Mann-Whitney U Test

The Mann-Whitney U test is a non-parametric statistical test used to determine whether there is a significant difference between the distributions of two independent groups. It is an alternative to the independent samples t-test when the assumptions of normality and homogeneity of variance are not met. The test compares the ranks of data points rather than the actual data values, making it suitable for ordinal data or continuous data that does not follow a normal distribution.

In the Mann-Whitney U test, the data from both groups are combined and ranked from lowest to highest. The ranks of the observations are then summed for each group, and the U statistic is calculated based on the difference in rank sums between the groups. A smaller U value indicates a greater difference between the groups.

In an independent t-test, the means of 2 samples are tested for statistical difference. In the Mann-Whitney U test (a.k.a. Wilcoxon rank-sum test in R), the variance (spread) of the two samples are tested rather than the means. The hypotheses for the Mann-Whitney U test are setup as follows:

Null Hypothesis: There is no difference between the groups ($H_1 = H_2$).

Alternative Hypothesis: This is a difference between the two groups ($H_1 \neq H_2$).

When using statistical software, the test provides a p-value to assess statistical significance. If the p-value is less than a chosen significance level (e.g., 0.05), the null hypothesis that the two groups have the same distribution is rejected, suggesting a significant difference between the groups. The Mann-Whitney U test is widely used when sample sizes are small or when data does not meet the assumptions of parametric tests.

Ranking Data for Nonparametric Rank Tests

Non-parametric rank tests are statistical tests that do not assume a specific distribution for the data. Instead, they rely on the ranks of the data points, rather than their actual values. This makes them robust to outliers and non-normality. To rank data for these tests, you follow these general steps:

1. **Combine Data**: If you're comparing multiple groups, combine all the data points into a single list.
2. **Assign Ranks:** Assign a rank to each data point based on its relative position in the combined list. The smallest value gets rank 1, the second smallest gets rank 2, and so on.
3. **Handle Ties**: If there are ties, assign the average rank to each tied data point.
    a. For example, if two data points are tied for the third and fourth positions, assign them a rank of 3.5.
    b. When ranking 3 items that are tied, you assign them the average of their ranks.
    c. For example, if you have three items that would normally rank 4th, 5th, and 6th, you would assign them all a rank of 5. This is because the average of 4, 5, and 6 is (4 + 5 + 6) / 3 = 5.
    d. This method ensures that the total sum of ranks remains the same, even when there are ties.

Once you've got the data ranked, you can apply the Mann-Whitney U Test to determine if there is a statistically significant difference between the groups.

Calculating the U Statistic

The U statistic is used in the Mann-Whitney U test, a non-parametric test that compares two independent groups. It measures the degree of overlap between the two groups' rankings.

Steps to calculate the U statistic:

1. **Rank the combined data:** Combine the data from both groups and rank them from smallest to largest.
2. **Sum the ranks of one group:** Sum the ranks of the data points from one of the groups.
3. **Calculate the U statistic:**
    - **Formula:** $U = n1 * n2 + (n1 * (n1 + 1)) / 2 - R1$
    - **Where:**
        - n1 = the number of observations in the first group
        - n2 = the number of observations in the second group
        - R1 = the sum of the ranks for the first group

Example: Consider the following data from two groups:

- **Group A:** 10, 12, 15, 18
- **Group B:** 8, 11, 13, 16

Ranked combined data: 8, 10, 11, 12, 13, 15, 16, 18

- n1 = 4 (number of observations in Group A)
- n2 = 4 (number of observations in Group B)
- R1 = 2 + 4 + 6 + 8 = 20 (sum of ranks for Group A)

Calculation:

- U = 4 * 4 + (4 * (4 + 1)) / 2 - 20
- U = 16 + 10 - 20
- U = 6

Interpretation: The U statistic is 6. To determine if this value is significant, you would compare it to a critical value from a Mann-Whitney U table. If the calculated U is less than or equal to the critical value, you can reject the null hypothesis and conclude that there is a significant difference between the two groups. There are two possible U values for any given dataset, the smaller value is the U test that you will compare to U critical.

Example One: Weight Loss Program

Imagine you want to compare the effectiveness of two different diets on weight loss. You have two independent groups: Group A (10 participants on Diet A) and Group B (11 participants on Diet B). After six weeks, you record the weight loss (in pounds) for each participant.

Group A = 4, 5, 5, 6, 7, 7, 8, 8, 9, 10
Group B = 10, 11, 11, 12, 12, 13, 13, 14, 14, 15, 16

GROUP	LBS LOST	RANK
A	4	1
A	5	2.5
A	5	2.5
A	6	4
A	7	5.5
A	7	5.5
A	8	7.5
A	8	7.5
A	9	9
A	10	10.5
B	10	10.5
B	11	12.5
B	11	12.5
B	12	14.5
B	12	14.5
B	13	16.5
B	13	16.5
B	14	18.5
B	14	18.5
B	15	20
B	16	21

Recall that our hypotheses are as follows:

Null Hypothesis: There is no difference between the groups.

Alternative Hypothesis: There is a difference between the groups.

The next step is to rank the data. First, combine the data of both groups into a single list, then apply the ranking procedures covered above to the data.

Notice that there are several data points that tie. For example, in group A, we have two participants that have a weight loss of 5 lbs. Since these values tie at count 2 and 3, we take the average and rank them both at 2.5.

A similar procedure is applied to the other ties in the data.

The next steps are to count the data points in each group and calculate the rank sums for each group.

Sample Size A: $n1 = 10$

Sample Size B: $n2 = 11$

Rank Sum Group A:

$R1 = 1 + 2.5 + 2.5 + 4 + 5.5 + 5.5 + 7.5 + 7.5 + 9 + 10 = 55.5$

Rank Sum Group B:

$R2 = 10.5 + 12.5 + 12.5 + 14.5 + 14.5 + 16.5 + 16.5 + 18.5 + 18.5 + 20 + 21 = 175.5$

Now it is time to calculate the U statistic for both groups.

Group A: U = n1 * n2 + (n1 * (n1 + 1)) / 2 - R1 = 10*11 + (10*(10+1))/2 - 55.5 = 109.5

Group B: U = n1 * n2 + (n2 * (n2 + 1)) / 2 - R2 = 10*11 + (11*(11+1))/2 - 175.5 = 0.5

Our U test value will be 0.5. Now, we compare that to a U critical obtained from a U table like the one found on this page: https://real-statistics.com/statistics-tables/mann-whitney-table/. Please note that the U tables are set up for specific alpha values, so we'll be looking for the table that is aligned to alpha = 0.5.

From that table, we see that the U-critical is 26. Since our U test value of 0.5 is less than this U-critical of 26, we will reject the null hypothesis. There is sufficient evidence to support the alternative hypothesis that the groups are different in this weight loss study.

Example Two: Teaching Methods

Two independent groups were subjected to different teaching methods to prepare for a licensing test. The scores on the test for the groups are as follows:

Group A: 78, 78, 80, 85, 88, 90, 92

Group B: 30, 65, 75, 77, 78, 80, 82, 99

Next, the data is combined and ranked. Notice that we have three scores of 78 in the list at counts 5, 6, 7. Since we have a three way tie here, we rank each of them as (5 + 6 + 7)/3 = 6. Notice that we have scores of 80 at counts 8 and 9, so these are ranked 8.5.

Group	Test Score	Rank
B	30	1
B	65	2
B	75	3
B	77	4
A	78	6
A	78	6
B	78	6
A	80	8.5
B	80	8.5
B	82	10
A	85	11
A	88	12
A	90	13
A	92	14
B	99	15

The next steps are to count the data points in each group and calculate the rank sums for each group.

Sample Size A: n1 = 7

Sample Size B: n2 = 8

Rank Sum Group A:

R1 = 6 + 6 + 8.5 + 11 + 12 + 13 + 14 = 70.5

Rank Sum Group B:

R2 = 1 + 2 + 3 + 4 + 6 + 8.5 + 10 + 15 = 49.5

Now it is time to calculate the U statistic for both groups.

Group A: U = n1 * n2 + (n1 * (n1 + 1)) / 2 - R1 = 7*8 + (7*(7+1))/2 - 70.5 = 13.5

Group B: U = n1 * n2 + (n2 * (n2 + 1)) / 2 - R2 = 7*8 + (8*(8+1))/2 - 49.5 = 42.5

From the U table, U-critical (alpha = .05) = 10

Since U test of 13.5 is greater than U critical of 10, we fail to reject the null hypothesis. There is not sufficient evidence to support the alternative hypothesis that the group scores are not different.

Performing a Mann-Whitney U Test (called Wilcoxon rank sum test) in Rstudio.

This is the R-code for the weight loss rank data set.

```
x1<-c(4,5,5,6,7,7,8,8,9,10)

x2<-c(10,11,11,12,12,13,13,14,14,15,16)

wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```

Here's the output:

```
Wilcoxon rank sum test with continuity correction

data:  x1 and x2

W = 0.5, p-value = 0.0001376

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

Interpretation:

Since this p-value is less than our alpha of .05, we will reject the null hypothesis in favor of the alternative hypothesis.

This the R-code for the scores rank data set.

```
x1<-c(78,78,80,85,88,90,92)

x2<-c(30,65,75,77,78,80,82,99)

wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```

Here's the output:

```
Wilcoxon rank sum test with continuity correction

data:  x1 and x2

W = 42.5, p-value = 0.1036

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

Interpretation:

Since this p-value is greater than our alpha of .05, we will fail to reject the null hypothesis. There is not sufficient evidence to support the alternative hypothesis.

NOTE: The text `alternative hypothesis: true location shift is not equal to zero` that R reports at the end of the test is just telling you what the alternative hypothesis is. It's not telling you the result of the test.

For the test result, look at the reported *p* value. Based on the *p* value and your selected alpha value, you either reject the null hypothesis or fail to reject the null hypothesis.