



INFERENTIAL STATISTICS - CHI-SQUARED TEST

The Chi-squared test and when to use each statistical test

Learning Approach

The Learning Approach requires that you do a practical (and also that you understand for the exam) using the **Chi-squared test**. This document will take you through the Chi-squared test in a step-by-step process, but first, we must look at *when* to use each test.

There are a number of factors which decide on whether a Spearman's Rank correlation coefficient, Mann-Whitney U or Chi-squared test should be used for a given set of data. Factors which must be taken into consideration are:

- the experimental design
- whether the study is looking to find a difference between the behaviour of groups (an experiment) or for a relationship between two variables (a correlation)
- the level of measurement

Experimental design

This section will briefly remind you of the experimental designs available, but for more detailed information, including an evaluation of each design, see **M5 Experimental Design**.

An **independent groups** design uses different participants for each condition in the study. A **repeated measures** design uses the same group of participants for each of the conditions. A **matched pairs** design uses different participants for each condition, but they are matched into pairs (or equivalent) based on important participant characteristics.

Each of these three experimental designs look to find *differences* between two variables, usually two groups of people in an experiment. A fourth design, although not an experimental design, is the **correlation design** which is used to try and find a *relationship* (correlation) between two variables of equal importance.

Levels of measurement

In AS Psychology, there are only three levels of measurement:

- **nominal data** is the lowest level of measurement
data sorted into categories or sets, there are no scores or numerical values, for example, responding "yes" or "no" to a questionnaire
- **ordinal data** is the middle level of measurement
ranked data – i.e. that that can be put into a hierarchy, e.g. rank out of 10
- **interval data** is the highest level of measurement
*also called **ratio data**, although not exactly the same, they are considered to be the same for the purpose of your course, these are mathematical data – so statistical figures which have numerical values*

Choosing the correct statistical test

These factors must all be taken into consideration. Before you can choose which test is to be used, you *must* know the experimental design (or if it is a correlation design) and the level of measurement. The table below explains when you would use each test.

	Spearman's Rank	Mann Whitney U	Chi-squared
Experimental design:	correlation design	independent groups	independent groups
Testing for:	a relationship	a difference	a difference
Level of measurement:	ordinal or interval	ordinal or interval	nominal

Degrees of freedom

The only other component of statistical tests that you need to understand are *N* and *df*. You will already know by now that *N* is just the number of participants (or scores). In the Mann-Whitney U test there are two different *N* values.

[youtube.com/c/MegaLecture/](https://www.youtube.com/c/MegaLecture/)

+92 336 7801123
aspsychology101.wordpress.com

The letters *df* stand for **degrees of freedom**. Whilst the Mann-Whitney U test and the Spearman's Rank test use *N*, the Chi-squared test of the Learning Approach does not, it uses *df*. This refers to the number of values in the final calculation that are free to vary. This is found by using the formula:

$$df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

for a table of data the Chi-squared test is being used for. You will come across degrees of freedom in a while.

Chi-squared test

We know by now that the Chi-squared test is chosen when the study uses an independent groups design, tests for a difference and has nominal (i.e. in categories) data. The experiment will use more than one condition on separate groups of participants. The example we will use for our Chi-squared test is people making donations. People entering a cathedral might be observed to see if they are more likely to make a donation if they believe someone is watching them. So the independent variable (IV) is whether the participants enter being observed or think they are unobserved and the dependent variable (DV) is whether or not they make a donation.

A Chi-squared test uses a contingency table called a **two-by-two table** which pairs the various conditions. It is common for a tally system to be used during the experiment. The two-by-two table for our study may look like the one below:

	observed	thought unobserved
donated		
not donated		

Each person would fall into one of the four conditions: they either donate knowing they are observed, donate unaware of being observed, don't donate knowing they're observed or don't donate not knowing they're being observed. The categories must be **mutually exclusive**, meaning they cannot fall into more than one of the categories.

The following instructions outline the Chi-squared test. The test is designed to see if the combinations of categories occur more frequently than would be expected by chance. It looks for an association between each combination of categories and predicts the value that would be expected for each cell in the contingency table. It then calculates how different the observed values were from the expected values, providing a probability for this occurring randomly.

Step 1: Put the raw data into a contingency table

	observed	thought unobserved	TOTAL
donated	50	30	80
not donated	20	60	80
TOTAL	70	90	160

Step 2: Transfer the totals for each cell into a table

Category	Observed value (O)
observed, donated	50
unobserved, donated	30
observed, not donated	20
unobserved, not donated	60

Step 3: Calculate the **expected value** of each cell from the original contingency table

The expected value is calculated by multiplying the row total by the column total of the cell in question, and then dividing that number by the grand total. For example, the expected value for “thought unobserved” and “not donated” would be the row total (80) multiplied by the column total (90), then that figure divided by the grand total (160) which would give us an answer of 45

Step 4: Minus the expected value from each observed value

Step 5: Square the figure for each answer of Step 4

Step 6: Divide the answers from Step 5 by the expected value

Category	Observed value (O)	Expected value (E)	O - E	(O - E) ²	$\frac{(O - E)^2}{E}$
observed, donated	50	$(80 \times 70) / 160 = 35$	15	225	6.43
observed, not donated	30	$(80 \times 90) / 160 = 45$	-15	225	5
unobserved, donated	20	$(80 \times 90) / 160 = 45$	-15	225	6.43
unobserved, not donated	60	$(80 \times 70) / 160 = 35$	15	225	5

Step 7: Use the formula for Chi-squared to calculate the answer of the Chi-squared test

The formula for Chi-squared is as below:

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

This just means totalling all of the values in the last column of our table above. So in our example, $\chi^2 = 22.86$

Step 8: Calculate the degree of freedom for your study

This is the number of rows - 1 multiplied by the number of columns - 1, so $(2 - 1) \times (2 - 1) = 1$

With the answer to the Chi-squared test and the degree of freedom calculated, we can then use the statistical tables to assess whether or not our findings are statistically significant. An example of one is shown below.

		Level of significance for one-tailed test			
		0.05	0.025	0.01	0.005
df =		Level of significance for two-tailed test			
		0.10	0.05	0.025	0.01
1	2.71	3.84	5.41	6.64	
2	4.60	5.99	7.82	9.21	

The Chi-squared test can be used to test either one-tailed or two-tailed tests, as the table shows.

The critical value must be equal to or less than our Chi-squared value (i.e. the result of the Chi-squared test must be the same as or more than the figure seen in the table).

Our hypothesis was two-tailed, as we did not specify which way the outcome would go, we merely hypothesised that the idea of observation would affect the donations being given. So we use the level of significance for a two-tailed test.

If we wanted to test to see if our findings are statistically significant for $p \leq 0.01$, we could do so. The critical value for a two-tailed test at $p \leq 0.01$ is 6.64 (remember the degree of freedom is 1). This has been highlighted in the table above. Our value was 22.86, which is certainly higher than 6.64, and so we can reject the null hypothesis for $p \leq 0.01$ and accept the alternative hypothesis.