



## Probability

As part of the inferential statistics section in the psychology course, students are required to understand **the role of probability**, when assessing if the results of a study occurred by chance, or not.

Probability is the likelihood of something happening. When probability is referred to in the exam, it is expressed as a decimal between zero (0) and one (1).

- 0 means that something is impossible
- 1 means that something is certain

If you flip a coin, what is the probability it will land on heads? It is a 50% chance, so the probability is 0.5.

In psychology when testing the probability of data, researchers are looking to see if the difference between the two data sets is due to random chance. The accepted cut-off points in deciding if a result is significant or not is 5%. This means a 0.05 probability the results happened by chance. In psychology the probability value is referred to as the p-value.

## Understanding the p-value

The p-value is based on whether the effect observed is due to chance or not.

If the p-value is small (less than 5%) it means that the results are unlikely due to chance and so are statistically significant. In this case the null hypothesis is rejected. The null hypothesis is that there will be no difference or effect.

## Activity 1: Signs

$>$	Greater Than
$\geq$	Greater Than or Equal To
$<$	Less Than
$\leq$	Less Than or Equal To

What does  $p \leq 0.05$  mean?

What does  $p > 0.05$  mean?

The 0.05 probability level is used to avoid making errors in rejecting/accepting the wrong hypothesis. There are times when a psychologist will use a 0.01 probability level ( $p \leq 0.01$ ). For example, during a medical trial where the researcher must be very confident results were not due to chance. However, a new area in psychology might lead a researcher to using a lenient probability level, such as 0.10 ( $p \leq 0.10$ ).



- A type 1 error (false positive) occurs when researchers incorrectly accept the null hypothesis when using a probability value that is too strict,  $p=0.01$ .
- A type 2 error (false negative) occurs when researchers incorrectly reject the null hypothesis when using a probability value that is too lenient,  $p=0.10$ .

## **Activity 2 – answer the following questions**

- Why do psychologists use the 0.05 level of probability?
- Why would a psychologist use the 0.01 level of probability?
- Why would a psychologist use the 0.10 level of probability?

Once a statistical test has been conducted and a calculated value has been obtained, this is then compared to a critical value.

This is an example of a critical value table for the Wilcoxon test (one of the statistical tests that could appear in an exam).

### **Critical values for the Wilcoxon Signed Ranks test**

<i>n</i>	Level of significance for a one-tailed test		
	0.05	0.025	0.01
	Level of significance for a two-tailed test		
	0.1	0.05	0.02
N=5	0	–	–
6	2	0	–
7	3	2	0
8	5	3	1
9	8	5	3
10	11	8	5
11	13	10	7
12	17	13	9

**The calculated value must be equal to or less than the critical value in this table for significance to be shown.**

How they work together:

1. Obtain a calculated value
2. Find the critical value
3. Then decide if the result is significant, which will depend on the statistical test.
  - If  $p \leq 0.05$ , the result is significant = reject the null hypothesis.
  - If  $p > 0.05$ , the result is not significant = accept the null hypothesis.



**Scenario:** A psychologist wants to know if people's memory is better after listening to relaxing music compared to non-relaxing music.

## Condition 1

- 10 participants are given a newspaper article to read. They have 10-minutes to read it.
- The participants then listen to classical music and after the song is finished, they are asked 10-questions about the article.

## Condition 2

- The same 10 participants are then given a second article to read. They have 10-minutes to read it.
- The participants then listen to rock music and after the song is finished, they are asked 10-questions about the article.

The researcher compares the performance of each condition in terms of recall. The accuracy score is recorded (based on the 10 questions) and the difference between classical and rock music is calculated.

- The calculated value for the Wilcoxon rank is calculated as 1
- The critical value where  $N=8$ ,  $P=0.05$  using a one-tailed test is 5.

The result is statistically significant ( $p \leq 0.05$ ) so we can reject the null hypothesis. There is less than a 5% probability that the results are due to random chance, so the effect is considered statistically significant.

Since the calculated value, 1 is less than the critical value 5, it means that classical music having a better influence on memory is unlikely due to chance, therefore the psychologist is 95% confident that relaxing music helped memory, in terms of recalling the information.



## Exam questions

1. A psychologist carried out a Spearman's rank test. They wanted to see if their results were significant at  $p \leq 0.05$ . Define what is meant by  $p \leq 0.05$ . (2)
2. Two students decide to calculate a Wilcoxon Signed Ranks test of difference on the data gathered from their investigation. They used  $p \leq 0.01$  level of significance. Describe what is meant by  $p \leq 0.01$  in relation to the students' investigation. (2)

3. A researcher found an observed/calculated value of 5.84 when  $df=1$  when they calculated the chi-squared test for their investigation. Explain whether their results were significant for a one-tailed (directional) hypothesis when  $p \leq 0.01$ . (2)

Critical values for chi-squared distribution

	Level of significance for a one-tailed test					
	0.10	0.05	0.025	0.01	0.005	0.0005
df	Level of significance for a two-tailed test					
	0.20	0.10	0.05	0.025	0.01	0.001
1	1.64	2.71	3.84	5.02	6.64	10.83
2	3.22	4.61	5.99	7.38	9.21	13.82
3	4.64	6.25	7.82	9.35	11.35	16.27
4	5.99	7.78	9.49	11.14	13.28	18.47
5	7.29	9.24	11.07	12.83	15.09	20.52
6	8.56	10.65	12.59	14.45	16.81	22.46
7	9.80	12.02	14.07	16.01	18.48	24.32
8	11.03	13.36	15.51	17.54	20.09	26.12
9	12.24	14.68	16.92	19.02	21.67	27.88
10	13.44	15.99	18.31	20.48	23.21	29.59

The calculated value must be equal to or exceed the critical value in this table for significance to be shown.

4. A Wilcoxon Signed Rank test was used to determine whether there was a significant difference between the scores in the two conditions. The calculated (T) value was 7 ( $T=7$ ) for a one-tailed test at  $p=0.05$  with  $N=10$ . Determine whether this result is significant or not. (1)

Critical values for the Wilcoxon Signed Ranks test

n	Level of significance for a one-tailed test		
	0.05	0.025	0.01
N=5	Level of significance for a two-tailed test		
	0.1	0.05	0.02
5	0	–	–
6	2	0	–
7	3	2	0
8	5	3	1
9	8	5	3
10	11	8	5
11	13	10	7
12	17	13	9

The calculated value must be equal to or less than the critical value in this table for significance to be shown.

5. Explain what is meant by  $p \geq 0.05$  in relation to this study. (2)
6. A medical researcher decides to use a  $p \leq 0.01$  level of significance on their results. State what is meant by ' $\leq$ '. (1)



## Mark scheme

1. AO1 (2 marks) Credit one mark for a partial definition. Credit two marks for a full definition. For example: The probability the results are due to chance is 0.05 (1). The probability the results are due to chance is equal to or less than five percent (2).  
*Credit other reasonable marking points.*
2. AO2 (2 marks) Credit two marks for an accurate description of  $p \leq 0.01$  in relation to the scenario.  
For example: The probability that the results of the students' memory experiment were 99% or more of the time due to the one-hour revision session (1) influencing the essay score awarded and 1% due to chance factors like access to additional reading materials (1).  
*Credit other reasonable marking points.*  
*Generic answers score 0 marks*
3. AO2 (1 mark), AO3 (1 mark) One mark for identifying the correct critical value (AO2) One mark for accurate judgement of difference (AO3)  
For example: • There is a significant difference in quality of family relationships whilst waiting for a diagnosis compared to having a diagnosis (1) as the observed value (5.84) for a one-tailed test when  $df=1$  is greater than the critical value (5.02) (1).  
*Answers must relate to the scenario.*  
*Credit other reasonable marking point*
4. AO2 (1 mark) One mark for correct determination of significance
5. The result is significant (calculated value of 7 is less than critical value of 11)  
Reject other answers.
6. AO1 (1 mark) AO2 (1 mark) One mark for accurate understanding of  $P \geq 0.05$  (AO1) One mark for explanation in relation to stimulus (AO2)  
For example, The probability that the results are due to chance is less than or equal to 5% (1). There is a less than 5% likelihood that the number of words recalled was due to chance rather than the environment (1).  
*Credit other reasonable marking points*
7. AO1 (1 mark) Credit one mark for an accurate statement.  
For example: Less than or equal to (1)  
*Credit other reasonable marking points.*