Lecture 36

Pooling Chi Square Tables and Paired Binary Data

In this lecture we will discuss the problem of combining data from separate populations into a single contingency table and discuss the Mantel Haenszel test for analyzing multiple 2× 2 tables together. We will also discuss the appropriate analog to the paired t-test for binary response data with binary predictors.

Simpson's Paradox

In the 1970's, it was observed that graduate programs at Berkeley, on the whole, admitted significantly fewer women than they did men. See the following table for the results for one time period. The difference is statistically significant.

	Applied	Accepted
Men	8442	44%
Women	4321	35%

This lead to the question of which departments were doing the discriminating. The following table is for six unnamed departments during this time period:

Department	Men		Women	
	Number Applicants	Percent Accepted	Number Applicants	Percent Accepted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

What happens is that most women choose to apply to departments with low admissions rates and many men choose to apply to departments with higher admission rates. There may well be discrimination going on, but it happens before Berkeley gets involved - or at least that was Berkeley's conclusion. See the Science article cited in the references and readings.

In order to analyze contingency tables from multiple populations, we can utilize the test described next:

The Mantel Haenszel Test

The Mantel Haenszel Test allows you to combine the results of several Chi-Square tables in one analysis and thus avoid the problem of pooling results from different populations. The following steps are used to conduct the test:

1. Choose one square for the analysis. The upper left hand square will work. Use this square in all your tables.
2. Calculate the Excess for this square. That is, calculated the observed minus the expected values for this square.
3. Calculate the variance as $\frac{R_1 R_2 C_1 C_2}{T^2(T-1)}$ where R_1 and R_2 are the row totals, C_1 and C_2 are the column totals, and T is the total total.
4. Sum the excesses. The standard error is the square root of the sum of the variances. Use this to form a standard normal test statistic.
5. If you want the pooled odds ratio, find $\frac{\sum \frac{n_{1,1}n_{2,2}}{T}}{\sum \frac{n_{1,2}n_{2,1}}{T}}$

The set up is that you will have multiple tables like the following:

	Column 1	Column 2	
Row 1	$n_{1,1}$	$n_{1,2}$	R_1
Row 2	$n_{2,1}$	$n_{2,2}$	R_2
	C_1	C_2	T

For instance, with the same-sex twin data, twins come in two types, monozygotic and dizygotic. So one might use the Mantel Haenszel test to determine whether men are over-represented among left-handers as follows:

Sex	zygosity	NRH	RH	Excess	Variance
Male	MZ	46	431	$46 - (477)(75)/932$ $= 7.6$	$(477)(455)(75)(857)/((932^2)(931))$ $= 17.25$
Female	MZ	29	426		
Male	DZ	73	743	$73 - (816)(123)/1577$ $= 9.4$	$(816)(761)(123)(1454)/((1577^2)(1576))$ $= 27.96$
Female	DZ	50	711		
Totals				17.0	45.21

$Z = 17.0/\sqrt{45.21} = 2.53$ with one-sided p-value $= 0.0057$.

McNemar's Test

The data discussed above is also obviously paired and we have been treating it as though the twins are uncorrelated in their handedness. McNemar's test provides a way to handle paired binary response data.

The following example is from Rice's text. The data come from a study by Johnson and Johnson. It was a follow up study on whether tonsillectomies increase the risk of contracting Hodgkin's disease. They collected data from 85 sibling pairs: one of the siblings had Hodgkin's and the other sibling was within 5 years of age but was disease free. The original paper analyzed the data with the following 2×2 table:

	Tonsillectomy	No tonsillectomy
Hodgkin's	41	44
Control	33	52

which lead to a χ^2 statistic of 1.53 which is quite insignificant. However, this analysis ignores the pairing of the data. Ignoring the pairing inflates the variance and leads to a lack of power. The analysis that addresses the pairing is McNemar's test which analyzes the following 2×2 table instead:

Hodgkin's patient	Disease-free Control Sibling	
	No tonsillectomy	Tonsillectomy
No tonsillectomy	37	7
Tonsillectomy	15	26

The null hypothesis in this case is the tonsillectomy's are just as likely in the Hodgkin's group as in the control group. Since there are equal numbers of Hodgkin's patients and controls, this would mean the second row total is the same as the second column total. Equivalently, the first row total should be the same as the first column total. These equalities will hold exactly when the off-diagonal entries are equal. Currently they are in proportion 7:15. Is that statistically equal?

The test is a Chi-Square test with 1 degrees of freedom but it is calculated as:

$$\chi^2 = \frac{(n_{1,2} - n_{2,1})^2}{n_{1,2} + n_{2,1}}$$

For the table above, the value is 2.91 which has a p-value of 0.09 which is suggestive of tonsillectomies being associated with Hodgkin's disease.

As another example, for the handedness data, the author's were interested in whether the observation that there is more left-handedness in younger people had to do with less cultural pressure to be right-handed today or if older left-handed individuals were more likely to die than their right-handed counterparts. This leads to the following, slightly peculiar, table that can be analyzed with McNemar's test. The data involves only 118 pairs of same-sex, opposite-handed twins.

	Left-Handed Twin	
Right-Handed Twin	Died first	Died second
Died First	0	67
Died Second	51	0

which leads to a Chi-Square statistic of 2.17 which is insignificant.

References and Readings

Exercises for Lecture 36

1. –

2. –