M13 Inferential Statistics – Mann-Whitney U Test

Levels of significance and the Mann-Whitney U test

Biological Approach

There are two types of statistic covered here. A **descriptive statistic** is used in all areas of the course. This includes measures of *central tendency* (e.g. mean, median and mode) and measures of *dispersion* (e.g. range). This section of the methodology, however, focuses on **inferential statistics**, those looking to draw inferences about the data, rather than just describe the results. In M9 Inferential Statistics – Spearman's Rank Correlation Coefficient you were introduced to the Spearman's test, which is one form of inferential statistical test, belonging here to the Psychodynamic Approach.

You will here learn about the **Mann-Whitney U test**, the statistical test of the Biological Approach; and later will meet the Chi-squared test of the Learning Approach. To see information on the Chi-squared test as well as a guide on how to choose the correct statistical test for a given scenario, see M15 Inferential Statistics – Chi-Squared Test.

Levels of significance

Statistical tests are designed to see if the null hypothesis (which says that the results of a study are due to chance) is true. So the tests are used to assess whether the findings were found by chance. Every study will have chance factors. The idea is to be able to decide what is down to chance and whether a relationship is significant.

In Psychology, anything that occurs due to chance in more than 1 in 20 cases is not accepted. For example if a study looks at whether or not women gossip more than men and you only study 20 women and more than 1 woman goes against the hypothesis, the null hypothesis cannot be rejected. Similarly, testing 100 people and more than 5 people going against your hypothesis means your alternative hypothesis cannot be accepted.

The **level of significance** shows the probability of the results being down to chance. This will be shown as a decimal value. Anything in Psychology must be at most 1 in 20 due to chance, so the maximum level of significance is 0.05 (the same as 1 in 20). Sometimes, acceptable levels may be even higher, such as 0.025 (1 in 40), 0.01 (1 in 100) or 0.005 (1 in 200). We use the letter p to show probability, so $p \leq 0.05$ means the probability of the results being due to chance are less than 0.05 (basically, 5%).

The choice of the researcher in which level of significance to use depends on what is being tested. A level of 0.05 is fairly lenient, whereas a level of 0.01 is quite strict. It a study has been done previously and the results proved not to be due to chance at 0.05, it might be worth repeating the second study at a stricter significance level. However, a new study which has not been seen before may wish to begin with a fairly gentle level of 0.05. The choice of level of significance can also depend on other factors, such as the seriousness of the consequences of the findings, such as with a new drugs trial or for a new educational scheme.

Mann-Whitney U test

The Mann-Whitney U test is used to see if findings are statistically significant for studies where an independent groups design has been used, and the type of data collected is *ordinal* or *interval*. You will meet more on types of data and when to choose each test in M15 Inferential Statistics – Chi-Squared Test.

Whereas the Spearman's Rank correlation coefficient is used to test for a relationship between two variables, both the Mann-Whitney U test and the Chi-squared test look for differences between two groups.

The formula for the test looks rather complicated, but as with the Spearman's test we can break it down into steps:

$$U_A = N_A N_B + \frac{N_A(N_A+1)}{2} - \Sigma R_A \qquad U_B = N_A N_B + \frac{N_B(N_B+1)}{2} - \Sigma R_B$$

In our example, we will use a study looking at males and females completing jigsaw puzzles. Eight males and nine females were asked to complete a jigsaw puzzle. The hypothesis is that males will be faster completing the jigsaws because they are better at visuospatial tasks. Each participant's 'score' was taken – the score is the amount of time, in seconds, it took the participant to complete the puzzle.

Participant Group 1 (Males)	Score	Participant Group 2 (Females)	Score
1	95	1	100
2	78	2	123
3	102	3	89
4	79	4	140
5	84	5	97
6	93	6	110
7	62	7	150
8	92	8	104
		9	96

Step 1: Rank the scores for the participants as a whole

Unlike with Spearman's where you rank each series of data by itself, here, we are going to rank the scores of all seventeen participants as if they were from one group. The lowest score (i.e. fastest time) gets rank 1, and so on

Step 2: Label the groups N_A and N_B and work out the value of N for each group

If one group is smaller than the other, the smaller group will be N_A

Step 3: Taking the groups separately, add together the ranks for each group

This is described as ΣR_A *for Group A, and* ΣR_B *for Group B*

Participant Group 1 (Males) $N_A = 8$	Score	Rank	Participant Group 2 (Females) $N_B = 9$	Score	Rank
1	95	8	1	100	11
2	78	2	2	123	15
3	102	12	3	89	5
4	79	3	4	140	16
5	84	4	5	97	10
6	93	7	6	110	14
7	62	1	7	150	17
8	92	6	8	104	13
	ΣR_A =	43	9	96	9
				ΣR_B =	110

Step 4: Use the formula to calculate a Mann-Whitney U test result for Group A

$U_A = N_A N_B + (N_A(N_A + 1))/2 - \Sigma R_A$ $\quad$ $U_A = (8 \times 9) + (8 \times 9)/2 - 43$ $\quad$ $U_A = (72 + 36) - 43$ $\quad$ $U_A = 65$

Step 5: Use the formula to calculate the result for Group B

$U_B = N_A N_B + (N_B(N_B + 1))/2 - \Sigma R_B$ $\quad$ $U_B = (8 \times 9) + (9 \times 10)/2 - 110$ $\quad$ $U_B = (72 + 45) - 110$ $\quad$ $U_B = 7$

Step 6: Take the smaller of U_A and U_B and label that value as U

In our example, 7 is smaller than 65, so U_B *becomes* U *(so in our example,* $U = 7$*)*

The value for U can then be checked against the critical value tables to see if the findings are statistically significant. U must be less than or equal to the critical value in the table. An exemplar table is shown here.

Level of significance $p \leq 0.05$

	N_B ▶	8	9	10	11	12
N_A	7	13	15	17	19	21
▼	8	15	18	20	23	26
	9	18	21	24	27	30
	10	20	24	27	30	33
	11	23	27	31	34	37

Our value for U was 7. The critical value is 18 (which has been highlighted in the table) as N_A was 8 and N_B was 9. Because 7 is less than 18, we can say that the results are statistically significant, and that they support the alternative hypothesis, that males are better at jigsaw puzzles than women.