Mann Whitney U Test

The **Mann Whitney U test** is used to compare two sets of data and find out if there is a significant difference between them. The test uses the median value (the value that lies at the mid-point when the results are in size order) of each set to make this decision.

Note: You can use the **Mann Whitney U test** for

- data that is not normally distributed around the mean
- data sets of fewer than twenty samples

Worked example:

Bedload was collected from two differing points in a river - Site 1 in the middle course and Site 2 in the lower course. Each sample of bedload was measured along its longest edge and the distance recorded. The researcher wanted to see if there was a significant difference between bedload size measurements in each area. They devised the following hypotheses to be tested through the Mann Whitney U test:

H_1 There is a significant difference between the size of the bedload found in the middle course and that found in the lower course.

H_0 There is no significant difference between the size of the bedload found in the middle course and that found in the lower course.

All the data that was collected (from both samples) is placed in size order from the smallest to the largest. One should distinguish between the data that has come from each set - in this case, samples from Site 1 (the middle course) have a '**1**' notation and those from Site 2 (the lower course) have a '**2**'.

These samples are then ranked. If more than one sample has the same value, they share a rank.

Course	**2**	**2**	**2**	**2**	**2**	**1**	**2**	**2**	**2**	**1**	**1**	**1**	**2**	**1**	**2**	**1**	**1**	**1**	**1**	**1**
Size	6	6	7	9	10	10	13	16	18	18	19	20	21	21	23	28	33	38	54	101
Rank	1.5	1.5	3	4	5.5	5.5	7	8	9.5	9.5	11	12	13.5	13.5	15	16	17	18	19	20

You then total the value of the ranks (Σr) for each set of data.

Middle course (Σr_1) : 5.5 + 9.5 + 11 + 12 + 13.5 + 16 + 17 + 18 + 19 + 20 = **141.5**

Lower course (Σr_2) : 1.5 + 1.5 + 3 + 4 + 5.5 + 7 + 8 + 9.5 + 13.5 + 15 = **68.5**

The U value is then calculated for each site:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \Sigma r_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \Sigma r_2$$

n_1 = number of samples from site 1

n_2 = number of samples from site 2

Σr_1 = total of rank values from site 1

Σr_2 = total of rank values from site 2

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \Sigma r_1$$

$$U_1 = 100 + \frac{110}{2} - 141.5$$

$$U_1 = 100 + 55 - 141.5$$

$$U_1 = 13.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \Sigma r_2$$

$$U_2 = 100 + \frac{110}{2} - 68.5$$

$$U_2 = 100 + 55 - 68.5$$

$$U_2 = 86.5$$

The **smaller** of the two U values is taken as the **calculated value**. On its own, this value is meaningless. It needs to be compared with a critical value for the number of samples taken to tell the researcher whether it is significant.

The critical value at a significance level of 0.05 can be found from the table below. The level of significance means the level to which one can be sure that the results are meaningful and did not occur just by chance. In this test, a 95% (or 0.05) level is appropriate.

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2							0	0	0	0	1	1	1	1	1	2	2	2	2
3				0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5		0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	0	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	0	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	0	2	4	6	8	10	13	15	17	19	22	24	26	38	31	34	36	38	41
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

From the significance table we can read off a critical value of **23**. As the **calculated value is less than the critical value, the null hypothesis can be rejected**.

Therefore it appears, in this case, that there is a significant difference between the size of the bedload found in the middle course and that found in the lower course.