Lesson 01: Mann-Whitney U Test

This lesson introduces the **Mann-Whitney U Test**, a non-parametric statistical method used to compare two independent groups. It highlights the test's purpose, key assumptions, and practical applications, focusing on its relevance in research where data may not meet the assumptions of parametric tests. By the end of the lesson, you will understand how to apply the Mann-Whitney U Test effectively.

1. Non-Parametric Tests

When data contain irregularities such as non-normal distributions or outliers, the validity of parametric tests can be compromised. This issue becomes more pronounced with small sample sizes or when key assumptions, like normality, are violated. In such cases, **non-parametric tests**—also known as **"assumption-free tests"**—provide a reliable alternative. By making fewer assumptions about the data, these tests are particularly effective in situations where parametric methods may fail to deliver accurate results.

Although robust (parametric) tests are generally preferable, non-parametric tests are still valuable, especially when using software with limited robust test options or when introducing statistical hypothesis testing concepts.

One of the defining features of non-parametric tests is their reliance on **ranking data** rather than analyzing raw scores. This ranking process involves assigning the smallest value a rank of 1, the next smallest a rank of 2, and so on. This approach helps reduce the impact of outliers and handle irregular data distributions by assigning ranks to values instead of focusing on their exact magnitudes. For instance, if the two highest scores in a dataset of 20 values are 30 and 60, the raw difference is 30, but in ranked form, they are just 1 rank apart (19 and 20), minimizing the effect of extreme values. However, this method comes with a trade-off: by ranking the data, we lose information about the precise differences between individual scores. Tis loss of detail can sometimes result in **reduced statistical power** compared to parametric tests.

But, what is statistical power?

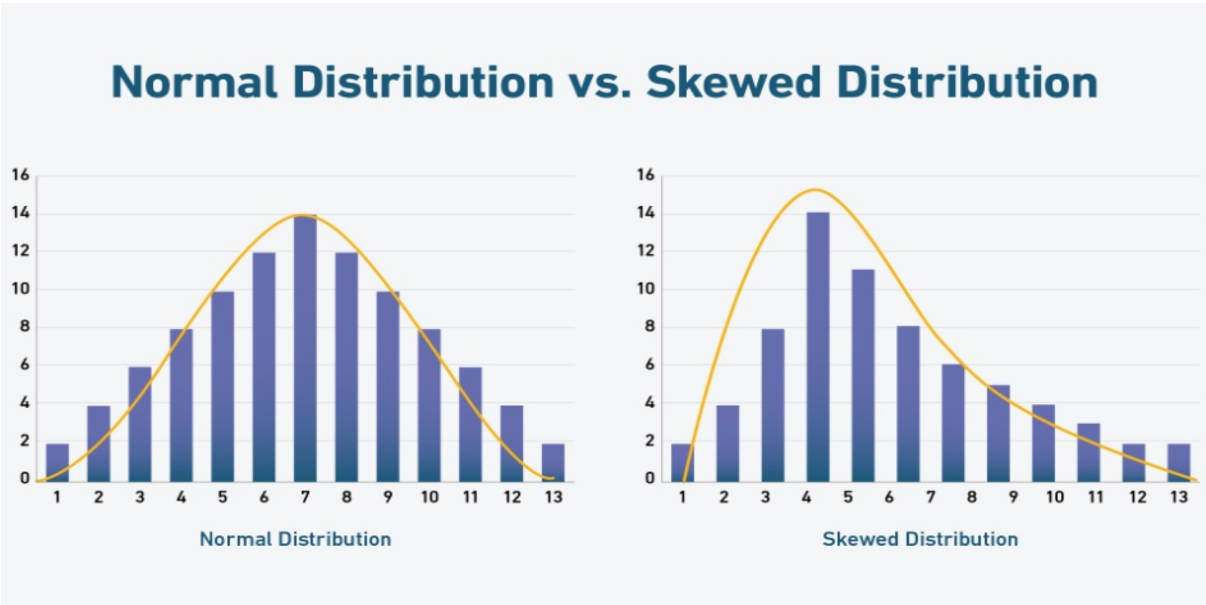
Statistical power refers to a test’s ability **to detect a real effect when it exists**. Parametric tests tend to have higher power when their assumptions—such as normality—are met. If those assumptions are satisfied, parametric tests are generally more effective at identifying genuine effects than non-parametric tests.

A key consideration in understanding power is its connection to the **Type I error rate**—the probability of falsely detecting a significant effect when none exists. Parametric tests are designed to maintain this error rate at 5% when the sampling distribution is normal, allowing for accurate power calculations. However, when the distribution is not normal, the Type I error rate may no longer be 5%, making it difficult to determine the true power of the test.

Although non-parametric tests are sometimes perceived as less powerful, this is primarily true when data meet the strict assumptions required by parametric tests. In situations where these assumptions are violated, non-parametric tests offer a more robust and reliable alternative, ensuring valid conclusions even with non-standard data distributions.

2. Mann-Whitney U Test

Mann-Whitney U test is invaluable for analyzing differences between two independent groups, particularly when the data is ordinal or continuous but not normally distributed. This is often the case in TEFL research, where data on language attitudes, learning strategies, or student engagement levels may not meet the assumptions of normality. It is a robust alternative to the independent t-test, particularly when the data is **skewed**; sample sizes are small; and/or assumptions of normality and equal variances are not met.



Examples:

Student Motivation: A researcher might use the Mann-Whitney U test to compare motivation levels between students in traditional classrooms and those in online environments. If motivation is measured using a Likert scale (ordinal data), this test can determine whether the differences are significant.

Language Proficiency: When evaluating the impact of two different teaching methods (e.g., task-based vs. grammar translation) on proficiency scores that don't follow a normal distribution, this test can identify which method yields higher **median** proficiency levels.

What is the Median:
The median is the middle value in a dataset when the numbers are arranged in ascending order. If there is an even number of data points, the median is the average of the two middle values.

While the **mean** (the arithmetic average of all the data points) is sensitive to extreme values (outliers), the **median** is resistant to outliers and provides a better measure of the "centre" in skewed distributions.

Relevance to Non-Parametric Tests:
Non-parametric tests often use the median because they are designed for data that may not meet the assumptions of normality or may contain outliers, making the median a more robust measure of central tendency.

2.1. KEY ASSUMPTIONS

Before using the Mann-Whitney U test, it's crucial to make sure your data meets certain conditions. Think of these as rules to ensure the test is appropriate for your research question:

A. Type of Data (Dependent Variable):

What you're measuring (the dependent variable) must be ordinal or continuous.
Ordinal: Think of ranked data, like satisfaction scales (e.g., "Strongly Disagree" to "Strongly Agree") or customer ratings. The categories have a clear order.
Continuous: Numerical data that can be measured on a scale, like time, test scores, or weight.

B. Groups to Compare (Independent Variable):

You need two distinct, separate groups. The groups must be categorical, meaning people fall into a category (e.g., male/female, employed/unemployed, smoker/non-smoker)

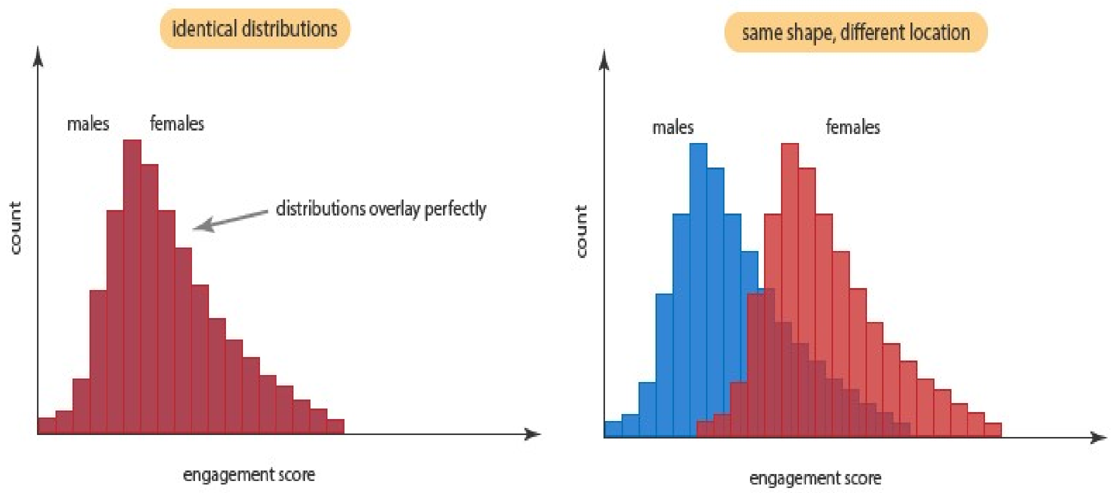
a. Independent Observations:

Observations within each group should not be related and there should be no relationship between observations from different groups. This means that each participant or measurement belongs to only one group, and there's no connection or influence between the data points. One participant can't be in both groups.

b. Distribution Shape (Advanced Concept):

While the Mann-Whitney U test is good for non-normally distributed data, it's also important to know the distribution shapes. Ideally, the shape of the distribution for each group should be similar. The test will detect shifts in distribution, so if distributions are different, your interpretation must acknowledge this. You need to consider if the distributions are similar to interpret the result correctly. This can mean that the location or spread of the two groups is significantly different.

To understand what this means, take a look at the diagram below:



In the two diagrams above, the distributions of scores for 'males' and 'females' share the same shape. In the left diagram, the male distribution (illustrated in blue in the right diagram) is not visible because the two distributions are identical—they completely overlap, with the blue male distribution positioned beneath the red female distribution. In the right diagram, although both distributions maintain the same shape, they differ in location, meaning one group's scores are systematically higher or lower than the other's.

When analyzing your own data, it is highly unlikely that the two distributions will be identical. However, they might have the same or a similar shape. If this is the case, you can use SPSS Statistics to perform a Mann-Whitney U test to compare the medians of your dependent variable (e.g., engagement scores) for the two groups of the independent variable (e.g., males and females). If the distributions have different shapes, the Mann-Whitney U test can only be used to compare mean ranks.

Therefore, when conducting a Mann-Whitney U test, it is essential to use SPSS Statistics to first determine whether the two distributions share the same shape or have different shapes. This involves additional steps in SPSS Statistics to ensure the appropriate interpretation of your results.

Using the wrong test gives you unreliable results. If your data doesn't fit these assumptions, the conclusions from the Mann-Whitney U test might be wrong.