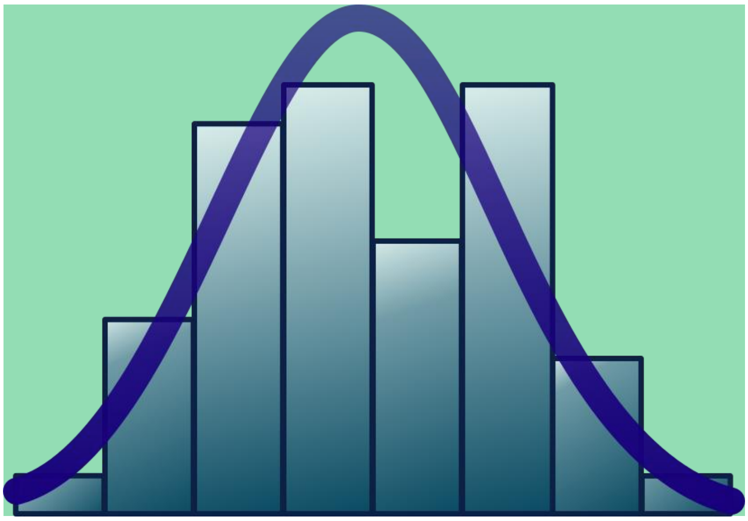
# Inferential Statistics Booklet



## Calculating and Interpreting the Tests

## OCR Psychology

# Mann Whitney U Test

The Mann-Whitney test is a non-parametric statistical test of **difference** that allows psychologists to determine if their results are significant. It is used in studies that have an **independent groups design**, where the data collected is at least **ordinal**.

## Example:

A study to investigate gender differences in the ability to memorise pictures of animals in short-term memory. Based on previous research it was predicted that females will recall significantly more pictures than males. This is a one-tailed prediction.

A field study was conducted in which 20 participants (10M, 10F) were selected by opportunity on a Wednesday afternoon in the college canteen. Once consent was gained they were issued with a copy of the memory test containing pictures of 15 animals. Each participant was tested individually and given one minute to memorise the pictures and a further one minute to immediately recall.

Maximum score on the memory test was 15.

Here are their results,

| Males | Rank (Ra) | Females | Rank (Rb) |
|---|---|---|---|
| 10 | **4.5** | 12 | **9** |
| 8 | **1.5** | 13 | **14** |
| 10 | **4.5** | 14 | **18** |
| 12 | **9** | 12 | **9** |
| 8 | **1.5** | 13 | **14** |
| 9 | **3** | 12 | **9** |
| 11 | **6** | 13 | **14** |
| 13 | **14** | 14 | **18** |
| 12 | **9** | 14 | **18** |
| 13 | **14** | 15 | **20** |
| **Mean = 10.6** | | **Mean = 13.2** | |
| | **Total = 67** | | **Total = 143** |

**The mean scores show that females (13.2) performed better on the memory recall test compared to males (10.6).**

a) Rank the data for each group and add up the total of each set of ranks (**remember to rank the groups as one whole data set with the lowest number receiving a rank of 1**). You have a total of 20 positions to give away. If participants share positions, allocate an average to each one.

b) Use the formula for Mann Whitney to calculate the value of U. Do each sum separately. Remember n = number of participants in that group (10M, 10F).

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

Or

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

**Ua = 67 – 10 (10+1)/ 2**

**Ua = 67 – 55**

**Ua = 12**

**Ub = 143 – 10 (10+1)/ 2**

**Ub = 143 – 55**

**Ub = 88**

**Therefore U = 12**

c) Use the critical value table to assess the significance in the findings based on a one-tailed test when **p<0.05**, and state whether the results are significant or not.

## Mann Whitney U Test - Critical Values Table (One-tailed test)

| $n_2$ | Probability Range | $n_1$ | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 3 | 0.05 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 4 | 5 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 | 11 |
| | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4 | 4 | 4 | 5 |
| 4 | 0.05 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 | 17 | 18 |
| | 0.01 | 0 | 0 | 0 | 1 | 1 | 2 | 3 | 3 | 4 | 5 | 5 | 6 | 7 | 7 | 8 | 9 | 9 | 10 |
| 5 | 0.05 | 1 | 2 | 4 | 5 | 6 | 8 | 9 | 11 | 12 | 13 | 15 | 16 | 18 | 19 | 20 | 22 | 23 | 25 |
| | 0.01 | 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 6 | 0.05 | 2 | 3 | 5 | 7 | 8 | 10 | 12 | 14 | 16 | 17 | 19 | 21 | 23 | 25 | 26 | 28 | 30 | 32 |
| | 0.01 | 0 | 1 | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 15 | 16 | 18 | 19 | 20 | 22 |
| 7 | 0.05 | 2 | 4 | 6 | 8 | 11 | 13 | 15 | 17 | 19 | 21 | 24 | 26 | 28 | 30 | 33 | 35 | 37 | 39 |
| | 0.01 | 0 | 1 | 3 | 4 | 6 | 7 | 9 | 11 | 12 | 14 | 16 | 17 | 19 | 21 | 23 | 24 | 26 | 28 |
| 8 | 0.05 | 3 | 5 | 8 | 10 | 13 | 15 | 18 | 20 | 23 | 26 | 28 | 31 | 33 | 36 | 39 | 41 | 44 | 47 |
| | 0.01 | 0 | 2 | 4 | 6 | 7 | 9 | 11 | 13 | 15 | 17 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 34 |
| 9 | 0.05 | 4 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | 36 | 39 | 42 | 45 | 48 | 51 | 54 |
| | 0.01 | 1 | 3 | 5 | 7 | 9 | 11 | 14 | 16 | 18 | 21 | 23 | 26 | 28 | 31 | 33 | 36 | 38 | 40 |
| 10 | 0.05 | 4 | 7 | 11 | 14 | 17 | 20 | 24 | 27 | 31 | 34 | 37 | 41 | 44 | 48 | 51 | 55 | 58 | 62 |
| | 0.01 | 1 | 3 | 6 | 8 | 11 | 13 | 16 | 19 | 22 | 24 | 27 | 30 | 33 | 36 | 38 | 41 | 44 | 47 |
| 11 | 0.05 | 5 | 8 | 12 | 16 | 19 | 23 | 27 | 31 | 34 | 38 | 42 | 46 | 50 | 54 | 57 | 61 | 65 | 69 |
| | 0.01 | 1 | 4 | 7 | 9 | 12 | 15 | 18 | 22 | 25 | 28 | 31 | 34 | 37 | 41 | 44 | 47 | 50 | 53 |
| 12 | 0.05 | 5 | 9 | 13 | 17 | 21 | 26 | 30 | 34 | 38 | 42 | 47 | 51 | 55 | 60 | 64 | 68 | 72 | 77 |
| | 0.01 | 2 | 5 | 8 | 11 | 14 | 17 | 21 | 24 | 28 | 31 | 35 | 38 | 42 | 46 | 49 | 53 | 56 | 60 |
| 13 | 0.05 | 6 | 10 | 15 | 19 | 24 | 28 | 33 | 37 | 42 | 47 | 51 | 56 | 61 | 65 | 70 | 75 | 80 | 84 |
| | 0.01 | 2 | 5 | 9 | 12 | 16 | 20 | 23 | 27 | 31 | 35 | 39 | 43 | 47 | 51 | 55 | 59 | 63 | 67 |
| 14 | 0.05 | 7 | 11 | 16 | 21 | 26 | 31 | 36 | 41 | 46 | 51 | 56 | 51 | 66 | 71 | 77 | 82 | 87 | 92 |
| | 0.01 | 2 | 6 | 10 | 13 | 17 | 22 | 26 | 30 | 34 | 38 | 43 | 47 | 51 | 56 | 60 | 65 | 69 | 73 |
| 15 | 0.05 | 7 | 12 | 18 | 23 | 28 | 33 | 39 | 44 | 50 | 55 | 61 | 66 | 72 | 77 | 83 | 88 | 94 | 100 |
| | 0.01 | 3 | 7 | 11 | 15 | 19 | 24 | 28 | 33 | 37 | 42 | 47 | 51 | 56 | 61 | 66 | 70 | 75 | 80 |
| 16 | 0.05 | 8 | 14 | 19 | 25 | 30 | 36 | 42 | 48 | 54 | 60 | 65 | 71 | 77 | 83 | 89 | 95 | 101 | 107 |
| | 0.01 | 3 | 7 | 12 | 16 | 21 | 26 | 31 | 36 | 41 | 46 | 51 | 56 | 61 | 66 | 71 | 76 | 82 | 87 |
| 17 | 0.05 | 9 | 15 | 20 | 26 | 33 | 39 | 45 | 51 | 57 | 64 | 70 | 77 | 83 | 89 | 96 | 102 | 109 | 115 |
| | 0.01 | 4 | 8 | 13 | 18 | 23 | 28 | 33 | 38 | 44 | 49 | 55 | 60 | 66 | 71 | 77 | 82 | 88 | 93 |
| 18 | 0.05 | 9 | 16 | 22 | 28 | 35 | 41 | 48 | 55 | 61 | 68 | 75 | 82 | 88 | 95 | 102 | 109 | 116 | 123 |
| | 0.01 | 4 | 9 | 14 | 19 | 24 | 30 | 36 | 41 | 47 | 53 | 59 | 65 | 70 | 76 | 82 | 88 | 94 | 100 |
| 19 | 0.05 | 10 | 17 | 23 | 30 | 37 | 44 | 51 | 58 | 65 | 72 | 80 | 87 | 94 | 101 | 109 | 116 | 123 | 130 |
| | 0.01 | 4 | 9 | 15 | 20 | 26 | 32 | 38 | 44 | 50 | 56 | 63 | 69 | 75 | 82 | 88 | 94 | 101 | 107 |
| 20 | 0.05 | 11 | 18 | 25 | 32 | 39 | 47 | 54 | 62 | 69 | 77 | 84 | 92 | 100 | 107 | 115 | 123 | 130 | 138 |
| | 0.01 | 5 | 10 | 16 | 22 | 28 | 34 | 40 | 47 | 53 | 60 | 67 | 73 | 80 | 87 | 93 | 100 | 107 | 114 |

**Rule:** If the ***calculated value*** is ***equal to*** or ***lower than*** the ***critical value*** then the results are ***significant at the 5% level of probability***.

**Critical value = 27 (when na =10, nb = 10, p<0.05)**

Complete the **significance statement** (delete some parts);

"The calculated value of U is **12**. This is **less than** the critical value of **27** (when p<0.05). Therefore this result **is significant**. There is a **less than** 5% probability that the results are due to chance."

d) State which hypothesis would be accepted and why. Interpret the result in terms of the context in the question.

**The research hypothesis would be accepted and the null rejected. This tells us that there is a significant difference between males and females and their ability to recall pictures of animals on a memory test.**

---

## Wilcoxon Test

The Wilcoxon test is a non-parametric statistical test of **difference** that allows a researcher to determine the significance of their findings. It is used in studies that have a **repeated measures** or **matched pairs design**, where the data collected is at least **ordinal**.

### Example:

A study to investigate if revision aids help with  memory recall. Based on previous research it was predicted that revision aids (cue cards) will significantly improve recall compared to not using any revision aids. This is a one-tailed prediction.

A lab experiment was conducted in which 10 participants (5M, 5F) were selected by volunteer sampling after an advert was placed in the sixth form library. Once consent was gained they were first allocated into condition 1 (no revision aid). Here they were given a short memory test containing 20 random words. Each participant was tested individually and given one minute to memorise the word list and a further one minute to immediately recall.

On completion they were given some revision aids (cue cards) and asked to take a short break to practice the techniques ready for the second test. One hour later they returned to the lab and completed condition 2. Here they were given 5 minutes to read the cue cards before being given a second memory test. This contained a list of 20 new words, for which they had one minute to memorise and a further minute to recall.

Maximum score on each memory test was 20.

Here are their results,

| Participant | 1 (no revision aid) | 2 (cue cards) | Difference | Rank of Difference |
|---|---|---|---|---|
| 1 | 12 | 20 | **-8** | **9** |
| 2 | 11 | 14 | **-3** | **3** |
| 3 | 10 | 15 | **-5** | **5** |
| 4 | 18 | 17 | **1** | **1** |
| 5 | 10 | 16 | **-6** | **6** |
| 6 | 14 | 12 | **2** | **2** |
| 7 | 13 | 13 | **0** | **0** |
| 8 | 16 | 20 | **-4** | **4** |
| 9 | 12 | 19 | **-7** | **7.5** |
| 10 | 18 | 11 | **7** | **7.5** |
| **Mean =** | **13.4** | **15.7** | | |
| | | | | **Rank Positive = 10.5**<br>**Rank Negative = 34.5** |

a) Calculate the mean for each condition (no revision aids versus cue cards) and state what the results tell us.

**The results show that on average recall of words is greater when revision aids like cue cards are used (15.7) compared to when they are not (13.4).**

b) Calculate the Wilcoxon test by completing the table and find the value of T (lowest of the ranks, either positive or negative). Ignore any scores with values of 0 or differences with a 0 score/ do not rank these. Also ignore the signs when calculating the difference.

Next add up the total for the rank scores that had a positive in the difference column, then add up the total for the rank scores that had a negative in the difference column.

**Rank Positive =** **10.5**

**Rank Negative =** **34.5**

**Take the lowest number. Therefore the value of T = 10.5**

c) Use the critical value table to assess the significance in the findings, when n =10, using a one-tailed test when **$p<0.05$**, and state whether the results are significant or not.

**Critical Values of the Wilcoxon Signed Ranks Test**

| n | Two-Tailed Test | | One-Tailed Test | |
|---|---|---|---|---|
| | α = .05 | α = .01 | α = .05 | α = .01 |
| 5 | -- | -- | 0 | -- |
| 6 | 0 | -- | 2 | -- |
| 7 | 2 | -- | 3 | 0 |
| 8 | 3 | 0 | 5 | 1 |
| 9 | 5 | 1 | 8 | 3 |
| 10 | 8 | 3 | 10 | 5 |
| 11 | 10 | 5 | 13 | 7 |
| 12 | 13 | 7 | 17 | 9 |
| 13 | 17 | 9 | 21 | 12 |
| 14 | 21 | 12 | 25 | 15 |
| 15 | 25 | 15 | 30 | 19 |
| 16 | 29 | 19 | 35 | 23 |
| 17 | 34 | 23 | 41 | 27 |
| 18 | 40 | 27 | 47 | 32 |
| 19 | 46 | 32 | 53 | 37 |
| 20 | 52 | 37 | 60 | 43 |
| 21 | 58 | 42 | 67 | 49 |
| 22 | 65 | 48 | 75 | 55 |
| 23 | 73 | 54 | 83 | 62 |
| 24 | 81 | 61 | 91 | 69 |
| 25 | 89 | 68 | 100 | 76 |
| 26 | 98 | 75 | 110 | 84 |
| 27 | 107 | 83 | 119 | 92 |
| 28 | 116 | 91 | 130 | 101 |
| 29 | 126 | 100 | 140 | 110 |
| 30 | 137 | 109 | 151 | 120 |

**Rule:** If the ***calculated value*** is ***equal to*** or ***lower than*** the ***critical value*** then the results are ***significant at the 5% level of probability***.

**The critical value (when n = 10, p<0.05) is 10.**

Complete the **significance statement** (delete some parts);

"The calculated value of the Wilcoxon is **10.5.** This is **more than** the critical value of **10** (when p<0.05). Therefore this result is **not significant**. There is a **more than** 5% probability that the results are due to chance."

d) State which hypothesis would be accepted and why. What does this tell us about the results, use the context in the question?

**The null hypothesis would be accepted and the research hypothesis rejected. This tells us that there is not a significant difference between the use of revision aids (cue cards) and not when completing a memory test of words.**

# Chi Squared Test

The chi-squared test is a non-parametric statistical test of **difference** or **association** that allows researchers to see if their results are significant. It is used for studies that have an **independent groups design**, where the data collected is **nominal** (in categories).

## Example:

A study to investigate whether students or teachers make healthier choices for lunchtime meals. A non-participant, covert, naturalistic observation was conducted, where event sampling was used to gather data from the sixth form canteen on a Monday lunchtime from 12-1pm. Students and teachers were both observed as to what options of food or drink they purchased during this time. Two observers were seated close by to the cashier where everyone had to pass. Both had a copy of the same behavioural categories, covering options for healthy food an unhealthy food sold at the canteen. Only food purchased was counted.

Below is the frequency table of students and teachers and whether they purchased healthy or unhealthy food options (observed values).

a) Label each cell with – A, B, C and D (cell A is done for you).

| | Students | Teachers | Total |
|---|---|---|---|
| Healthy Choices | 12 (A) | 22 (B) | **34** |
| Unhealthy Choices | 26 (C) | 5 (D) | **31** |
| **Column total** | **38** | **27** | **65** |

**The table is showing that students are more likely to make unhealthy choices and teachers are more likely to make healthy choice**

b) Add up the totals for each row and each column in the **contingency table** above. The grand total (shaded in grey) will total the same whether it's added from rows or columns.

c) Use this table to fill in the gaps by completing the steps below:

| Cell | Observed value (O) | Expected value (E) | O-E | $(O\text{-}E)^2$ | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|---|
| A | **12** | **19.88** | **-7.88** | **62.09** | **3.12** |
| B | **22** | **14.12** | **7.88** | **62.09** | **4.40** |
| C | **26** | **18.12** | **7.88** | **62.09** | **3.43** |
| D | **5** | **12.88** | **-7.88** | **62.09** | **4.82** |

1. For each cell in the original contingency table, calculate the expected value (E) as follows;

$$\frac{rowtotalxcolumntotal}{grandtotal}$$

**Cell A) 34x38/ 65 = 19.88**

**Cell B) 34x27/ 65 = 14.12**

**Cell C) 31x38/ 65 = 18.12**

**Cell D) 31x27/ 65 = 12.88**

2. For each cell subtract the expected value from the observed value (O-E).
3. Square the values for the previous step $(O\text{-}E)^2$.
4. Divide the result of the previous step by E.
5. Add up all the values in the last column (shaded grey) to give a final value for Chi-squared $\chi^2$.

This is represented by the formula;

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \quad \mathbf{15.77}$$

d) To decide if our calculated value for $\chi^2$ is significant, you need to work out the degrees of freedom for the contingency table using the following formula;

**df = (no. rows - 1) x (no. columns – 1)**

**df = (2-1) x (2-1)**

**df = 1**

Calculate these for this study (look at the first raw data table).

Once you have these two values, you can use the $\chi^2$ table to check if the value for $\chi^2$ is higher than the critical value given in the table. If it is, then the result is significant at the level given. Make sure you look at the probability level of 5% ($p < 0.05$).

| TABLE IV | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Chi-Square ($\chi^2$) Distribution | | | | | | | | |
| | Area to the Right of Critical Value | | | | | | | |
| Degrees of Freedom | 0.99 | 0.975 | 0.95 | 0.90 | 0.10 | 0.05 | 0.025 | 0.01 |
| 1 | — | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 |
| 2 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 |
| 3 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 |
| 4 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 |
| 5 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.071 | 12.833 | 15.086 |
| 6 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 |
| 7 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 |
| 8 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 |
| 9 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 |
| 10 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 |
| 11 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 |
| 12 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 |
| 13 | 4.107 | 5.009 | 5.892 | 7.042 | 19.812 | 22.362 | 24.736 | 27.688 |
| 14 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 |
| 15 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 |
| 16 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 |
| 17 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 |
| 18 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 |
| 19 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 |
| 20 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 |
| 21 | 8.897 | 10.283 | 11.591 | 13.240 | 29.615 | 32.671 | 35.479 | 38.932 |
| 22 | 9.542 | 10.982 | 12.338 | 14.042 | 30.813 | 33.924 | 36.781 | 40.289 |
| 23 | 10.196 | 11.689 | 13.091 | 14.848 | 32.007 | 35.172 | 38.076 | 41.638 |
| 24 | 10.856 | 12.401 | 13.848 | 15.659 | 33.196 | 36.415 | 39.364 | 42.980 |
| 25 | 11.524 | 13.120 | 14.611 | 16.473 | 34.382 | 37.652 | 40.646 | 44.314 |
| 26 | 12.198 | 13.844 | 15.379 | 17.292 | 35.563 | 38.885 | 41.923 | 45.642 |
| 27 | 12.879 | 14.573 | 16.151 | 18.114 | 36.741 | 40.113 | 43.194 | 46.963 |
| 28 | 13.565 | 15.308 | 16.928 | 18.939 | 37.916 | 41.337 | 44.461 | 48.278 |
| 29 | 14.257 | 16.047 | 17.708 | 19.768 | 39.087 | 42.557 | 45.722 | 49.588 |
| 30 | 14.954 | 16.791 | 18.493 | 20.599 | 40.256 | 43.773 | 46.979 | 50.892 |

**Rule:** If the ***calculated value*** is ***equal to*** or ***greater than*** the ***critical value*** then the results are ***significant at the 5% level of probability***.

Complete the **significance statement** (delete some parts);

"The calculated value of $\chi^2$ is **15.77**. This is **less more** than the critical value of **3.841** (when **$p<0.05$**). Therefore this result **is significant**. There is a **less than** 5% probability that the results are due to chance."

f) State which hypothesis would be accepted and why. What does this tell us about the results, use the context in the question?

**The research hypothesis would be accepted and the null rejected. This tells us that there is a significant difference between students and teachers and the healthy or unhealthy food choices they make.**

# Sign Test

The sign test is a non-parametric statistical test of **difference** that allows a researcher to determine the significance of their investigation. It is used in studies that have used a **repeated measures design**, where the data collected is **nominal**.

## Example:

A study to see whether students prefer horror or comedy films. It was predicted that more students would prefer comedy films. A sample of 10 students were selected by opportunity in the college study area. Once consent was obtained, they were asked two questions.

Each participant was asked if they liked horror films and then if they liked comedy films. By doing this nominal data was gathered by recording YES or NO to each question. This was a repeated measures design as each participant was asked two questions.

The results are in the table below:

**Table (1) Results to show participants preference for horror or comedy films.**

| Participant | Horror Films (Condition A) | Comedy Films (Condition B) |
|---|---|---|
| 1 | yes | no |
| 2 | no | yes |
| 3 | no | yes |
| 4 | yes | no |
| 5 | no | yes |
| 6 | no | no |
| 7 | yes | no |
| 8 | no | yes |
| 9 | no | yes |
| 10 | no | no |

a) Go to the next table and add the appropriate signs to the participant scores; either positive (+) or negative (-) using the key below.

**KEY:** If condition A (horror) is YES put a minus and if condition B (comedy) is a YES put a plus.

As the prediction is that comedy films will be preferred they are allocated the plus sign.

Ignore participants with the same answer (this does not tell us which film they prefer).

The signs are allocated like this as it supports our hypothesis. We think participants will be more likely to say yes to the comedy films.

**(Yes in Horror Films = minus sign, Yes in Comedy Films = plus sign)**

| Participant | Horror Films (Condition A) | Comedy Films (Condition B) | Sign (+/ -) |
|---|---|---|---|
| 1 | yes | no | - |
| 2 | no | yes | + |
| 3 | no | yes | + |
| 4 | yes | no | - |
| 5 | no | yes | + |
| 6 | no | no | |
| 7 | yes | no | - |
| 8 | no | yes | + |
| 9 | no | yes | + |
| 10 | no | no | |

**N.B. Ignore any results where the scores are the same in both conditions.**

**Remove these participants from the sample size, therefore now N= 8**

b) Now count the number of each positive and negative sign assigned to each participant's scores.

(+) TOTAL = **5**

(-) TOTAL = **3**

The smallest of signs (+ or -) is the overall binomial test result. **The observed value of S = 3**

Now look for the 0.05 level of significance and find the critical value in the table below;

| N | 0.05 | 0.01 |
|---|---|---|
| 5 | 0 | |
| 6 | 0 | 0 |
| 7 | 0 | 0 |
| 8 | 1 | 0 |
| 9 | 1 | 1 |
| 10 | 1 | 1 |
| 11 | 2 | 1 |
| 12 | 2 | 2 |
| 13 | 3 | 2 |
| 14 | 3 | 2 |
| 15 | 3 | 3 |

<u>**Rule:**</u> If the ***calculated value*** is ***equal to*** or ***lower than*** the ***critical value*** then the results are ***significant at the 5% level of probability***.

**NB:**

N = number of participants whose scores were used. In a sign test you ignore any results were there is the same score in both conditions for example, “no no” or “yes yes”.

Therefore in this example, the number of participants scores used would be **8 participants.**

**Therefore, the critical sign test value = 1**

c) Write a **statement of significance** (delete some parts);

“The calculated value of S is **3**. This is **more than** the critical value of **1** (when $p<0.05$). Therefore this result is **not significant**. There is a **more than** 5% probability that the results are due to chance.”

d) Does this mean the study was significant? Explain in terms of the context in the question.

**This means that there was not a significant difference between the types of films preferred by students (horror or comedy).**

## Spearman’s Rho Test

Spearman’s rho is a non-parametric statistical test of **correlation** that allows a researcher to determine the significance of their investigation. It is used in studies that are looking for a relationship, where the data is at least **ordinal**.

## Example:

An investigation was carried out to see if there was a relationship between self-ratings of aggression and the number of siblings a person has. It was predicted that having more siblings can make people more aggressive, as they are more likely to have to compete within the family.

Participants were asked to volunteer for a study during a two-day period in the sixth form. A poster was placed in the library and the canteen asking for males and females to complete a short questionnaire. 15 students volunteered and were each asked to rate themselves on a scale 0-5 for overall aggression (0 = never aggressive, 5 = always aggressive). They were also asked to give the number of siblings they had. Ordinal data from the two questions was collected to analyse the relationship between the two variables.

The results are in the table below:

| Student | A (self-rating of aggression) | Rank A | B (number of siblings) | Rank B | Rank of Difference (d) | Rank of Difference Squared (d2) |
|---|---|---|---|---|---|---|
| 1 | 3 | **11** | 3 | **13** | **-2** | **4** |
| 2 | 2 | **7** | 2 | **8** | **-1** | **1** |
| 3 | 5 | **15** | 4 | **15** | **0** | **0** |
| 4 | 1 | **3.5** | 2 | **8** | **-4.5** | **20.25** |
| 5 | 2 | **7** | 2 | **8** | **-1** | **1** |
| 6 | 3 | **11** | 3 | **13** | **-2** | **4** |
| 7 | 2 | **7** | 1 | **4** | **3** | **9** |
| 8 | 0 | **1.5** | 0 | **2** | **-0.5** | **0.25** |
| 9 | 2 | **7** | 2 | **8** | **-1** | **1** |
| 10 | 2 | **7** | 2 | **8** | **-1** | **1** |
| 11 | 1 | **3.5** | 0 | **2** | **1.5** | **2.25** |
| 12 | 0 | **1.5** | 0 | **2** | **-0.5** | **0.25** |
| 13 | 4 | **13.5** | 3 | **13** | **0.5** | **0.25** |
| 14 | 3 | **11** | 2 | **8** | **3** | **9** |
| 15 | 4 | **13.5** | 2 | **8** | **5.5** | **30.25** |
| | | | | | | **Sum = 83.5** |

a) Calculate the ranks for each group **(remember to rank the data as two separate groups- do aggression, then do siblings)**. Remember you have 15 positions to give away, and some scores will share positions. Rank all scores, even if they are 0.

Next calculate the difference between the ranks (d) and finally calculate the last column (d2). You will need the sum of this column (d2) to substitute into the formula below (shaded in grey).

b) Calculate the Spearman's Rho test by using the formula to find the value of rho (Rs). It is easier to leave the 1 until you have calculated the other part of the equation.

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

**rs = 1 – 6x83.5/ 15(225-1)**

**rs = 1 – 501/ 3360**

**rs = 1 – 0.149**

**rs = 0.851**

You should have a number between -1 and +1. This is the correlation coefficient = **0.851**

c) Use the critical value table to assess the significance in the findings when n =15 using a one-tailed test when **p<0.05**, and state whether the results are significant or not.

**Critical values for Spearman's rank**

| | Level of significance for a one-tailed test | | | | |
|---|---|---|---|---|---|
| | **0.05** | **0.025** | **0.01** | **0.005** | **0.0025** |
| | Level of significance for a two-tailed test | | | | |
| *N* | **0.10** | **0.05** | **0.025** | **0.01** | **0.005** |
| **5** | 0.900 | 1.000 | 1.000 | 1.000 | 1.000 |
| **6** | 0.829 | 0.886 | 0.943 | 1.000 | 1.000 |
| **7** | 0.714 | 0.786 | 0.893 | 0.929 | 0.964 |
| **8** | 0.643 | 0.738 | 0.833 | 0.881 | 0.905 |
| **9** | 0.600 | 0.700 | 0.783 | 0.833 | 0.867 |
| **10** | 0.564 | 0.648 | 0.745 | 0.794 | 0.830 |
| **11** | 0.536 | 0.618 | 0.709 | 0.755 | 0.800 |
| **12** | 0.503 | 0.587 | 0.678 | 0.727 | 0.769 |
| **13** | 0.484 | 0.560 | 0.648 | 0.703 | 0.747 |
| **14** | 0.464 | 0.538 | 0.626 | 0.679 | 0.723 |
| **15** | 0.446 | 0.521 | 0.604 | 0.654 | 0.700 |
| **16** | 0.429 | 0.503 | 0.582 | 0.635 | 0.679 |
| **17** | 0.414 | 0.485 | 0.566 | 0.615 | 0.662 |
| **18** | 0.401 | 0.472 | 0.550 | 0.600 | 0.643 |
| **19** | 0.391 | 0.460 | 0.535 | 0.584 | 0.628 |
| **20** | 0.380 | 0.447 | 0.520 | 0.570 | 0.612 |
| **21** | 0.370 | 0.435 | 0.508 | 0.556 | 0.599 |
| **22** | 0.361 | 0.425 | 0.496 | 0.544 | 0.586 |
| **23** | 0.353 | 0.415 | 0.486 | 0.532 | 0.573 |
| **24** | 0.344 | 0.406 | 0.476 | 0.521 | 0.562 |
| **25** | 0.337 | 0.398 | 0.466 | 0.511 | 0.551 |
| **26** | 0.331 | 0.390 | 0.457 | 0.501 | 0.541 |
| **27** | 0.324 | 0.382 | 0.448 | 0.491 | 0.531 |
| **28** | 0.317 | 0.375 | 0.440 | 0.483 | 0.522 |
| **29** | 0.312 | 0.368 | 0.433 | 0.475 | 0.513 |
| **30** | 0.306 | 0.362 | 0.425 | 0.467 | 0.504 |

**Rule:** If the ***calculated value*** is ***equal to*** or ***greater than*** the ***critical value*** then the results are ***significant at the 5% level of probability***.

d) Write a **statement of significance** (delete some parts);

"The calculated value of Rho is **0.851**. This is **more than** the critical value of **0.446** (when p<0.05). Therefore this result **is significant.** There is a **less than** 5% probability that the results are due to chance."

e) Does this mean the study was significant? Explain in terms of the context in the question.

**The results tell us that there is a strong, positive correlation between self-ratings of aggression and the number of siblings a person has. As one variable increases, so does the other.**