BIOS 6244 Analysis of Categorical Data
October 26, 2005 Lecture

Cochran-Mantel-Haenszel Methods (Sec. 3.2)

In this section, we will discuss the following:

(1) A test of conditional independence,
(2) A test of homogeneous association, and
(3) A summary measure of partial association

Example (meta analysis examining smoking vs. lung cancer)

Consider Table 3.3, p. 60, in our textbook:

Table 3.3 Chinese Smoking and Lung Cancer Study, with Information Relevant to Cochran–Mantel–Haenszel Test

		Lung Cancer				
City	Smoking	Yes	No	Odds Ratio	μ_{11k}	Var(n_{11k})
Beijing	Smokers	126	100	2.20	113.0	16.9
	Nonsmokers	35	61			
Shanghai	Smokers	908	688	2.14	773.2	179.3
	Nonsmokers	497	807			
Shenyang	Smokers	913	747	2.18	799.3	149.3
	Nonsmokers	336	598			
Nanjing	Smokers	235	172	2.85	203.5	31.1
	Nonsmokers	58	121			
Harbin	Smokers	402	308	2.32	355.0	57.1
	Nonsmokers	121	215			
Zhengzhou	Smokers	182	156	1.59	169.0	28.3
	Nonsmokers	72	98			
Taiyuan	Smokers	60	99	2.37	53.0	9.0
	Nonsmokers	11	43			
Nanchang	Smokers	104	89	2.00	96.5	11.0
	Nonsmokers	21	36			

Source: Based on data in Z. Liu, Smoking and lung cancer in China, *Intern. J. Epidemiol.*, *21*: 197–201 (1992). Reprinted with permission of Oxford University Press.

These data summarize the results from 8 different case-control studies conducted in China that examined the association between smoking and lung cancer. The confounding (or control) variable to be considered here is the city in which the study was conducted. Subjects in different cities may vary in terms of relevant characteristics such as socioeconomic status, which may result in heterogeneity among the cities in terms of the E-D association between smoking and lung cancer. Thus, we wish to control for the possible confounding effect of "city."

Cochran-Mantel-Haenszel Test (Sec. 3.2.1)

For a 2x2xK table like the one represented in Table 3.3, the null hypothesis that X and Y are conditionally independent, given Z, means that $OR_{XY(k)} = 1$ for k = 1, 2, …, K. The standard sampling models for 2x2xK tables treat the observed cell counts $\{n_{ijk}\}$ as one of the following:

(1) Independent Poisson random variables (RV's),
(2) Multinomial RV's with fixed n,
(3) Multinomial RV's with fixed sample size for each of the K partial tables, with counts in different partial tables being independent, or
(4) Independent binomial samples within each partial table with row totals fixed.

In the k'th partial table, the row totals are denoted by $\{n_{1+k}, n_{2+k}\}$ and the column totals by $\{n_{+1k}, n_{+2k}\}$. Given both sets of totals, all 4 samping schemes mentioned above yield a hypergeometric distribution for the count n_{11k} in the k'th partial table. As we saw when discussing Fisher's exact test in Sec. 2.6.1, the count in this cell (1st row, 1st column) determines all of the other counts in the partial table, assuming that the row and column totals are known for that layer. As with Fisher's exact test for a 2x2 table, the $\{n_{11k}\}$ will be used to determine the test statistic for a 2x2xK table.

Under the null hypothesis of conditional independence,

$$E(n_{11k}) = \mu_{11k} = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

$$\text{Var}(n_{11k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n^2_{++k}(n_{++k}-1)}. \qquad (8)$$

When the true $OR_{XY(k)} > 1$ in partial table k, we expect to observe $n_{11k} - \mu_{11k} > 0$. The test statistic for the test of conditional independence combines these differences across all K partial tables. If $OR_{XY(k)} >> 1$ in every partial table, then the sum of such differences will be a large positive number; if $OR_{XY(k)} << 1$ in every partial table, then the sum of such differences will be a large negative number.

The test statistic proposed independently by Cochran and by Mantel & Haenszel combines the information from the K partial tables as follows:

$$\text{CMH} = \frac{\left[\sum_k (n_{11k} - \mu_{11k})\right]^2}{\sum_k \text{Var}(n_{11k})} = \frac{\left[\sum_k n_{11k} - \sum_k \mu_{11k}\right]^2}{\sum_k \text{Var}(n_{11k})}. \qquad (9)$$

This is called the *Cochran-Mantel-Haenszel (CMH) statistic* (or sometimes just the *Mantel-Haenszel statistic*). It has an approximate $\chi^2(1)$ distribution under the null hypothesis when n is "large enough."

The CMH test statistic takes larger values when $n_{11k} - \mu_{11k}$ is consistently positive [$OR_{XY(k)} > 1$] or consistently negative [$OR_{XY(k)} < 1$] for all tables, rather than positive for some and negative for others. *The CMH test is inappropriate when the association between X and Y varies dramatically across the partial tables* (i.e., when several $OR_{XY(k)} >> 1$ and several $OR_{XY(k)} << 1$). It works best when the X-Y association is "similar" in each partial table.

When the true association $OR_{XY(k)}$ is about the same in each partial table, the CMH test is more powerful than performing separate tests of the X-Y association in each table. We have already seen that, for partial tables exhibiting Simpson's paradox, we cannot rely on tests based on the marginal table since the results will not be consistent with those based on the partial tables.

Example, cont. (meta-analysis of smoking vs. lung cancer)

Each of the studies in Table 3.3 matched cases of lung cancer with controls not having lung cancer and then recorded whether each subject had ever been a smoker. Thus, in each layer (≡ city), we should treat the column totals as fixed and the counts within each column as a binomial sample. We wish to test the hypothesis of conditional independence between smoking and lung cancer, i.e., that $OR_{XY(k)} = 1$ for all k = 1, 2, …, K.

For each study, Table 3.3 contains $\widehat{OR}_{XY(k)}$, $\mu_{11k} = E(n_{11k})$, and $Var(n_{11k})$, calculated using the formulas in Equation (8) above. In each study, there was a moderate (1.59 – 2.85) positive association between smoking and lung cancer, so it is appropriate to use the CMH test statistic.

Applying the formula in Equation (9) above, we obtain

$$\sum_k n_{11k} = 2930, \ \sum_k \mu_{11k} = 2562.5, \ \sum_k Var(n_{11k}) = 482.1$$

$$\Rightarrow \qquad CMH = \frac{(2930 - 2562.5)^2}{482.1} = 280.1.$$

Using a $\chi^2(1)$ distribution to calculate the approximate the p-value, we find $p < .0001$. Thus, we conclude that there is extremely strong evidence that smoking and lung cancer are not independent and that $OR_{XY} > 1$.

Note that the evidence against independence is far greater for the combined data than in any one study individually. This is one of the strengths of meta-analysis.

As in any statistical inference situation, it is not enough to just test the null hypothesis and report a p-value. We must also pay attention to estimation of the unknown parameter. If we reject the null hypothesis of conditional independence [i.e., $OR_{XY(k)} = 1$ for all layers] using the CMH test, then we should proceed to estimating the true values of the conditional OR's.

The 1st step is to test the null hypothesis of homogeneous association, i.e., test

$$H_0: \ OR_{XY(1)} = OR_{XY(2)} = \ \ldots = OR_{XY(K)}.$$

If there is insufficient evidence to reject H_0, then we can conclude that it is plausible that all of the conditional OR's are equal to some common value OR_{XY}. We can then find an approximate 95% CI(OR_{XY}).

Testing Homogeneity of Conditional OR's (Sec. 3.2.4)

To test H_0: $OR_{XY(1)} = OR_{XY(2)} = \ldots = OR_{XY(K)}$, we compare the observed cell counts in the 2x2xK table with the expected cell counts under H_0 using the usual Pearson χ^2 test statistic:

$$X^2 = \sum_i \sum_j \sum_k \frac{\left(n_{ijk} - \mu_{ijk}\right)^2}{\mu_{ijk}}.$$

Under H_0, we estimate μ_{ijk} by using $\{\hat{\mu}_{11k}, \hat{\mu}_{12k}, \hat{\mu}_{21k}, \hat{\mu}_{22k}\}$, which denote the expected cell frequencies in the k'th partial table that has the same marginal totals as the observed partial table, yet has $OR_{XY(k)}$ equal to the Mantel-Haenszel estimate $\widehat{OR}_{MH}$ of a common odds ratio. (See Section 3.2.3.)

The closer the n_{ijk} are to the $\hat{\mu}_{ijk}$ from a table with $OR_{XY(k)} = \widehat{OR}_{MH}$, the smaller the value of X^2 and the smaller the evidence against H_0.

The test based on X^2 above is called the *Breslow-Day test* and has an approximate $\chi^2(K-1)$ distribution, which is valid as long as $\hat{\mu}_{ijk} \geq 5$ in at least 80% of the cells.

Calculation of the $\{\hat{\mu}_{ijk}\}$ for the Breslow-Day test statistic is rather complex and will not be covered in this course. The Breslow-Day test is available in SAS, SPSS, and StatXact.

Example, cont. (meta-analysis of smoking vs. lung cancer)

For the data in Table 3.3, the Breslow-Day test statistic is $X^2 = 5.2$, and d.f. = 8 – 1 = 7. The approximate p-value for the Breslow-Day test is therefore

$$\text{p-value} \approx \Pr(X^2 \geq 5.2) = .636.$$

Thus, we have no evidence against the null hypothesis of homogeneous X-Y association and, therefore, we can proceed to the estimation of the common conditional OR_{XY}.

Estimation of a Common Conditional OR (Sec.3.2.3)

The Mantel-Haenszel estimator of the common conditional OR is given by

$$\widehat{OR}_{MH} = \frac{\sum_k \frac{n_{11k} n_{22k}}{n_{++k}}}{\sum_k \frac{n_{12k} n_{21k}}{n_{++k}}}. \qquad (10)$$

Note that the numerator consists of the sum of main diagonal cross products weighted by the inverse of the total sample size for that layer and the denominator consists of the same weighted sum of the off-diagonal cross-products.

As in our approximate confidence interval procedures for an unconditional OR in Section 2.3.3, we base our approximate confidence interval estimation of OR_{MH} on the approximate null distribution of $\log\left(\widehat{OR}_{MH}\right)$. To derive an approximate CI for $\log(OR_{MH})$, we need the ASE of $\log\left(\widehat{OR}_{MH}\right)$. However, this has a rather complex form, so it will not be presented in this course. However, SAS, SPSS and StatXact can all perform these calculations.

For the data in Table 3.3,

$$\widehat{OR}_{MH} = \frac{\frac{(126)(61)}{322} + \ldots + \frac{(104)(36)}{250}}{\frac{(35)(100)}{322} + \ldots + \frac{(21)(89)}{250}} = 2.17.$$

The ASE of $\log\left(\widehat{OR}_{MH}\right)$ =.046. Thus, an approximate 95% CI for $\log(OR_{MH})$ is given by

$$\log(2.17) \pm 1.96(.046) = .777 \pm .091 = (.686, .868).$$

Back-transforming yields an approximate 95% CI(OR_{MH}): $(e^{.686}, e^{.868}) = (1.98, 2.38)$.

Note how narrow this approximate interval is. It enables us to conclude that the odds of lung cancer for smokers are about twice those of non-smokers. Of course, this precise estimation of the common OR is not surprising since the total sample size for all studies in Table 3.3 is 8,419.

Example, revisiting (ulcer drugs)

It is interesting to examine the effect of using the M-H estimate of the common OR with data that exhibit Simpson's paradox. Consider the data on the effectiveness of 2 ulcer drugs obtained at 2 different treatment sites (see pp. 48-49 of these notes, from the October 24, 2005 lecture):

		Cured	Not Cured
Site 1	Drug A	40	160
	Drug B	30	170

		Cured	Not Cured
Site 2	Drug A	85	15
	Drug B	300	100

Recall that the conditional odds ratios were $\widehat{OR}_{XY(1)} = 1.42$ and $\widehat{OR}_{XY(2)} = 1.89$, and that the marginal odds ratio was $\widehat{OR} = .58$. These OR's provide evidence of Simpson's Paradox. The Mantel-Haenszel estimate of the common conditional OR is

$$\widehat{OR}_{MH} = \frac{\frac{(40)(170)}{400} + \frac{(85)(100)}{500}}{\frac{(30)(160)}{400} + \frac{(15)(300)}{500}} = 1.62$$

Note that this is almost identical to the mean of the conditional OR's from the 2 sites. It is a much more representative measure of the effectiveness of Drug A vs. Drug B and is consistent with the results obtained at each of the 2 sites. The effect of Simpson's paradox has been "removed" in this measure of the overall effectiveness of Drug A.

Some Caveats Related to Mantel-Haenszel Analysis (Sec. 3.2.5)

(1) Sometimes the χ^2 approximation for the Breslow-Day test statistic is not very accurate even when n is fairly large. An adjustment proposed by Tarone that improves the χ^2 approximation involves subtracting the following term from the usual value of the test statistic:

$$\frac{\left[\sum_k \left(n_{11k} - \widehat{\mu_{11k}}\right)\right]^2}{\sum_k \left[\frac{1}{\widehat{\mu}_{11k}} + \frac{1}{\widehat{\mu}_{12k}} + \frac{1}{\widehat{\mu}_{21k}} + \frac{1}{\widehat{\mu}_{22k}}\right]^{-1}}.$$

This adjustment is usually minor, but can have a substantial impact if any of the $\hat{\mu}_{ijk}$ are extremely small.

(2) It is common in case-control studies (as in Table 3.3) to match cases and controls on one or more potential confounding variables (age, gender, etc.). This means that, technically, the columns in the 2x2 tables in each layer no longer represent independent binomial samples. In this case, Mantel-Haenszel methods are only approximate. Methods that take the dependence between cases and controls into account are covered in Chapter 9 in our text.

(3) The methods of this chapter can be generalized to IxJxK tables (see Sec. 7.3 of our text).

(4) Another test of the homogeneity of conditional OR's is covered in Sec. 6.5.1 of our text.