Chi-square tests

(Chapter 11 of the text)

A motivating example

A survey of attitudes to premarital sex among different religious groups asked 1579 people the question "When is premarital sex wrong?" They were asked to respond "Always or almost always", or "Sometimes", or "Never". They were also asked whether they were Protestant, Catholic or Jewish. (Those who were none of these are not included.) The results were as follows:

	Always or almost always	Sometimes	Never	Total
Protestant	472	227	384	1083
Catholic	99	120	226	445
Jewish	3	14	34	51
Total	574	361	644	1579

A natural question to ask is whether this proves that people of different religions have different views about premarital sex.

As in any hypothesis testing problem, we start by formulating a null hypothesis that says there is no difference. Then, if that hypothesis is rejected with a sufficiently small P-value, we will conclude that there really is a difference.

In this case, the simplest way to think about this is to look at the overall proportions of people giving each response ("always", "sometimes" or "never"). From the column totals in the previous table, we derive the answers as $\frac{574}{1579}$=0.3635, $\frac{361}{1579}$=0.2286 and $\frac{644}{1579}$=0.4079.

Next, we calculate the "expected number" of responses we would see in each cell of the table if the null hypothesis was correct. For example, among the 1083 Protestants, if they were distributed in the proportions .3635, .2286, .4079, then the actual numbers would be 393.7, 247.6, 441.8. It doesn't matter that these are not whole numbers, because at this stage, it is a theoretical calculation that leads to the following full table of expected values:

	Always or almost always	Sometimes	Never	Total
Protestant	393.7	247.6	441.8	1083
Catholic	161.8	101.7	181.5	445
Jewish	18.5	11.7	20.8	51
Total	574	361	644	1579

We now work out

$$
\begin{aligned}
\chi^2 &= \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \\
&= \frac{(472-393.7)^2}{393.7} + \frac{(227-247.6)^2}{247.6} + \frac{(384-441.8)^2}{441.8} \\
&\quad + \frac{(99-161.8)^2}{161.8} + \frac{(120-101.7)^2}{101.7} + \frac{(226-181.5)^2}{181.5} \\
&\quad + \frac{(3-18.5)^2}{18.5} + \frac{(14-11.7)^2}{11.7} + \frac{(34-20.8)^2}{20.8} \\
&= 15.572 + 1.714 + 7.562 + 24.375 + 3.293 \\
&\quad + 10.910 + 12.986 + 0.452 + 8.377 \\
&= 85.241.
\end{aligned}
$$

The symbol χ^2 (pronounced "chi-squared") is used for the result of this sum, because when H_0 is true, it has a distribution known as the χ^2 distribution.

In order to use the χ^2 distribution, we need to compute the *degrees of freedom*. The rule here is: for a table with r rows and c columns,

$$df = (r-1)(c-1).$$

So in this case, $r = c = 3$ and $df = 2 \times 2 = 4$.

The next step is to look in a table of the χ^2 distribution — in our text, Table C or page A4 of Appendix A. In this case, looking to the row with $df = 4$ we see values 5.39, 7.78, 9.49, 11.14, 13.28, 14.86, 18.47 corresponding to right-tail probabilities 0.25, 0.1, 0.05, 0.025, 0.01, 0.005, 0.001. The value 85.241 lies well beyond the range of this table — in other words, the P value is much smaller than 0.001, indicating a very highly significant effect. In other words, there is a definite association between religious affiliation and how people respond to a questionnaire about premarital sex — it wasn't just a chance association.

df	Right-tail Probability 0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.31
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.43	104.21	112.32
80	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	98.65	107.57	113.15	118.14	124.12	128.30	137.21
100	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Footnote. Could make exact calculation in Excel, using CHIDIST. P-value is 1.35×10^{-17}.

General principles behind the χ^2 test

Either a *test of independence* or a *test of homogeneity*.

Don't worry about the distinction between these two.

Steps in the test:

1. Determine the expected values when the null hypothesis of independence is correct. Typically

$$\text{EXPECTED VALUE} = \frac{\text{ROW SUM} \times \text{COLUMN SUM}}{\text{TOTAL SUM}}$$

2. Calculate $\chi^2 = \sum \frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}}$

3. Calculate $df = (r-1)(c-1)$ where r and c are the number of rows and columns in the table.

4. Use Table C to determine the P value for the χ^2 value closest to the value calculated in step 2. Note that in this situation, we always use the one-sided (right-)tail probability.

5. Interpret the result — the smaller the value of P, the more significant the result.

However for this result to be valid, one of the requirements is that the expected number in each cell should be at least 5 (see page 496 of the text). We'll see an example later where this requirement makes a difference.

The χ^2 test for the skin lesions in marathon runners

Category	Referred for treatment	Not referred for treatment	Total
Non-runners	14	196	210
Light trainers	5	73	78
Moderate trainers	13	88	101
Heavy trainers	6	25	31
Total	38	382	420

Under the null hypothesis that there is no association between running status and the need for treatment, the probability that a given individual is referred for treatment is still $\frac{38}{420} = 0.09048 = 1 - 0.90952$. This leads to the following table of expected values:

Category	Referred for treatment	Not referred for treatment	Total
Non-runners	19	191	210
Light trainers	7.075	70.943	78
Moderate trainers	9.138	91.862	101
Heavy trainers	2.805	28.195	31
Total	38	382	420

Table of (Observed $-$ Expected)^2/Expected:

Category	Referred for treatment	Not referred for treatment
Non-runners	1.316	0.131
Light trainers	0.600	0.060
Moderate trainers	1.632	0.162
Heavy trainers	3.640	0.362

The total χ^2 is 7.903, with $df = (4-1) \times (2-1) = 3$. Looking it up in the χ^2 table, we find that the value associated with a .05 right tail probability is 7.81. Since 7.903 is greater than this, the right-tail probability is smaller, in other words, just under 0.05. Because of that, we can say that the value is just significant.

Also, we can see that the biggest contribution to χ^2 is 3.640 corresponding to the "heavy trainers, referred for treatment" category. In other words, the evidence for a significant effect is almost entirely due to that group.

Two comments about this test:

1. The exact P-value according to the chi-squared distribution is 0.048 (function CHIDIST in Excel)

2. The condition that the expected values should all be at least 5 is violated by the "heavy trainers, referred for treatment" category (mean is 2.805). However, the alternative "Fisher's exact test" leads to a P-value 0.049 (instead of 0.048), so it makes no difference in practice.

Further comments and caveats

1. If we accept the null hypothesis, that does not mean the null hypothesis is necessarily correct. Often, it just signifies that we didn't have enough data to reject it.

2. Likewise, even if we reject the null hypothesis, that doesn't mean we have proved a causal connection. Other issues such as lurking variables may still be relevant. We have shown is that the association isn't just due to chance — there could still be other explanations.

3. In reality, we use these methods and there are still a lot of false positives (i.e. rejecting the null hypothesis when we shouldn't). Issue of *multiple testing* — if we apply these procedures many times, roughly 5% of cases when there is really no effect will be misidentified as having established an effect.

Student Party Affiliations

	2009	2010
Democrat	35	26
Independent	12	13
Republican	21	42

Expected values under null hypothesis:

	2009	2010
Democrat	27.8	33.2
Independent	11.4	13.6
Republican	28.8	34.2

Chi-squared statistic is 7.29 with df=2. Very close to 0.025 tail probability.

P-value about 0.025. Differences are statistically significant.

Sex Education Study

Treatment group	Had sex? Yes	Had sex? No
12-hour comprehensive	39	53
8-hour comprehensive	40	57
Safer Sex	44	41
Abstinence	31	64
Control	41	47

Chi-squared statistic is 7.42 with df=4. P-value about 0.115.

(Fisher exact test: 0.114.)

When all five treatment groups are compared, the differences are not statistically significant.

The British General Election
(Guardian/ICM Poll, Unweighted)

Date	Conservative	Labour	Liberal Democrat	Other
16/18 April	213	194	209	59
9/11 April	223	234	131	85
1/3 April	237	242	127	63

Table of $\frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}}$:

0.66	4.07	17.78	1.52
0.01	0.49	3.95	3.67
0.85	1.76	5.02	0.47

Chi-squared statistic is 40.3 with df=6. P-value about 0.0000004.

The trend towards the Liberal Democrats appears to be real.

THE END