Chi Square Analysis

When do we use chi square?

More often than not in psychological research, we find ourselves collecting scores from participants. These data are usually *continuous measures*, and might be scores on a questionnaire or psychological scale, reaction time data or memory scores, for example. And when we have this kind of data, we will usually use it to look for mean differences on scores between or within groups (e.g. using t-tests or ANOVAs), or perhaps to look for relationships between different types of scores that we have collected (e.g. correlation, regression).

However sometimes we do not have this kind of data. Sometimes data will be a lot simpler than this, instead consisting only of frequency data. In these cases participants do not contribute scores for analysis; instead they each contribute to a "head count" within different grouping categories. This kind of data is known as *categorical data*, examples of which could be gender (male or female) or university degree classifications (1, 2:1, 2:2, 3, pass or fail) – or any other variable where each participant falls into one category. When the data we want to analyse is like this, a chi-square test, denoted χ^2, is usually the appropriate test to use.

What does a chi-square test do?

Chi-square is used to test hypotheses about the distribution of observations in different categories. The null hypothesis (H_o) is that the observed frequencies are the same as the expected frequencies (except for chance variation). If the observed and expected frequencies are the same, then $\chi^2 = 0$. If the frequencies you observe are different from expected frequencies, the value of χ^2 goes up. The larger the value of χ^2, the more likely it is that the distributions are significantly different.

…but what does this mean in English?

To try and explain this a little better, let's think about a concrete example. Imagine that you were interested in the relationship between road traffic accidents and the age of the driver. We could randomly obtain records of 60 accidents from police archives, and see how many of the drivers fell into each of the following age-categories: 17-20, 21-30, 31-40, 41-50, 51-60 and over 60. If there is no relationship between accident-rate and age, then the drivers should be equally spread across the different age-bands (i.e. there should be similar numbers of drivers in each category). This would be the *null hypothesis*. However, if younger drivers are more likely to have accidents, then there would be a large number of accidents in the younger age-categories and a low number of accidents in the older age-categories.

So… say we actually collected this data, and found that out of 60 accidents, there were 25 individuals aged 17-20, 15 drivers aged 21-30 and 5 cases in each of the other age groups. This data would now make up our set of ***observed*** frequencies.

We might now ask: are these observed frequencies similar to what we might expect to find by chance, or is there some non-random pattern to them? In this particular case, from just looking at the frequencies it seems fairly obvious that a larger proportion of the accidents involved younger drivers. However, the question of whether this distribution could have just occurred by chance is yet to be answered. The Chi-Square test helps us to decide this by comparing our observed frequencies to the frequencies that we might ***expect*** to obtain purely by chance.

It is important to note at this point, that that Chi square is a very versatile statistic that crops up in lots of different circumstances. However, for the purposes of this handout we will only concentrate on two applications of it:

- **Chi-Square "Goodness of Fit" test:** This is used when you have categorical data for **one** independent variable, and you want to see whether the distribution of your data is similar or different to that expected (i.e. you want to compare the *observed* distribution of the categories to a theoretical *expected* distribution).

- **Chi-Square Test of Association between two variables:** This is appropriate to use when you have categorical data for **two** independent variables, and you want to see if there is an association between them.

Chi-Square "Goodness of Fit" test

This is used when you have **one independent variable**, and you want to compare an observed frequency-distribution to a theoretical expected frequency-distribution.

For the example described above, there is a single independent variable (in this example "**age group**") with a number of different levels (17-20, 21-30, 31-40, 41-50, 51-60 and over 60). The statistical question is: do the frequencies you actually observe differ from the expected frequencies by more than chance alone?

In this case, we want to know whether or not our observed frequencies of traffic accidents occur equally frequently for the different ages groups (so that our theoretical frequency-distribution contains the same number of individuals in each of the age bands).

The way in which we would collate this data would be to use a *contingency table*, containing both the observed and expected frequency information.

	Age band						
	17-20	21-30	31-40	41-50	51-60	over 60	*Total:*
Observed frequency of accidents	25	15	5	5	5	5	*60*
Expected frequency of accidents	10	10	10	10	10	10	*60*

To work out whether these two distributions are significantly different from one another, we use the following Chi-square formula:

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

This translates into:

χ^2 = sum of (i.e., across categories) (observed frequency – expected frequency)^2 (divided by) expected frequency

This may look complicated, but really it just means that you have to follow four simple steps, which are described on the next page.

Step One
Take each observed frequency and subtract from it its associated expected frequency (i.e., work out (O-E)):

25-10 = **15** 15-10 = **5** 5-10 = **-5** 5-10 = **-5** 5-10 = **-5** 5-10 = **-5**

Step Two
Square each value obtained in step 1 (i.e., work out $(O\text{-}E)^2$):

225 **25** **25** **25** **25** **25**

Step Three
Divide each of the values obtained in step 2, by its associated expected frequency (i.e., work out $\frac{(O\text{-}E)^2}{E}$):

$\frac{225}{10}$ = **22.5** $\frac{25}{10}$ = **2.5** $\frac{25}{10}$ = **2.5** $\frac{25}{10}$ = **2.5** $\frac{25}{10}$ = **2.5** $\frac{25}{10}$ = **2.5**

Step Four
Add together all of the values obtained in step 3, to get your value of Chi-Square:

χ^2 = 22.5 + 2.5 + 2.5 + 2.5 + 2.5 + 2.5 = **35**

Assessing the size of our obtained Chi-Square value:

What you do, in a nutshell...

(a) Work out how many "degrees of freedom" (d.f.) you have.
(b) Decide on a probability level.
(c) Find a table of "critical Chi-Square values" (in most statistics textbooks).
(d) Establish the critical Chi-Square value for *this* particular test, and compare to your obtained value.

If your obtained Chi-Square value is **bigger** than the one in the table, then you conclude that **your obtained Chi-Square value is too large to have arisen by chance**; it is more likely to stem from the fact that there were real differences between the observed and expected frequencies. In other words, contrary to our null hypothesis, the categories did *not* occur with similar frequencies.

If, on the other hand, your obtained Chi-Square value is **smaller** than the one in the table, you conclude that there is no reason to think that the observed pattern of frequencies is not **due simply to chance** (i.e., we retain our initial assumption that the discrepancies between the observed and expected frequencies are due merely to random sampling variation, and hence we have no reason to believe that the categories did not occur with equal frequency).

For our worked example...

(a) First we work out our degrees of freedom. For the Goodness of Fit test, this is simply the number of categories minus one. As we have six categories, there are 6-1 = **5** degrees of freedom.

(b) Next we establish the probability level. In psychology, we use **p < 0.05** as standard – and this is represented by the **5% column**.

(c) We now need to consult a table of "critical values of Chi-Square". Here's an excerpt from a typical table:

Degrees of Freedom	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	12.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

Source: Adapted from p.112 of Sir R.A. Fisher, *Statistical Methods fro Research Workers* (Edinburgh: Oliver and Boyd, 1958).

(d) The values in each column are "critical" values of Chi-Square. These values would be expected to occur by chance with the probability shown at the top of the column. The relevant value for *this* test is found at the intersection of the appropriate d.f. row and probability column. As our obtained Chi-Square has 5 d.f., we are interested in the values in the **5 d.f.** row. As the probability level is **p <.05**, we then need to look in the **5% column** (as .05 represents a chance level of 5 in 100… or 5%) to find the critical value for this statistical test. In this case, the critical value is **11.07**.

Finally, we need to compare our *obtained* Chi-Square to the critical value. If the obtained Chi-Square is larger than a value in the table, it implies that it is unlikely to have occurred by chance. Our obtained value of **35** is much larger than the critical value of 11.07. We can therefore be relatively confident in concluding that our *observed* frequencies are significantly different from the frequencies that we would *expect* to obtain if all categories were equally distributed. In other words, age *is* related to the amount of road traffic accidents that occur.

<u>Chi-Square Test of Association between two variables</u>

The second type of chi square test we will look at is the Pearson's chi-square test of association. You use this test when you have categorical data for **two** independent variables, and you want to see if there is an association between them.

For this example, let's stick with the theme of driving, but this time consider gender performance on driving tests. This time we have two categorical variables: Gender (two levels: Male vs Female) and Driving Test Outcome (two levels: Pass vs Fail).

In this case, the statistical question we want to know the answer to is whether driving test outcome is related to the gender of the person taking the test. Or in other words, we want to know if males show a different pattern of pass/fail rates than females.

To answer this question, we would start off by putting out data into a *contingency table*, this time containing only the *observed* frequency information. We can then use this table to calculate *expected* frequencies.

In this case, imagine the pattern of driving test outcomes looked like this:

	Male	**Female**
Pass	7	11
Fail	13	9

In this example, simply looking at the observed frequencies gives us an idea that the pattern of driving test outcomes may be different for the genders. It seems females have a more successful pass/fail rate than males. However, to test whether this observed difference is significant, we need to look at the outcome of a Chi-Square test. As with the one-variable Chi-Square test, our aim is to see if the pattern of observed frequencies is significantly different from the pattern of frequencies which we would expect to see by chance - i.e., what we would expect to obtain if there was no relationship between the two variables in question. With respect to the example above, "no relationship" would mean that the pattern of driving test performance for males was no different to that for females.

The Chi-Square formula is exactly the same as for the one-variable test described earlier; the only difference is in how you calculate the expected frequencies.

Step 1: Add numbers across columns and rows. Calculate total number in chart.

	Male	**Female**	
Pass	7	12	**= 19**
Fail	13	8	**= 21**
	= 20	**= 20**	<u>**= 40**</u>

Step 2: Calculate expected numbers for each individual cell (i.e. the frequencies we would expect to obtain if there were no association between the two variables). You do this by multiplying row sum by column sum and dividing by total number.

$$\textbf{Expected Frequency} = \frac{\textbf{Row Total x Column Total}}{\textbf{Grand Total}}$$

For example: using the first cell in table (Male/Pass);

$$\frac{19 \times 20}{40} = 9.5$$

…and the cell below (Male/Fail):

$$\frac{21 \times 20}{40} = 10.5$$

Do this for each cell in the table above.

Step 3: Now you should have an observed number and expected number for each cell. The observed number is the number already in 1st chart. The expected number is the number found in the last step (step 2). Redo the contingency table, this time adding in the expected frequencies in brackets below the obtained frequencies:

	Male	Female	
Pass	7 *(9.5)*	12 *(9.5)*	total=19
Fail	13 *(10.5)*	8 *(10.5)*	total=21
	total=20	total=20	grand total =40

Step 4: Now calculate Chi Square using the same formula as before:

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

...or...

$$\chi^2 = \text{Sum of } \frac{(\text{Observed - Expected})^2}{\text{Expected}}$$

Calculate this formula for each cell, one at a time. For example, cell #1 (Male/Pass):

Observed number is: **7**
Expected number is: **9.5**

Plugging this into the formula, you have: $\frac{(7 - 9.5)^2}{9.5} = 0.6579$

Continue doing this for the rest of the cells.

Step 5: Add together all the final numbers for each cell, obtained in Step 4. **There are 4 total cells, so at the end you should be adding four numbers together for you final Chi Square number.**

In this case, you should have:

0.6579 + 0.6579 + 0.5952 + 0.5952 = 2.5062

So, **0.095** is our obtained value of Chi-Square; it is a single-number summary of the discrepancy between our obtained frequencies, and the frequencies which we would expect if there was no association between our two variables. *The bigger this number, the greater the difference between the observed and expected frequencies.*

Step 6: Calculate degrees of freedom (*df*):

(Number of Rows – 1) x (Number of Columns – 1)
(2 – 1) x (2 – 1)
1 x 1
=1 *df* (degrees of freedom)

Assessing the size of our obtained Chi-Square value:

The procedure here is the same as for the Goodness of Fit test. We just need:

(a) Our "degrees of freedom" (d.f.) ✓
(b) A suitable probability level (p=0.05 in psychology) ✓
(c) A table of "critical Chi-Square values" ✓
(d) Establish the critical Chi-Square value for this test (at the intersection of the appropriate d.f. row and probability column), and compare to your obtained value.

As before, if the obtained Chi-Square value is **bigger** than the one in the table, then you conclude that **your obtained Chi-Square value is too large to have arisen by chance**. This would mean that the two variables are likely to be related in some way. **NB** - the Chi-Square test merely tells you that there is *some relationship* between the two variables in question: it does not tell you what that relationship is, and most importantly, it does not tell you anything about the causal relationship between the two variables.

If, on the other hand, your obtained Chi-Square value is **smaller** than the one in the table, you cannot reject the null hypothesis. In other words, you would conclude that your variables are unlikely to be associated.

For this example: Using the chart below, at the p = 0.05 significance level, with 1 *df*, the critical value can be established as **3.84**. Therefore, in order to reject the null hypothesis, the final answer to the Chi Square must be **greater or equal to 3.84**.

Degrees of Freedom	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	12.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

Source: Adapted from p.112 of Sir R.A. Fisher, *Statistical Methods fro Research Workers* (Edinburgh: Oliver and Boyd, 1958).

The Chi Square calculation above was **2.5062.** This number is ***less*** than the critical value of **3.84**, so in this case the null hypothesis cannot be rejected. In other words, there does not appear to be a significant association between the two variables: males and females have a statistically similar pattern of pass/fail rates on their driving tests.

Assumptions of the Chi-Square test

For the results of a Chi-Square test to be reliable, the following assumptions must hold true:

1. Your data are a random sample from the population about which inferences are to be made.
2. Observations must be independent: each subject must give one and only one data point (i.e., they must contribute to one and only one category).
3. Problems arise when the expected frequencies are very small. As a rule of thumb, Chi- Square should not be used if more than 20% of the expected frequencies have a value of less than 5 (it does not matter what the observed frequencies are). You can get around this problem in two ways: either combine some categories (if this is meaningful, in your experiment), or obtain more data (make the sample size bigger).