Learning the Mann-Whitney U Test: A Guide to Non-Parametric Hypothesis Testing

Authored by

Mohammed looti

November 9, 2025

The **Mann-Whitney U test**, also known as the **Wilcoxon rank-sum test**, is a foundational procedure within nonparametric statistics. This powerful tool is specifically designed to determine whether there is a statistically significant difference between the distributions of two independent samples. It is invaluable in research settings where the data cannot confidently be assumed to follow a normal distribution, or when researchers are constrained by small sample sizes (typically defined as $n < 30$).

Unlike its parametric counterpart, the Mann-Whitney U test relies on ranking the data rather than using the raw data values directly. This makes it the formally recognized **nonparametric equivalent** to the independent samples t-test, offering a robust and reliable alternative when the stringent assumptions of parametric testing are violated by the collected data.

When to Apply the Mann-Whitney U Test

The Mann-Whitney U test is the most appropriate statistical choice whenever an investigator seeks to compare measurements derived from two distinct and independent groups. This test is particularly necessary when the variable of interest is measured on an ordinal scale or when preliminary analysis indicates that the data distribution is far from normal, rendering the t-test invalid.

Consider situations where data points are likely skewed or contain extreme outliers that would disproportionately influence a mean-based test. By converting raw scores into ranks, the Mann-Whitney U test mitigates the impact of these issues, providing a more balanced assessment of central tendency or distributional differences between the two groups.

You need to compare the entry salaries of five graduates from University A versus five graduates from University B, knowing that the underlying salary distribution in that industry is **not normally distributed**.
You are evaluating differences in weight loss outcomes between two small groups (12 people on Diet A and 10 people on Diet B), and the weight loss data collected is demonstrably **not normally distributed**.
You wish to determine if the test scores of 8 students in Class A differ from those of 7 students in Class B, assuming the underlying distribution of academic scores is **non-normal**.

In all these examples, we are comparing two independent sets of measurements, dealing with relatively small sample sizes, and/or the data fails to meet the assumption of normal distribution.

Consequently, the Mann-Whitney U test is ideally suited for analysis, provided its core underlying assumptions are satisfied.

Core Assumptions of the Mann-Whitney U Test

Before embarking on the calculation of the U statistic, it is essential to confirm that the collected data meets the following three fundamental assumptions. Adherence to these criteria ensures the statistical validity and proper interpretation of the Mann-Whitney U test results.

Measurement Scale: The dependent variable--the outcome being measured--must be captured on an ordinal scale or a continuous scale. **Ordinal variables** involve data that can be meaningfully ranked (e.g., satisfaction ratings, Likert scales). **Continuous variables** include precise numerical measurements like time, age, or standardized test scores.
Independence of Observations: All individual observations collected from both samples must be **mutually independent**. This critical assumption requires that the two comparison groups are distinct from each other, and that the value obtained from any single participant does not influence, nor is influenced by, the value obtained from another participant.
Distributional Shape: Although the data does not need to be normally distributed, the **shape** of the distributions for the two independent groups should be approximately the same. If the shapes differ dramatically (e.g., one is heavily skewed right and the other is heavily skewed left), the test should be interpreted cautiously, as it may only reliably compare the medians rather than the overall distributions.

Once these necessary preliminary assumptions have been thoroughly verified, researchers can confidently proceed with the mathematical steps required to calculate the Mann-Whitney U statistic and determine the outcome of the hypothesis test.

Step-by-Step Guide to the Mann-Whitney U Test Procedure

Conducting the Mann-Whitney U test follows the rigorous, standardized five-step framework commonly employed in statistical hypothesis testing. This systematic structure guarantees a logical, objective, and verifiable conclusion regarding the research question.

Formulate the Hypotheses.

The Mann-Whitney U test is typically employed as a two-sided test, seeking to establish any general difference between the population distributions of the two groups being compared.

H0: The two population distributions are equal (The Null Hypothesis).

Ha: The two population distributions are not equal (Alternative Hypothesis).

Set the Significance Level (α).

The researcher must explicitly predetermine the significance level (alpha, α). This value establishes the probability threshold below which the researcher is willing to reject the null hypothesis, often set at 0.05 (5%) or 0.01 (1%).

Calculate the Test Statistic (U).

The test statistic, denoted as U, is derived by calculating two separate values, U1 and U2, and selecting the minimum value as the final U statistic. These calculations rely on first ranking all observations across the combined data set from both groups:

$U1 = n1n2 + n1(n1+1)/2 - R1$

$U2 = n1n2 + n2(n2+1)/2 - R2$

In these formulas, n1 and n2 represent the respective sample sizes of the two groups. Crucially, R1 and R2 represent the **sum of the ranks** that correspond specifically to the observations belonging to sample 1 and sample 2.

The subsequent practical examples will clearly illustrate the process of combining data, handling tied ranks, and subsequently calculating R1 and R2 to determine the U statistic.

Compare U to the Critical Value.

The calculated U statistic must then be compared against the appropriate tabulated critical value. This critical value is determined by consulting a Mann-Whitney U table, which relies on the specific sample sizes (n1 and n2) and the predetermined significance level (α). The decision rule is straightforward: if the calculated U value is less than or equal to the critical value, the null hypothesis must be rejected.

Interpret the Findings.

The final and most important step involves translating the statistical decision (whether to reject or fail to reject H0) into meaningful, real-world conclusions that directly address the initial research question and context.

Illustrative Calculations of the Mann-Whitney U Test

The following two detailed examples provide a hands-on demonstration of the manual computational steps required to successfully perform the **Mann-Whitney U test**, from combining the data to reaching a final interpretation.

Example 1: Assessing a New Drug for Panic Attacks

A clinical trial was organized to investigate the effectiveness of a new pharmaceutical drug in reducing the frequency of panic attacks. Twelve patients were randomly assigned to one of two equal groups (n=6 each): one group received the **New Drug**, and the other received a **Placebo**. The table below records the total number of panic attacks experienced by each patient over a dedicated one-month period.

NEW DRUG	PLACEBO
3	4
5	8
1	6
4	2
3	1
5	9

We must conduct a Mann-Whitney U test to ascertain if the frequency of panic attacks significantly differs between the two treatment groups. We will utilize a standard .05 level of significance for this test.

1. State the hypotheses.

H0: The distribution of panic attacks for the New Drug population is equal to that of the Placebo population.

Ha: The distribution of panic attacks for the two populations are not equal.

2. Determine a significance level to use for the hypothesis.

As stipulated by the research design, the chosen **significance level (α) is 0.05**.

3. Find the test statistic.

To calculate U, we must first determine the sums of the ranks (R1 and R2). This involves pooling all 12 observations and ranking them from the smallest (rank 1) to the largest (rank 12). Note that the sample sizes are equal: n1 = 6 and n2 = 6.

The formulas for calculating the U values are:

U1 = n1n2 + n1(n1+1)/2 - R1

U2 = n1n2 + n2(n2+1)/2 - R2

The initial data grouped by treatment:

NEW DRUG	PLACEBO
3	4
5	8
1	6
4	2
3	1
5	9

Total Sample (Sorted): 1, 1, 2, 3, 3, 4, 4, 5, 5, 6, 8, 9

Ranks Assigned (Handling Ties): 1.5, 1.5, 3, 4.5, 4.5, 6.5, 6.5, 8.5, 8.5, 10, 11, 12

R1 (Sum of ranks for New Drug) = 1.5 + 4.5 + 6.5 + 8.5 + 4.5 + 8.5 = **34**

R2 (Sum of ranks for Placebo) = 1.5 + 3 + 6.5 + 10 + 11 + 12 = **44**

Next, we substitute n1=6, n2=6, R1=34, and R2=44 into the U formulas:

U1 = 6(6) + 6(6+1)/2 - 34 = 36 + 21 - 34 = **23**

U2 = 6(6) + 6(6+1)/2 - 44 = 36 + 21 - 44 = **13**

The test statistic (U) is the smaller of the two values: **U = 13**.

Note: The ranking process, especially handling ties, is a crucial element of the nonparametric approach.

4. Reject or fail to reject the null hypothesis.

Consulting the Mann-Whitney U table for n1 = 6, n2 = 6, and a two-tailed α = 0.05, we find that the critical value is 5:

n1 \ n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		1	2	3	5	6	7	10	11	13	14	16	17	19	21	22	24	25	27
7		1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	7	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Since our calculated test statistic (U = 13) is **greater than** the critical value (5), we **fail to reject the null hypothesis**.

5. Interpret the results.

Based on the evidence gathered, we conclude that there is insufficient statistical evidence at the 0.05 level of significance to claim that the distribution of panic attacks differs significantly between the placebo group and the group receiving the new drug.

Example 2: Study Time and Academic Performance

Researchers designed an experiment to assess the impact of a mandatory 30 minutes of daily study time over one week on subsequent academic test scores. Fifteen students were randomly allocated to either the **Study Group** (n1=8) or the **No-Study Group** (n2=7). The resulting test scores for all participants are provided in the table below.

STUDY	NO-STUDY
89	88
92	93
94	95
96	75
91	72
99	80
84	81
90	

We will perform the Mann-Whitney U test to determine if there is a statistical difference in the resultant test scores between the two groups, employing a more conservative .01 level of significance.

1. State the hypotheses.

H0: The two population distributions of test scores are equal (no effect of studying).

Ha: The two population distributions of test scores are not equal (studying has an effect).

2. Determine a significance level to use for the hypothesis.

For this analysis, the problem dictates the use of a conservative **significance level (α) of 0.01**.

3. Find the test statistic.

We must calculate R1 and R2 by combining and ranking all 15 scores. Our sample sizes are n1=8 (Study Group) and n2=7 (No-Study Group).

The formulas remain:

U1 = n1n2 + n1(n1+1)/2 - R1

U2 = n1n2 + n2(n2+1)/2 - R2

The data samples used for ranking:

STUDY	NO-STUDY
89	**88**

STUDY	NO-STUDY
92	**93**
94	**95**
96	**75**
91	**72**
99	**80**
84	**81**
90	

Total Sample (Sorted): 72, 75, 80, 81, 84, 88, 89, 90, 91, 92, 93, 94, 95, 96, 99

Ranks Assigned: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

R1 (Sum of ranks for Study Group) = 5+7+8+9+10+12+14+15 = **80**

R2 (Sum of ranks for No-Study Group) = 1+2+3+4+6+11+13 = **40**

Now, we calculate U1 and U2 using n1=8, n2=7, R1=80, and R2=40.

U1 = 8(7) + 8(8+1)/2 - 80 = 56 + 36 - 80 = **12**

U2 = 8(7) + 7(7+1)/2 - 40 = 56 + 28 - 40 = **44**

The test statistic is the minimum value: **U = 12**.

*Note: The formula U1 + U2 must always equal n1n2 (12 + 44 = 56, and 8*7 = 56).*

4. Reject or fail to reject the null hypothesis.

Using n1 = 8, n2 = 7, and a significance level of 0.01 (two-tailed), the Mann-Whitney U table indicates that the critical value is 6:

n1 \ n2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3								0	0	0	1	1	1	2	2	2	2	3	3
4					0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5				0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6			0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7			0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8			1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9		0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10		0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11		0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	46
12		1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13		1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14		1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15		2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16		2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17		2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18		2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	0	3	8	13	18	24	30	36	42	46	54	60	67	73	79	86	92	99	105

Since our calculated test statistic (U = 12) is **greater than** the critical value (6), we **fail to reject the null hypothesis**.

5. Interpret the results.

Based on the Mann-Whitney U test results, we do not have sufficient statistical evidence at the 0.01 level of significance to conclude that the test scores of the students who were instructed to study daily are statistically different from the test scores of the students who were in the control group.

Additional Resources

For statistical practitioners or data scientists who wish to implement the
Mann-Whitney U test
using common programming environments, such as Python or R, accessing pre-built functions is the most efficient approach.

How to Perform a Mann-Whitney U Test in Python