Lecture 35

Categorical Responses: Chi-Square and Fisher's Exact tests

In this lecture, we will discuss two additional analyses of 2 by 2 contingency tables: The Chi-square test (which is like using ANOVA to test the equality of two population means) and Fisher's exact test, which works well for small sample sizes.

We will again use the following summary data of handedness in twins born between 1900 and 1910 in Denmark.

Sex	Right handed	Not right handed
Male	1174	119
female	1137	79

Chi-Square Tests

A Chi-Square test is like an ANOVA for categorical responses. It tests whether the rows are distributed equally over the columns and whether, symmetrically, the columns are distributed equally over the rows. That is the same as a test for independence between the row and column categories. The statistic is the sum of (observe - expected)$^{\mathbf{2}}$/expected, where the expected value is the row total times the column total over the total for the table. That is, the expected value for each cell is the row proportion times the column total or, symmetrically, the column proportion times the row total. If the contingency table has $\boldsymbol{r}$ row categories and $\boldsymbol{c}$ column categories, then the degrees of freedom for the resulting Chi-Square statistic is $\boldsymbol{(r-1)\times(c-1)}$. For our $\mathbf{2\times 2}$ tables, the degrees of freedom is just $\mathbf{1}$. The reason is that proportions add to 1, so if you know the first $\boldsymbol{c-1}$ column proportions, then you know the last and if you know the first $\boldsymbol{r-1}$ row proportions, then you know the last.

In our example, the calculations proceed as follows:

Observed Values

Sex	Right handed	Not right handed	Row totals
Male	1174	119	1293
female	1137	79	1216
Column totals	2311	198	2509

Expected Values

Sex	Right handed	Not right handed	Row totals
Male	$\mathbf{(1293\times 2311)/2509=1191}$	$\mathbf{(1293\times 198)/2509=102}$	1293
female	$\mathbf{(1216\times 2311)/2509=}$ 1120	$\mathbf{(1216\times 198)/2509=96}$	1216
Column totals	2311	198	2509

Chi-Square Values (Observed - Expected)^2/Expected

Sex	Right handed	Not right handed
Male	0.242	2.820
female	0.257	2.998

So the Chi-Square statistic is $\chi^2 = 6.3$ with a p-value of 0.012. This test is mathematically equivalent to our test of two proportions. Note that $\chi^2 = 6.3 = (2.51)^2 = Z^2$ which is entirely analogous with the F statistic being the square of the t-statistic in One-way ANOVA with two groups.

Other analogies with ANOVA hold as well. The Chi-Square test generalizes to more than a binary predictor. It also generalizes to more than a binary response. However, having a significant Chi-Square test does not tell you where any differences might lie. One way to find out it to look at the table holding the Chi-Square values and see where the largest contributors to the test statistic lie. In this case, they come from the non-right handed column - male twins are over-represented in this group and female twins are correspondingly under-represented.

Fisher's Exact Test

The last test we will discuss was developed by Fisher at a tea party (I kid you not). The situation at that party was a strange one that rarely happens in science - both marginal distributions were fixed in advance and known to everyone involved. The interpretation of Fisher's exact test to other testing situations is that it is a conditional test where what you are given are the marginal totals. You might wonder why one would want such a test and p-values since surely your own marginals are not (both) fixed in advanced. However, my recommendation is to use Fisher's test whenever you have small sample sizes: it is reasonably conservative, has been proven to work well, and Fisher is a reliable genius in the murky science that is statistics.

At any rate, here is the set-up for Fisher's development of his test. A woman at the tea party claimed she could tell whether the milk was poured into her tea first (before the tea) or second (after the tea was poured). Fisher devised a test on the spot using 8 tea cups - 4 with the milk poured first and 4 with the milk poured second. The lady and Fisher and all other parties knew this arrangement.

Here are the results:

	Reality		
Guess	Milk first	Milk second	
Milk first	3	1	4
Milk Second	1	3	4
	4	4	8

Fisher's exact test is based on the hypergeometric distribution, typically described as a ball and urn problem. An urn contains A black balls and B white balls. You draw out n without replacement. What is the probability you draw k black balls?

$$P(k) = \frac{\binom{A}{k}\binom{B}{n-k}}{\binom{A+B}{n}}$$

Now, we just have to figure out what all these balls and colors are. We can do it either way, but suppose the colors are reality - that is $\boldsymbol{A}$ is the 4 cups of tea with the milk really poured first and $\boldsymbol{B}$ is the 4 cups of tea with the milk really poured second. Then $\boldsymbol{n}$ is the 4 cups of tea the lady chose to guess that the milk is poured first (the other 4 have to then be her guess for the milk being poured second.) So what is the probability she could have guessed **3** or more correctly?

$$\boldsymbol{P}(\mathbf{3}) + \boldsymbol{P}(\mathbf{4}) = \frac{\binom{\mathbf{4}}{\mathbf{3}}\binom{\mathbf{4}}{\mathbf{1}}}{\binom{\mathbf{8}}{\mathbf{4}}} + \frac{\binom{\mathbf{4}}{\mathbf{4}}\binom{\mathbf{4}}{\mathbf{0}}}{\binom{\mathbf{8}}{\mathbf{4}}} = \frac{\mathbf{17}}{\mathbf{70}} = \mathbf{.2429}$$

Fisher concluded that the lady could not tell which was poured first. The lady, had she been statistically sophisticated, could have complained about the lack of power of the test give the very small sample size.

References and Readings

Exercises for Lecture 35

1. –
2. –