scores are handled correctly by assigning the average of the ranks they would have otherwise occupied.

Automating this meticulous ranking process within **Excel** requires constructing a sophisticated array formula. This formula must simultaneously identify the relative position of each score within the entire combined sample while also ensuring that proper adjustments are made for ties. The most effective methodology typically involves nested use of the `COUNTIF` and `RANK.AVG` functions within a structured lookup environment to guarantee precise assignment of the combined **ranks** to the original data points.

For example, to calculate the rank for the first value in the Treated group (Cell C2), the formula must reference the full range of both the Treated data (A2:A13) and the Control data (B2:B13). Importantly, all data ranges must be defined as absolute references (e.g., $A$2:$A$13) to prevent them from shifting incorrectly when the formula is copied. The initial complex formula setup, designed to accurately determine these combined **ranks**, is demonstrated in the image below:


D2 =RANK(A2, $A$2:$B$13, 1)+(COUNT($A$2:$B$13)+1-RANK(A2, $A$2:$B$13, 1)-RANK(A2, $A$2:$B$13, 0))/2

	A	B	C	D	E
1	Treated	Untreated		Treated (ranks)	Untreated (ranks)
2	24	20		18.5	
3	25	23			
4	21	21			
5	22	25			
6	23	18			
7	18	17			
8	17	18			
9	28	24			
10	24	20			
11	27	24			
12	21	23			
13	23	19			

Despite the initial complexity of this formula, the efficiency of **Excel** means it only needs to be entered correctly into the first calculation cell (C2). Once verified, the user can simply drag the formula down the entire 'Rank Treated' column (C2:C13) and then copy it across and down for the 'Rank Control' column (D2:D13). This action automatically and accurately calculates the combined **ranks** for every single data point in both groups, resulting in the structured data required for the next step: the derivation of the test statistic components:

	A	B	C	D	E	F
1	Treated	Untreated		Treated (ranks)	Untreated (ranks)	
2	24	20		18.5	7.5	
3	25	23		21.5	14.5	
4	21	21		10	10	
5	22	25		12	21.5	
6	23	18		14.5	4	
7	18	17		4	1.5	
8	17	18		1.5	4	
9	28	24		24	18.5	
10	24	20		18.5	7.5	
11	27	24		23	18.5	
12	21	23		10	14.5	
13	23	19		14.5	6	
14						
15						
16						
17						
18						

Step 3: Calculation of the U Test Statistic Components

Following the successful assignment of combined **ranks**, the next critical phase involves calculating the constituent elements needed to derive the **U test statistic**. This stage requires four distinct measurements: calculating the total sum of **ranks** for each group (R_1 and R_2), and determining the exact sample size for each group (n_1 and n_2). These values collectively form the basis for determining the relative difference in medians.

The sum of the **ranks** (R_1 and R_2) is straightforwardly calculated by applying the `SUM()` function to the respective rank columns (C and D). Similarly, the sample sizes (n_1 and n_2) are quickly obtained using the `COUNT()` function on the original data columns (A and B). These foundational measures are then incorporated into the formulas for the intermediate U statistics.

The **U test statistic** must be calculated separately for each group (U_1 and U_2) using the following specific formulas, which mathematically relate the group's observed rank sum to the rank sum that would be expected under the null hypothesis:

$U_1 = n_1 n_2 + frac{n_1(n_1+1)}{2} - R_1$
$U_2 = n_1 n_2 + frac{n_2(n_2+1)}{2} - R_2$

For the purposes of hypothesis testing, the overall **U test statistic** used in the subsequent steps is defined as the smaller of the two calculated values: $U = min(U_1, U_2)$. The visualization below displays the layout and resulting intermediate values within **Excel** after these calculations have been successfully executed:

	A	B	C	D	E	F	G	H
1	**Treated**	**Untreated**		**Treated (ranks)**	**Untreated (ranks)**			
2	24	20		18.5	7.5			
3	25	23		21.5	14.5			
4	21	21		10	10			
5	22	25		12	21.5			
6	23	18		14.5	4			
7	18	17		4	1.5			
8	17	18		1.5	4			
9	28	24		24	18.5			
10	24	20		18.5	7.5			
11	27	24		23	18.5			
12	21	23		10	14.5			
13	23	19		14.5	6			
14								
15				**R1**	172	=SUM(D2:D13)		
16				**R2**	128	=SUM(E2:E13)		
17								
18				**N1**	12	=COUNT(A2:A13)		
19				**N2**	12	=COUNT(B2:B13)		
20								
21				**U1**	50	=E18*E19+E18*(E18+1)/2-E15		
22				**U2**	94	=E18*E19+E19*(E19+1)/2-E16		
23								
24				**U**	50	=MIN(E21, E22)		
25								
26								
27								

Step 4: Normal Approximation and P-Value Calculation

A critical step in deriving a conclusion is the approximation of the U distribution. When the sample sizes are adequately large (the established rule of thumb suggests $n > 8$ for both groups, which is true in our example), the distribution of the **U test statistic** closely mirrors the standard normal distribution. This relationship permits us to transform the calculated U value into a standardized **z test statistic**. This standardization is essential because it allows us to precisely determine the statistical probability of observing our data if the null hypothesis were true--a probability represented by the P-value.

This conversion process requires two preparatory calculations: determining the theoretical mean (μ_U) and the standard deviation (σ_U) of the U distribution, assuming the null hypothesis holds true (i.e., that the groups are statistically similar):

Mean: $\mu_U = \frac{n_1 n_2}{2}$
Standard Deviation: $\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

Once these parameters are calculated within **Excel**, the standardized **z test statistic** is derived using the standard formula: $Z = frac{U - mu_U}{sigma_U}$. The final, critical step is obtaining the corresponding two-tailed P-value. This is achieved using **Excel**'s `NORMSDIST` function, which compares the calculated Z value against the standard normal curve to quantify the probability of the observed outcome.

The image provided below illustrates the final block of calculations, clearly showing the derived **z test statistic** and the resulting critical P-value, the value that will ultimately dictate the statistical decision regarding the fuel treatment's effect:

	A	B	C	D	E	F	G	H	I	J
1	Treated	Untreated		Treated (ranks)	Untreated (ranks)					
2	24	20		18.5	7.5					
3	25	23		21.5	14.5					
4	21	21		10	10					
5	22	25		12	21.5					
6	23	18		14.5	4					
7	18	17		4	1.5					
8	17	18		1.5	4					
9	28	24		24	18.5					
10	24	20		18.5	7.5					
11	27	24		23	18.5					
12	21	23		10	14.5					
13	23	19		14.5	6					
14										
15				R1	172	=SUM(D2:D13)				
16				R2	128	=SUM(E2:E13)				
17										
18				N1	12	=COUNT(A2:A13)				
19				N2	12	=COUNT(B2:B13)				
20										
21				U1	50	=E18*E19+E18*(E18+1)/2-E15				
22				U2	94	=E18*E19+E19*(E19+1)/2-E16				
23										
24				U	50	=MIN(E21, E22)				
25										
26				z	-1.270170592	=(E24-E18*E19/2)/SQRT(E18*E19*(E18+E19+1)/12)				
27				p	0.20402387	=NORM.DIST(E26, 0, 1, TRUE)*2				
28										

Interpreting Results and Drawing the Conclusion

The fundamental goal of performing the Mann-Whitney U test is to rigorously evaluate the central tenet of the null hypothesis (H_0). The null hypothesis asserts that the two populations--the treated group and the control group--share identical distributions, meaning that the true median **mpg** values are statistically equivalent.

For this specific fuel treatment analysis, the calculated P-value resulting from our **Excel** calculations is **0.20402387**. To render a conclusive statistical decision, this P-value must be compared against the predetermined level of statistical significance level ($alpha$). By convention, $alpha$ is most often set at 0.05. The standard decision rule mandates that if the P-value is lower than $alpha$, we must reject H_0.

In this case, the calculated P-value of 0.20402387 is substantially larger than the selected significance level of 0.05. Consequently, we are compelled to fail to reject the null hypothesis. This statistical outcome demonstrates that, based on the collected sample data and the rigorous nature of the **Mann-Whitney U test**, the researchers lack sufficient statistical evidence to definitively conclude that the true median **mpg** performance differs between the automobiles that received the experimental fuel treatment and those that did not. The evidence suggests the observed difference is likely due to random chance, not the treatment itself.