

The χ^2 Test

Contingency Tables: A contingency table is a sort of Venn Diagram which shows how a population splits according to two factors. We would like to see whether or not the two factors are independent of each other. The procedure is to take a table of observed frequencies and, from this, draw up a table of *expected* frequencies on the assumption of independence. For example, suppose we take a group of 60 people, and we are interested to find out if those who are over 21 are more likely to pass their driving test than those who are under 21. The observed frequencies are shown in the table below:

	Over 21	Under 21	Total
Pass	36	12	48
Fail	4	8	12
Total	40	20	60

Working out the expected frequencies: We can see that one-fifth of the group failed the test: we would therefore expect 8 out of the 40 over-21s to fail. 8 is the *expected* frequency, and the calculation to work it out is $\frac{40 \times 12}{60}$; in general, the calculation is $\frac{\text{row total} \times \text{column total}}{\text{grand total}}$. (Note that the numbers do not necessarily come out as integers). The table of expected frequencies is calculated as:

	Over 21	Under 21	Total
Pass	32	16	48
Fail	8	4	12
Total	40	20	60

The newly calculated figures are shown in bold


We can see that the number of over 21's who passed the test is more than we expected – could it be this high by chance alone, or could it be that the two factors (age and pass rate) are in fact related?

χ^2 test: We now need a statistic which measures by how much the observed frequencies differ from the expected frequencies. This is called the χ^2 statistic and is calculated as $\chi^2 = \sum \frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$. So, taking each pair of corresponding values, we have:

$$\chi^2 = \frac{(36 - 32)^2}{32} + \frac{(12 - 16)^2}{16} + \frac{(4 - 8)^2}{8} + \frac{(8 - 4)^2}{4} = 7.5$$

On its own, this doesn't tell us much. The next move is to go to a table of *critical values* which tells us at which point "could have happened by chance" becomes "unlikely to have happened by chance." The table of critical values starts like this:

Degrees of freedom	0.9	0.95	0.975	0.99
1	2.705	3.841	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

 Find out how to enter the observed frequencies into your GDC. It can then calculate the expected frequencies and the value of χ^2 .