Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
https://simfit.org.uk
https://simfit.silverfrost.com

The Mann-Whitney U test is a sort of nonparametric equivalent of the unpaired t test that is used to examine the relative size of observations in two data sets, say X and Y without assumptions about the distributions.

To be precise, the user has two samples (i.e. vectors X and Y) with m and n observations

$$X = (x_1, x_2, \ldots, x_m)$$
$$Y = (y_1, y_2, \ldots, y_n)$$

where the ranks of the two sets of observations within a combined, i.e. pooled, data set can be consulted to see if is is reasonable to conclude that either

- data values in both samples are similar,
- data values in sample X tend to be smaller than those in sample Y, or
- data values in sample X tend to be larger than those in sample Y.

The test is weak unless large samples are used, and is further weakened by ties within the data, that is, multiple observations with the same value.

From the main SIMFIT menu select [A/Z], choose to open the SIMFIT nonparametric testing program **rstest**, then analyze the test files provided to obtain the following results.

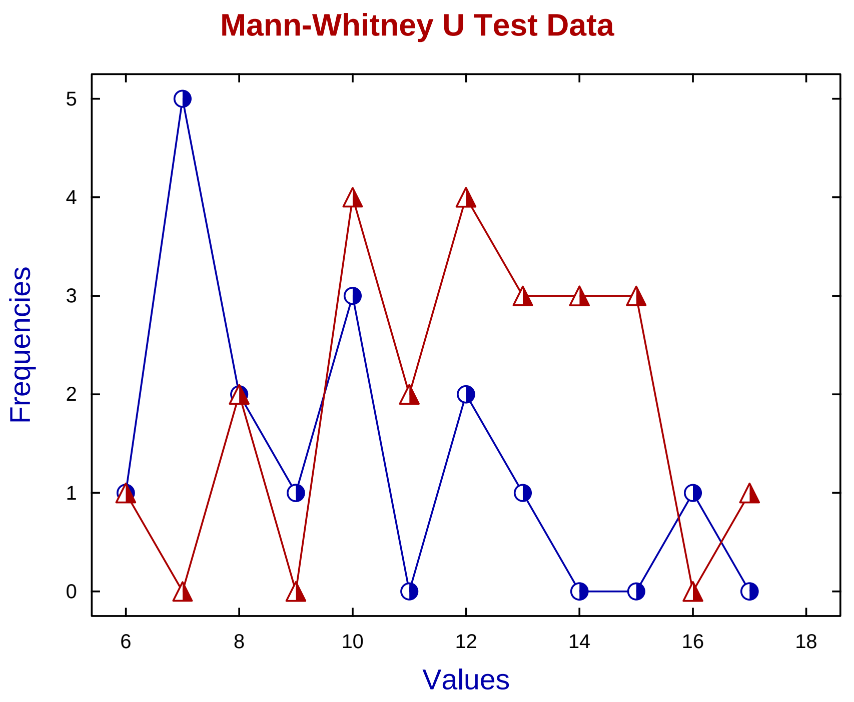
Wilcoxon-Mann-Whitney U test		
X-data: g08ahf.tf1 (Mann-Whitney U test)		
Y-data: g08ahf.tf2 (Mann-Whitney U test)		
X sample size	16	
Y sample size	23	
U	86.00	
Z	-2.804	
H_0: $F(x)$ is equal to $G(y)$ (x and y are comparable)		
as null hypothesis against the alternatives:-		
H_1: $F(x)$ is not equal to $G(y)$ (x and y not comparable)		
p	0.0050	Reject H_0 at 1% significance level
H_2: $F(x) > G(y)$ (x tend to be smaller than y)		
p	0.0025	Reject H_0 at 1% significance level
H_3: $F(x) < G(y)$ (x tend to be larger than y)		
p	0.9977	

Note that U is the Mann-Whitney test statistic which is used to calculate an exact p value, while Z is an approximately normal test statistic and, using SIMFIT program **normal**, we find that $P(Z \leq -2.804) = 0.0025$.

To understand how to interpret the meaning of the above two-tail and one-tail test statistics you can just look at a table of frequencies. This is easily constructed using SIMFIT program **editmt** to rearrange the samples into increasing order as follows, where bracketed values are frequencies.

X	6(1)	7(5)	8(2)	9(1)	10(3)	11(0)	12(2)	13(1)	14(0)	15(0)	16(1)	17(0)
Y	6(1)	7(0)	8(2)	9(0)	10(4)	11(2)	12(4)	13(3)	14(3)	15(3)	16(0)	17(1)

Alternatively, the frequencies can be plotted, as lines and symbols by first using SimFIT program **makfil** to generate plotting files, followed by SimFIT program **simplot** to create the following plot which emphasizes the test results, i.e. the most likely conclusion is that X-sample values tend to be smaller than the Y-sample values.



Using the built-in data editor in **simplot** to move X leftwards and Y rightwards to prevent overlapping, then replacing symbols by bars and suppressing the lines gives the next alternative way to plot the data.