

## Statistics: 2.3 The Mann-Whitney U Test

---

Rosie Shier. 2004.

### 1 Introduction

The Mann-Whitney U test is a non-parametric test that can be used in place of an unpaired t-test. It is used to test the null hypothesis that two samples come from the same population (i.e. have the same median) or, alternatively, whether observations in one sample tend to be larger than observations in the other. Although it is a non-parametric test it does assume that the two distributions are similar in shape.

### 2 Carrying out the Mann-Whitney U test

Suppose we have a sample of  $n_x$  observations  $\{x_1, x_2, \dots, x_n\}$  in one group (i.e. from one population) and a sample of  $n_y$  observations  $\{y_1, y_2, \dots, y_n\}$  in another group (i.e. from another population).

The Mann-Whitney test is based on a comparison of every observation  $x_i$  in the first sample with every observation  $y_j$  in the other sample. The total number of pairwise comparisons that can be made is  $n_x n_y$ .

If the samples have the same median then each  $x_i$  has an equal chance (i.e. probability  $\frac{1}{2}$ ) of being greater or smaller than each  $y_j$ .

So, under the null hypothesis  $H_0 : P(x_i > y_j) = \frac{1}{2}$   
and under the alternative hypothesis  $H_1 : P(x_i > y_j) \neq \frac{1}{2}$

We count the number of times an  $x_i$  from sample 1 is greater than a  $y_j$  from sample 2. This number is denoted by  $U_x$ . Similarly, the number of times an  $x_i$  from sample 1 is smaller than a  $y_j$  from sample 2 is denoted by  $U_y$ . Under the null hypothesis we would expect  $U_x$  and  $U_y$  to be approximately equal.

Procedure for carrying out the test:

1. Arrange all the observations in order of magnitude.
2. Under each observation, write down  $X$  or  $Y$  (or some other relevant symbol) to indicate which sample they are from.
3. Under each  $x$  write down the number of  $y$ s which are to the left of it (i.e. smaller than it); this indicates  $x_i > y_j$ . Under each  $y$  write down the number of  $x$ s which are to the left of it (i.e. smaller than it); this indicates  $y_j > x_i$

4. Add up the total number of times  $x_i > y_j$  — denote by  $U_x$ . Add up the total number of times  $y_j > x_i$  — denote by  $U_y$ . Check that  $U_x + U_y = n_x n_y$ .
5. Calculate  $U = \min(U_x, U_y)$
6. Use statistical tables for the Mann-Whitney U test to find the probability of observing a value of  $U$  or lower. If the test is one-sided, this is your p-value; if the test is a two-sided test, double this probability to obtain the p-value.

NOTE: If the number of observations is such that  $n_x n_y$  is large enough ( $> 20$ ), a normal approximation can be used with  $\mu_U = \frac{n_x n_y}{2}$ ,  $\sigma_U = \sqrt{\frac{n_x n_y (N + 1)}{12}}$ , where  $N = n_x + n_y$ .

**Dealing with ties:** It is possible that two or more observations may be the same. If this is the case we can still calculate  $U$  by allocating half the tie to the  $X$  value and half the tie to the  $Y$  value. However, if this is the case then the normal approximation must be used with an adjustment to the standard deviation. This becomes:

$$\sigma_U = \sqrt{\frac{n_x n_y}{N(N-1)} \times \left[ \frac{N^3 - N}{12} - \sum_{j=1}^g \frac{t_j^3 - t_j}{12} \right]}$$

where  $N = n_x + n_y$   
 $g$  = the number of groups of ties  
 $t_j$  = the number of tied ranks in group  $j$

Note that the Mann-Whitney U test is statistically equivalent to the Wilcoxon rank sum test (not to be confused with the Wilcoxon **signed** rank sum test, which is for paired data).

Example:

The following data shows the age at diagnosis of type II diabetes in young adults. Is the age at diagnosis different for males and females?

Males: 19 22 16 29 24

Females: 20 11 17 12

Solution:

1. Arrange in order of magnitude

Age	11	12	16	17	19	20	22	24	29
M/F	F	F	M	F	M	F	M	M	M
$M > F$			2		3		4	4	4
$F > M$	0	0		1		2			

2. Affix M or F to each observation (see above).
3. Under each M write the number of Fs to the left of it; under each F write the number of Ms to the left of it (see above).
4.  $U_M = 2 + 3 + 4 + 4 + 4 = 17$        $U_F = 0 + 0 + 1 + 2 = 3$
5.  $U = \min(U_M, U_F) = 3$
6. Using tables for the Mann-Whitney U test we get a two-sided p-value of  $p = 0.11$

7. If we use a normal approximation we get:

$$z = \frac{U - \frac{n_x n_y}{2}}{\sqrt{\frac{n_x n_y (N + 1)}{12}}} = \frac{3 - 10}{\sqrt{50/3}} = -1.715$$

This gives a two-sided p-value of  $p = 0.09$ .

The exact test and the normal approximation give similar results. We would conclude that there is no real evidence that the age at diagnosis is different for males and females, although the results are borderline and the lack of statistical significance in this case may just be due to the very small sample. The actual median age at diagnosis is 14.5 years for females and 22 for males, which is quite a substantial difference. In this case it would be advisable to conduct a larger study.

### 3 Carrying out the Mann-Whitney U test in SPSS

- Choose **Analyze**
- Select **Nonparametric Tests**
- Select **2 Independent Samples**
- Highlight your test variable (in our example this would be age) and click on the arrow to move this into the **Test Variable List** box
- Highlight the grouping variable and click on the arrow to move this into the **Grouping Variable** box.
- Click on **Define Groups** and type in the codes that indicate which group an observation belongs to (in our example, the codes which indicate whether a subject is male or female). Click on **Continue**
- Under **Test Type** make sure that **Mann-Whitney U** is selected
- If you want exact probabilities, click on **Exact**, choose **Exact**, then **Continue**
- Click on **OK**

The output will look like this:

#### Ranks

	Sex	N	Mean Rank	Sum of Ranks
Age	Male	5	6.40	32
	Female	4	3.25	13

#### Test Statistics

	Age
Mann-Whitney U	3.000
Wilcoxon W	13.000
z	-1.715
Asymp. Sig. (2-tailed)	0.086
Exact Sig. [2*(1-tailed Sig.)]	0.111
Exact Sig. (2-tailed)	0.111
Exact Sig. (1-tailed)	0.056
Point Probability	0.024

We are interested in the exact p-value (“Exact Sig (2-tailed)”) and the p-value based on the normal approximation (“Asymp Sig (2-tailed)”). If the normal approximation is appropriate then these should be roughly similar.