**Mann-Whitney U Test**

(source: <http://blog.excelmasterseries.com/2010/09/mann-whitney-u-test-done-in-excel.html> retrieved 3/24/2014)

Spring 2017 (file no longer at URL)

The Mann-Whitney U Test is a great substitute for the two-sample unpaired t test when it can't be determined that both samples come from normally-distributed populations or if ordinal data is being analyzed. The two-sample unpaired t -test is used to decide if whether there is a difference between the means of two groups or, equivalently, whether or not both groups come from the same population. The two-sample unpaired t-test test is also used to test whether two populations have the same means. The Mann-Whitney U Test compares the sum of ranks and not the means of the two samples.

Major Uses of the Mann-Whitney U Test

The Mann-Whitney U Test is a nonparametric alternative for the two-sample independent t-test. It is therefore often used in the following situations:

1) Determining if there is a difference between two independent groups in one measured variable - An examples of such a test might be determining whether men and women have different satisfaction rating or different white blood cell counts.

2) Determining if two independent groups respond differently to the same treatment - The samples taken would be the "after" scores or the "change" score (the difference between before and after) for the subjects in each independent group. An example of such a test might be determining whether men and women displayed different levels of improvement after undergoing the same training program.

3) Determining if two independent groups of similar subjects respond differently to different treatments - The samples taken would be the "after" scores or the "change" score (the difference between before and after) for the subjects in each independent group. An example of such a test might be determining whether independent groups of similar men displayed different levels of improvement after undergoing the different training programs.

- As mentioned above, the Mann-Whitney U Test is an excellent alternative for the two-sample, unpaired t-test. In this case, each sample is independent of the other. If, on the other hand, the sample pairs are not independent of each other, e.g., before-and-after data pairs, the Sign Test and the Wilcoxon Signed Rank Test would be good nonparametric alternatives. If a one-sample t test is being conducted, i.e., comparing a sample to a known value such as a population, the Sign Test and the Wilcoxon Signed Rank Test would suitable parametric alternatives instead of the Mann Whitney U Test.

- The Mann-Whitney U Test remains the logical choice when the data for the two-sample unpaired t-test are ordinal. Ordinal data are data in which the order matters but the difference between them is undefined. An example of ordinal data is a customer satisfaction ratings scale that goes from 1 to 10. Because the Mann-Whitney U Test works well with ordinal data, it is uniquely well suited to evaluate whether ordinal data such as personal preferences, attitudes, and behaviors have changed. The data can be ordinal but cannot be nominal. Nominal data are simply names or categories of things whose order or rank do not have any meaning, unlike ordinal data.

- The Mann-Whitney U Test compares the sum of ranks and is therefore less likely to produce spurious results than the t-test (which compares means) when outliers are present.

- When normality holds, the Mann-Whitney U Test has an (asymptotic) efficiency of about 0.95 when compared to the t test. For data whose distribution is significantly different than the normal distribution and also for sufficiently large sample sizes, the Mann-Whitney U Test can be considerably more efficient than the t-test.

- Overall, the robustness makes the Mann-Whitney U Test more widely applicable than the t test, and for large samples from the normal distribution, the efficiency loss compared to the t test is only 5%, so Mann-Whitney U Test is often the default test for comparing ratio, interval, and ordinal data having similar distributions.

Test procedures are different for large and small numbers of samples for the MWW (Mann-Whitney U Test/Wilcoxon Rank Sum) Test because the test statistic's distribution converges to the normal distribution as the number of samples goes beyond 10. Parametric tools based on the normal distribution can therefore be used when the number of samples exceeds 10.

Assumptions of the Mann-Whitney U Test

Even though the Mann-Whitney U test is a non-parametric test, there are several assumptions that should be met for test results to be considered valid:

1. The samples drawn from the population are random.

2. Independence within the samples and mutual independence is assumed. All samples in one group must have taken from different subjects or sources than all samples in the other group. The Mann-Whitney U test therefore cannot be used to test whether there is a difference between before-and-after results from the same subjects. The Mann-Whitney U Test can, however, be used to test if there is a difference between "after" results and "change" results (the difference between before and after) between independent groups of subjects.

3. The measured variable that is being tested (the dependent variable) can be continuous (ratio or interval) or categorical. Categorical variables must be ordinal and not nominal. A customer satisfaction rating scale is an example ordinal data. Nominal data are data categories whose order does not have meaning.

4. The independent variable that determines what from group each measurement is from must be a categorical variable that is nominal. An example of a nominal variable would be gender.

5. Populations from which samples were drawn are assumed to have similar distribution shape. The shape of a histogram of each sample will provide an approximation of the shape of the distribution of the population from which the sample was derived. The shapes of the histograms of the two sample groups should be reasonably similar to each other for the Mann-Whitney U Test to produce valid results. The smaller the sample size, the more difficult it is to observe whether the samples are distributed in similar shapes.

Samples are considered large for the Mann-Whitney U Test if sample sizes (n1 and n2) are at least 10 for both samples. When sample size is at least 10 for both samples, the distribution of the test statistic, U, can be approximated by the normal distribution.