

equals X^2 . (Note: For 2×2 tables, X^2 simplifies to

$$X^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{(n_{1+}n_{2+}n_{+1}n_{+2})}$$

Adjusted residuals are identical for Poisson and binomial sampling. The Pearson residual defined here refers to Poisson sampling, and a different Pearson residual applies for binomial sampling; see Section 5.3.3.)

2.33. Formula (2.4.3) has alternative expression $X^2 = n \sum (p_{ij} - p_{i+}p_{+j})^2 / p_{i+}p_{+j}$. For a particular set of $\{p_{ij}\}$, X^2 is directly proportional to n . Hence, X^2 can be large when n is large, regardless of whether the association is practically important. Explain why chi-squared tests, like other tests, simply indicate the degree of evidence against a hypothesis and do not give information about the strength of association. ("Like fire, the chi-square test is an excellent servant and a bad master," Sir Austin Bradford Hill, *Proc. R. Soc. Med.*, 58: 295-300 (1965).)

2.34. Let Z denote a standard normal variate. Then Z^2 has a chi-squared distribution with $df = 1$. A chi-squared variate with degrees of freedom equal to df has representation $Z_1^2 + \dots + Z_{df}^2$, where Z_1, \dots, Z_{df} are independent standard normal variates. Using this, show that if Y_1 and Y_2 are independent chi-squared variates with degrees of freedom df_1 and df_2 , then $Y_1 + Y_2$ has a chi-squared distribution with $df = df_1 + df_2$.

CHAPTER 3

Three-Way Contingency Tables

An important part of most research studies is the choice of control variables. In studying the effect of an explanatory variable X on a response variable Y , one should "control" covariates that can influence that relationship. That is, one should use some mechanism to hold such covariates constant while studying the effect of X on Y . Otherwise, an observed effect of X on Y may simply reflect effects of those covariates on both X and Y . This is particularly true for observational studies, where one does not have the luxury of randomly assigning subjects to different treatments.

To illustrate, suppose a study considers effects of passive smoking; that is, the effects on a nonsmoker of living with a smoker. To analyze whether passive smoking is associated with lung cancer, a cross-sectional study might compare lung cancer rates between nonsmokers whose spouses smoke and nonsmokers whose spouses do not smoke. In doing so, we should attempt to control for age, socioeconomic status, or other factors that might relate both to whether one's spouse smokes and to whether one has lung cancer. Unless we control such variables, results will have limited usefulness. Suppose spouses of nonsmokers tend to be younger than spouses of smokers, and suppose younger people are less likely to have lung cancer. Then, a lower proportion of lung cancer cases among spouses of nonsmokers may simply reflect their lower average age.

Including control variables in an analysis requires a multivariate rather than a bivariate analysis. This chapter generalizes the methods of Chapter 2 regarding two-way contingency tables to multi-way tables. The main topic is analyzing the association between two categorical variables X and Y , while controlling for effects of a possibly confounding variable Z . We do this by studying the X - Y relationship at fixed, constant levels of Z . For simplicity, the examples refer to three-way tables with binary responses. Later chapters treat more general cases as well as the use of models to perform statistical control.

Section 3.1 shows that the association between two variables may change dramatically under a control for another variable. Sections 3.2 and 3.3 present inferential methods for such associations; Section 3.2 presents large-sample methods and Section 3.3 discusses exact inference for small samples.

3.1 PARTIAL ASSOCIATION

We begin this chapter by discussing statistical control and the types of relationships one can encounter in performing such control with multivariate categorical data. We illustrate basic concepts for a response variable *Y*, an explanatory variable *X*, and a single control variable *Z*, all of which are categorical. A three-way contingency table displays counts for the combinations of levels of the three variables.

3.1.1 Partial Tables

Two-way cross-sectional slices of the three-way table cross classify *X* and *Y* at separate levels of the control variable *Z*. These cross sections are called *partial tables*. They display the *X-Y* relationship at fixed levels of *Z*, hence showing the effect of *X* on *Y* while controlling for *Z*. The partial tables remove the effect of *Z* by holding its value constant.

The two-way contingency table obtained by combining the partial tables is called the *X-Y marginal table*. Each cell count in the marginal table is a sum of counts from the same cell location in the partial tables. The marginal table, rather than controlling *Z*, ignores it. The marginal table contains no information about *Z*. It is simply a two-way table relating *X* and *Y*. Methods for two-way tables, discussed in Chapter 2, do not take into account effects of other variables.

The associations in partial tables are called *conditional associations*, because they refer to the effect of *X* on *Y* conditional on fixing *Z* at some level. Conditional associations in partial tables can be quite different from associations in marginal tables. In fact, it can be misleading to analyze only a marginal table of a multi-way contingency table, as the following example illustrates.

3.1.2 Death Penalty Example

Table 3.1 is a 2 × 2 × 2 contingency table—two rows, two columns, and two layers—from an article that studied effects of racial characteristics on whether individuals convicted of homicide receive the death penalty. The 674 subjects classified in Ta-

Table 3.1 Death Penalty Verdict by Defendant's Race and Victims' Race

| Victims' Race | Defendant's Race | Death Penalty | | Percentage Yes |
|---------------|------------------|---------------|-----|----------------|
| | | Yes | No | |
| White | White | 53 | 414 | 11.3 |
| | Black | 11 | 37 | 22.9 |
| Black | White | 0 | 16 | 0.0 |
| | Black | 4 | 139 | 2.8 |
| Total | White | 53 | 430 | 11.0 |
| | Black | 15 | 176 | 7.9 |

Source: M. L. Radelet and G. L. Pierce, *Florida Law Rev.* 43: 1-34 (1991). Reprinted with permission of the *Florida Law Review*.

ble 3.1 were the defendants in indictments involving cases with multiple murders in Florida between 1976 and 1987. The variables in Table 3.1 are *Y* = "death penalty verdict," having categories (yes, no), and *X* = "race of defendant" and *Z* = "race of victims," each having categories (white, black). We study the effect of defendant's race on the death penalty verdict, treating victims' race as a control variable. Table 3.1 has a 2 × 2 partial table relating defendant's race and the death penalty verdict at each level of victims' race.

For each combination of defendant's race and victims' race, Table 3.1 lists and Figure 3.1 displays the percentage of defendants who received the death penalty. We use these to describe the conditional associations between defendant's race and the death penalty verdict, controlling for victims' race. When the victims were white, the death penalty was imposed 22.9% — 11.3% = 11.6% more often for black defendants than for white defendants. When the victim was black, the death penalty was imposed 2.8% more often for black defendants than for white defendants. Thus, *controlling* for victims' race by keeping it fixed, the percentage of "yes" death penalty verdicts was higher for black defendants than for white defendants.

The bottom portion of Table 3.1 displays the marginal table for defendant's race and the death penalty verdict. We obtain it by summing the cell counts in Table 3.1 over the two levels of victims' race, thus combining the two partial tables (e.g., 11 + 4 = 15). We see that, overall, 11.0% of white defendants and 7.9% of black defendants received the death penalty. *Ignoring* victims' race, the percentage of "yes"

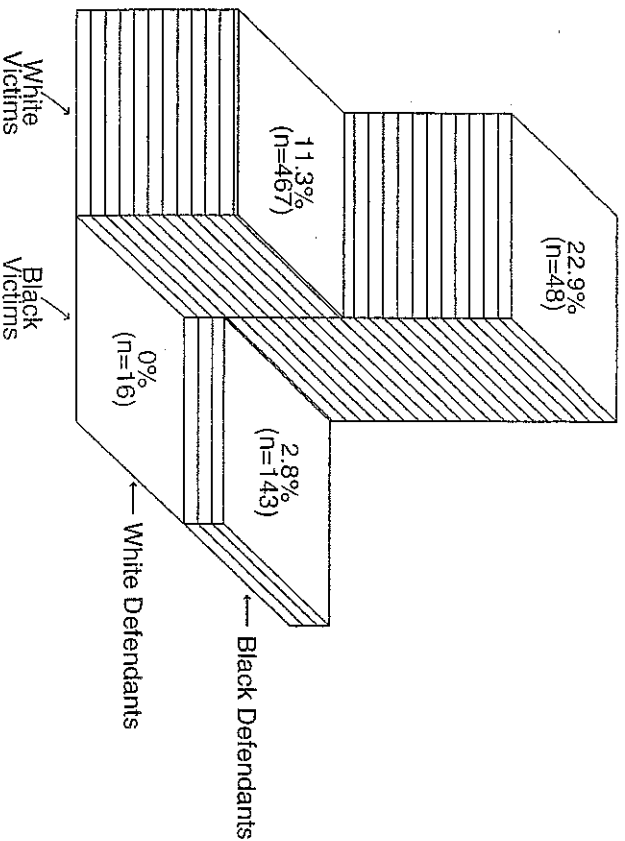


Figure 3.1 Percent receiving death penalty.

death penalty verdicts was lower for black defendants than for white defendants. The association reverses direction compared to the partial tables.

Why does the association between death penalty verdict and defendant's race differ so much when we ignore vs. control victims' race? This relates to the nature of the association between the control variable, victims' race, and each of the other variables. First, the association between victims' race and defendant's race is extremely strong. One can verify that the marginal table relating these variables has odds ratio $(467 \times 143)/(48 \times 16) = 87.0$; the odds that a white defendant had white victims are estimated to be 87.0 times the odds that a black defendant had white victims. Second, the percentages in Table 3.1 show that, regardless of defendant's race, the death penalty was considerably more likely when the victims were white than when the victims were black. So, whites are tending to kill whites, and killing whites is more likely to result in the death penalty. This suggests that the marginal association should show a greater tendency for white defendants to receive the death penalty than do the conditional associations. In fact, Table 3.1 shows this pattern.

Figure 3.2 may clarify why the conditional associations differ so from the marginal association. For each defendant's race, the figure plots the proportion receiving the death penalty at each level of victims' race. Each proportion is labeled by a letter

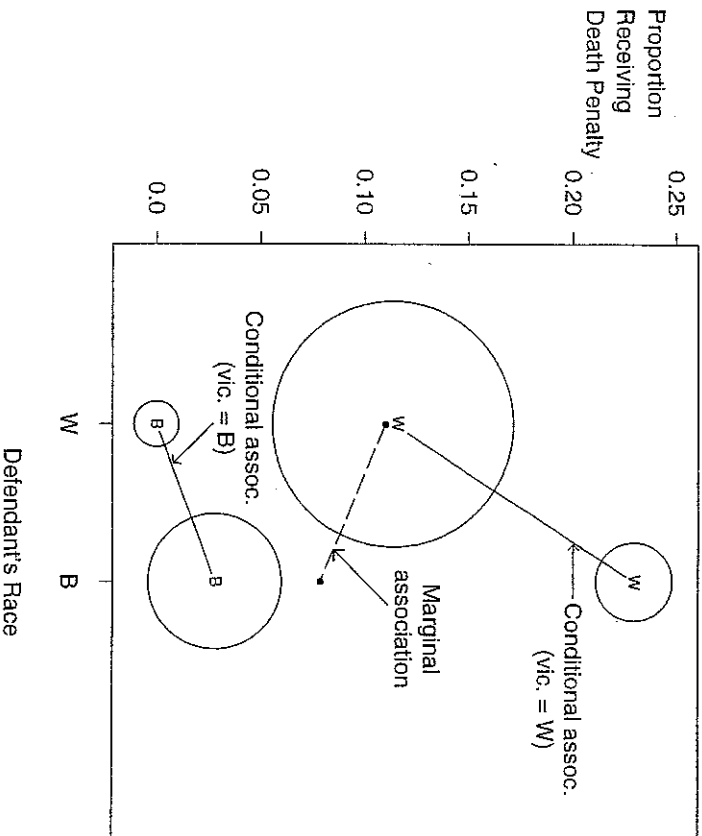


Figure 3.2 Proportion receiving death penalty by defendant's race, controlling and ignoring victim's race.

symbol giving the level of victims' race. Surrounding each observation is a circle having area proportional to the number of observations at that combination of defendant's race and victims' race. For instance, the W in the largest circle represents a proportion of .113 receiving the death penalty for cases with white defendants and white victims. That circle is largest, because the number of cases at that combination $(53 + 414 = 467)$ is larger than at the other three combinations. The next largest circle relates to cases in which blacks kill blacks.

We control for victims' race by comparing circles having the same victims' race letter at their centers. The line connecting the two W circles has a positive slope, as does the line connecting the two B circles. Controlling for victims' race, this reflects a higher chance of the death penalty for black defendants than white defendants. When we add results across victims' race to get a summary result for the marginal effect of defendant's race on the death penalty verdict, the larger circles having the greater number of cases have greater influence. Thus, the summary proportions for each defendant's race, marked on the figure by periods, fall closer to the center of the larger circles than the smaller circles. A line connecting the summary marginal proportions has negative slope, indicating that white defendants are more likely than black defendants to receive the death penalty.

The result that a marginal association can have different direction from the conditional associations is called *Simpson's paradox*. This result applies to quantitative as well as categorical variables.

3.1.3 Conditional and Marginal Odds Ratios

One can describe marginal and conditional associations using odds ratios. We illustrate for $2 \times 2 \times K$ tables, where K denotes the number of levels of a control variable, Z . Let $\{n_{ijk}\}$ denote observed frequencies and let $\{\mu_{ijk}\}$ denote their expected frequencies.

Within a fixed level k of Z ,

$$\theta_{XY(k)} = \frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}} \quad (3.1.1)$$

describes conditional X - Y association. It is the ordinary odds ratio computed for the four expected frequencies in the k th partial table. We refer to the odds ratios for the K partial tables as the X - Y conditional odds ratios.

The conditional odds ratios can be quite different from marginal odds ratios, for which the third variable is ignored rather than controlled. The X - Y marginal table has expected frequencies $\{\mu_{ij+}\} = \sum_k \mu_{ijk}$ obtained by summing over the levels of Z . The X - Y marginal odds ratio is

$$\theta_{XY} = \frac{\mu_{11+}\mu_{22+}}{\mu_{12+}\mu_{21+}}.$$

Similar formulas with cell counts substituted for expected frequencies provide sample estimates of $\theta_{XY(k)}$ and θ_{XY} .

We illustrate by computing sample conditional and marginal odds ratios for the association between defendant's race and the death penalty. From Table 3.1, the estimated odds ratio in the first partial table, for which victims' race is white, equals

$$\hat{\theta}_{XY(1)} = \frac{53 \times 37}{414 \times 11} = 0.43.$$

The sample odds for white defendants receiving the death penalty were 43% of the sample odds for black defendants. In the second partial table, for which victim's race is black, the estimated odds ratio equals $\hat{\theta}_{XY(2)} = (0 \times 139)/(16 \times 4) = 0.0$, since the death penalty was never given to white defendants having black victims.

Estimation of the marginal odds ratio for defendant's race and the death penalty uses the 2×2 marginal table in Table 3.1, collapsing over victims' race. The estimate equals $(53 \times 176)/(430 \times 15) = 1.45$. The sample odds of the death penalty were 45% higher for white defendants than for black defendants. Yet, we just observed that those odds were smaller for a white defendant than for a black defendant, within each level of victims' race. This reversal in the association when we control for victims' race illustrates Simpson's paradox. (Problems 5.16 and 6.3 consider further analyses of these data.)

3.1.4 Marginal versus Conditional Independence

Consider the true relationship between X and Y , controlling for Z . If X and Y are independent in each partial table, then X and Y are said to be *conditionally independent, given Z*. All conditional odds ratios between X and Y then equal 1. Conditional independence of X and Y , given Z , does not imply marginal independence of X and Y . That is, when odds ratios between X and Y equal 1 at each level of Z , the marginal odds ratio may differ from 1.

The expected frequencies in Table 3.2 show a hypothetical relationship among three variables: $Y =$ response (success, failure), $X =$ drug treatment (A, B), and $Z =$ clinic (1, 2). The conditional association between X and Y at the two levels of

Table 3.2 Conditional Independence Does Not Imply Marginal Independence

| Clinic | Treatment | Response | |
|--------|-----------|----------|---------|
| | | Success | Failure |
| 1 | A | 18 | 12 |
| | B | 12 | 8 |
| 2 | A | 2 | 8 |
| | B | 8 | 32 |
| Total | A | 20 | 20 |
| | B | 20 | 40 |

Z is described by the odds ratios

$$\theta_{XY(1)} = \frac{18 \times 8}{12 \times 12} = 1.0$$

$$\theta_{XY(2)} = \frac{2 \times 32}{8 \times 8} = 1.0.$$

Given clinic, response and treatment are conditionally independent. The marginal table adds together the tables for the two clinics. The odds ratio for that marginal table equals $\theta_{XY} = (20 \times 40)/(20 \times 20) = 2.0$, so the variables are not marginally independent.

Why are the odds of a success twice as high for treatment A as treatment B when we ignore clinic? The conditional X - Z and Y - Z odds ratios give a clue. The odds ratio between Z and either X or Y , at each fixed level of the other variable, equals 6.0. For instance, the X - Z odds ratio at the first level of Y equals $(18)(8)/(12)(2) = 6.0$. The conditional odds (given response) of receiving treatment A are six times higher at clinic 1 than clinic 2, and the conditional odds (given treatment) of success are six times higher at clinic 1 than at clinic 2. Clinic 1 tends to use treatment A more often, and clinic 1 also tends to have more successes. For instance, if patients who attend clinic 1 tend to be in better health or tend to be younger than those who go to clinic 2, perhaps they have a better success rate than subjects in clinic 2 regardless of the treatment received.

It is misleading to study only the marginal table, concluding that successes are more likely with treatment A than with treatment B . Subjects within a particular clinic are likely to be more homogeneous than the overall sample, and response is independent of treatment in each clinic.

3.1.5 Homogeneous Association

There is *homogeneous X-Y association* in a $2 \times 2 \times K$ table when

$$\theta_{XY(1)} = \theta_{XY(2)} = \dots = \theta_{XY(K)}.$$

The conditional odds ratio between X and Y is then identical at each level of Z . Thus, the effect of X on Y is the same at each level of Z , and a single number describes the X - Y conditional associations. Conditional independence of X and Y is the special case in which each conditional odds ratio equals 1.0.

Homogeneous X - Y association in an $I \times J \times K$ table means that any conditional odds ratio formed using two levels of X and two levels of Y is the same at each level of Z . When X - Y conditional odds ratios are identical at each level of Z , the same property holds for the other associations. For instance, the conditional odds ratio between two levels of X and two levels of Z is identical at each level of Y . Homogeneous association is a symmetric property, applying to any pair of the variables viewed across the levels of the third. When it occurs, there is said to be *no interaction* between two variables in their effects on the third variable.

When homogeneous association does not exist, the conditional odds ratio for any pair of variables changes across levels of the third variable. For $X = \text{smoking (yes, no)}$, $Y = \text{lung cancer (yes, no)}$, and $Z = \text{age (<45, 45-65, >65)}$, suppose $\theta_{XY(1)} = 1.2$, $\theta_{XY(2)} = 2.8$, and $\theta_{XY(3)} = 6.2$. Then, smoking has a weak effect on lung cancer for young people, but the effect strengthens considerably with age.

The estimated conditional odds ratios for Table 3.1 are $\hat{\theta}_{XY(1)} = 0.43$ and $\hat{\theta}_{XY(2)} = 0.0$. The values are not close, but the second estimate is unstable because of the zero cell count. If we add $\frac{1}{2}$ to each cell count, we obtain 0.94 for the second estimate. Because the second estimate is so unstable and because further variation can occur from sampling variability, these data do not necessarily contradict homogeneous association. The next section shows how to check whether sample data are consistent with homogeneous association or conditional independence.

3.2 COCHRAN-MANTTEL-HAENSZEL METHODS

This section introduces inferential analyses for three-way tables. We present a test of conditional independence and a test of homogeneous association for the K conditional odds ratios in $2 \times 2 \times K$ tables. We also show how to combine the sample odds ratios from the K partial tables into a single summary measure of partial association.

Analyses of conditional association are relevant in most applications having multivariate data. To illustrate, we analyze Table 3.3, which summarizes eight studies in China about smoking and lung cancer. The smokers and nonsmokers are the levels

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

| City | Smoking | Lung Cancer | | Odds Ratio | μ_{11k} | $\text{Var}(n_{11k})$ |
|-----------|------------|-------------|-----|------------|-------------|-----------------------|
| | | Yes | No | | | |
| 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 | | | |
| Taiwan | 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.

of a group classification X , the two possible outcomes (yes, no) for lung cancer are the levels of a response variable Y , and different cities are levels of a control variable Z . Subjects may vary among cities on relevant characteristics such as socioeconomic status, which may cause heterogeneity among the cities in smoking rates and in the lung cancer rate. Thus, we investigate the association between X and Y while controlling for Z .

3.2.1 The Cochran-Mantel-Haenszel Test

For $2 \times 2 \times K$ tables, the null hypothesis that X and Y are conditionally independent, given Z , means that the conditional odds ratio $\theta_{XY(z)}$ between X and Y equals 1 in each partial table. The standard sampling models treat the cell counts as (1) independent Poisson variates, or (2) multinomial counts with fixed overall sample size, or (3) multinomial counts with fixed sample size for each partial table, with counts in different partial tables being independent, or (4) independent binomial samples within each partial table with row totals fixed. In partial table k , the row totals are $\{n_{1+k}, n_{2+k}\}$, and the column totals are $\{n_{+1k}, n_{+2k}\}$. Given both these totals, all these sampling schemes yield a hypergeometric distribution (Section 2.6.1) for the count n_{11k} in the cell in the first row and first column. That cell count determines all other counts in the partial table. The test statistic utilizes this cell in each partial table.

Under the null hypothesis, the mean and variance of n_{11k} are

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

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

When the true odds ratio $\theta_{XY(z)}$ exceeds 1.0 in partial table k , we expect to observe $(n_{11k} - \mu_{11k}) > 0$. The test statistic combines these differences across all K tables. When the odds ratio exceeds 1.0 in every partial table, the sum of such differences tends to be a relatively large positive number; when the odds ratio is less than 1.0 in each table, the sum of such differences tends to be a relatively large negative number. The test statistic summarizes the information from the K partial tables using

$$CMH = \frac{[\sum_k (n_{11k} - \mu_{11k})]^2}{\sum_k \text{Var}(n_{11k})}. \quad (3.2.1)$$

This is called the *Cochran-Mantel-Haenszel (CMH) statistic*. It has a large-sample chi-squared distribution with $df = 1$.

The *CMH* statistic takes larger values when $(n_{11k} - \mu_{11k})$ is consistently positive or consistently negative for all tables, rather than positive for some and negative for others. This test is inappropriate when the association varies dramatically among the partial tables. It works best when the X - Y association is similar in each partial table.

The *CMH* statistic combines information across partial tables. When the true association is similar in each table, this test is more powerful than separate tests within

each table. It is improper to combine results by adding the partial tables together to form a single 2×2 marginal table for the test. Simpson's paradox (Section 3.1.2) revealed the dangers of collapsing three-way tables.

3.2.2 Lung Cancer Meta Analysis Example

Table 3.3 summarizes eight case-control studies in China about smoking and lung cancer. Each study matched cases of lung cancer with controls not having lung cancer and then recorded whether each subject had ever been a smoker. In each partial table, we treat the counts in each column as a binomial sample, with column total fixed. We test the hypothesis of conditional independence between smoking and lung cancer, which states that the true odds ratio equals 1.0 for each city.

Table 3.3 reports the sample odds ratio for each table and the expected value and variance of the number of lung cancer cases who were smokers (the count in the first row and first column), under this hypothesis. In each table, the sample odds ratio shows a moderate positive association, so it makes sense to combine results through the *CMH* statistic. We obtain $\sum_k n_{11k} = 2930$, $\sum_k \mu_{11k} = 2562.5$, and $\sum_k \text{Var}(n_{11k}) = 482.1$, for which $CMH = (2930.0 - 2562.5)^2 / 482.1 = 280.1$, with $df = 1$. There is extremely strong evidence against conditional independence ($P < .0001$). This is not surprising, given the large sample size for the combined studies ($n = 8419$).

A statistical analysis that combines information from several studies is called a *meta analysis*. The meta analysis of Table 3.3 provides stronger evidence of an association than any single partial table gives by itself.

3.2.3 Estimation of Common Odds Ratio

It is more informative to estimate the strength of association than simply to test a hypothesis about it. When the association seems stable across partial tables, we can estimate an assumed common value of the K true odds ratios.

In a $2 \times 2 \times K$ table, suppose that $\theta_{XY(1)} = \dots = \theta_{XY(K)}$. The *Mantel-Haenszel estimator* of that common value equals

$$\hat{\theta}_{MH} = \frac{\sum_k (n_{11k} n_{22k} / n_{++k})}{\sum_k (n_{12k} n_{21k} / n_{++k})}. \quad (3.2.2)$$

The standard error for $\log(\hat{\theta}_{MH})$ has a complex formula (Agresti (1990), p. 236), so we shall not report it here. Some software (such as SAS-PROC FREQ) computes a standard error, and one can also obtain an estimate and standard error using logit models (Section 5.4.4).

For the Chinese smoking studies summarized in Table 3.3, the Mantel-Haenszel odds ratio estimate equals

$$\hat{\theta}_{MH} = \frac{(126)(61)/(322) + \dots + (104)(36)/250}{(35)(100)/(322) + \dots + (21)(89)/(250)} = 2.17.$$

The estimated standard error of $\log(\hat{\theta}_{MH}) = \log(2.17) = 0.777$ equals 0.046. An approximate 95% confidence interval for the common log odds ratio is $0.777 \pm 1.96 \times 0.046$, or $(0.686, 0.868)$, corresponding to $(\exp(.686), \exp(.868)) = (1.98, 2.38)$ for the odds ratio. The odds of lung cancer for smokers equal about twice the odds for non-smokers. Such odds ratios are typically much larger in Western society. They may be lower in China partly because, until recent years, pipes with long stems were more common than cigarettes.

If the true odds ratios are not identical but do not vary drastically, $\hat{\theta}_{MH}$ still provides a useful summary of the K conditional associations. Similarly, the *CMH* test is a powerful summary of evidence against the hypothesis of conditional independence, as long as the sample associations fall primarily in a single direction.

3.2.4 Testing Homogeneity of Odds Ratios*

One can test the hypothesis that the odds ratio between X and Y is the same at each level of Z ; that is, $H_0 : \theta_{XY(1)} = \dots = \theta_{XY(K)}$. This is a test of homogeneous association for $2 \times 2 \times K$ tables.

Let $\{\hat{\mu}_{11k}, \hat{\mu}_{12k}, \hat{\mu}_{21k}, \hat{\mu}_{22k}\}$ denote estimated expected frequencies in the k th partial table that have the same marginal totals as the observed data, yet have odds ratio equal to the Mantel-Haenszel estimate $\hat{\theta}_{MH}$ of a common odds ratio. The test statistic, called the *Breslow-Day statistic*, has the Pearson form

$$\sum \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}, \quad (3.2.3)$$

where the sum is taken over all cells in the table. The closer the cell counts fall to the values having a common odds ratio, the smaller the statistic and the less the evidence against H_0 .

Calculation of the $\{\hat{\mu}_{ijk}\}$ satisfying a common odds ratio is complex and not discussed here since standard software (such as PROC FREQ in SAS) reports this statistic. The Breslow-Day statistic is an approximate chi-squared statistic with $df = K - 1$. The sample size should be relatively large in each partial table, with $\{\hat{\mu}_{ijk} \geq 5\}$ in at least about 80% of the cells.

For the meta analysis of smoking and lung cancer (Table 3.3), software reports a Breslow-Day statistic equal to 5.2, based on $df = 7$, for which $P = .64$. This evidence does not contradict the hypothesis of equal odds ratios. We are justified in summarizing the conditional association by a single odds ratio for all eight partial tables.

3.2.5 Some Caveats*

To ensure that the distribution of the Breslow-Day statistic (3.2.3) converges to chi-squared as the sample size increases, R. Tarone showed (*Biometrika*, 72: 91-95 (1985)) that one must adjust it by subtracting

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

This adjustment is usually minor. For Table 3.3, the correction is less than 0.01, making no difference to the conclusion.

Each partial table in Table 3.3 refers to a case-control study. Because each subject having lung cancer was matched with one or more controls not having lung cancer, the two columns are not truly independent binomial samples. In practice, whether a case had ever been a smoker is probably essentially independent of whether a control had ever been a smoker, so the different columns in each table should be similar to independent binomial samples. If for each case-control pair we had information about whether each subject had been a smoker, we could form a 2×2 table relating whether the control had ever been a smoker (yes, no) to whether the case had ever been a smoker (yes, no). Then specialized methods are available for comparing cases and controls on the proportions who ever smoked that take into account potential dependence. Chapter 9 discusses some of these. Positive correlations between proportions yield smaller P-values than we get by treating the samples as independent.

Section 6.5.1 discusses an alternative test of homogeneity of odds ratios, based on models. Section 7.3 presents generalizations of the *CHM* test for $I \times J \times K$ tables.

3.3 EXACT INFERENCE ABOUT CONDITIONAL ASSOCIATIONS*

The chi-squared tests presented in the previous section, like chi-squared tests of independence for $I \times J$ tables, are large-sample tests. The true sampling distributions converge to chi-squared as the sample size n increases. It is difficult to provide general guidelines about how large n must be. The tests' adequacy depends more on the two-way marginal totals than on counts in the separate partial tables. For the *CMH* statistic, for instance, cell counts in the partial tables can be small (which often happens when K is large), but the X - Y marginal totals should be relatively large.

In practice, small sample sizes are not problematic, since one can conduct *exact* inference about conditional associations. For instance, exact tests of conditional independence generalize Fisher's exact test for 2×2 tables.

3.3.1 Exact Test of Conditional Independence for $2 \times 2 \times K$ Tables

For $2 \times 2 \times K$ tables, conditional on the marginal totals in each partial table, the Cochran-Mantel-Haenszel test of conditional independence depends on the cell counts through $\sum_k n_{11k}$. Exact tests use $\sum_k n_{11k}$ in the way they use n_{11} in Fisher's exact test for 2×2 tables. Hypergeometric distributions in each partial table determine probabilities for $\{n_{11k}, k = 1, \dots, K\}$. These determine the distribution of their sum.

The null hypothesis of conditional independence states that all conditional odds ratios $\{\theta_{XY(k)}\}$ equal 1. A "positive" conditional association corresponds to the one-sided alternative to this hypothesis, $\theta_{XY(k)} > 1$. The P-value then equals the right-tail probability that $\sum_k n_{11k}$ is at least as large as observed, for the fixed marginal totals. For the one-sided alternative $\theta_{XY(k)} < 1$, the P-value equals the left-tail probability that $\sum_k n_{11k}$ is no greater than observed. Two-sided alternatives can use a two-tail probability of those outcomes that are no more likely than the observed one.

Exact tests of conditional independence are computationally highly intensive. They require software for practical implementation. We used StatXact in the following example.

3.3.2 Promotion Discrimination Example

Table 3.4 refers to U.S. government computer specialists of similar seniority considered for promotion from classification level GS-13 to level GS-14. The table cross classifies promotion decision by employee's race, considered for three separate months. We test conditional independence of promotion decision and race. The table contains several small counts. The overall sample size is not small ($n = 74$), but one marginal count (collapsing over month of decision) equals zero, so we might be wary of using the *CMH* test.

We first use the one-sided alternative of an odds ratio less than 1. This corresponds to potential discrimination against black employees, their probability of promotion being lower than for white employees. Fixing the row and column marginal totals in each partial table, the test uses n_{11k} , the first cell count in each. For the margins of the partial tables in Table 3.4, n_{111} can range between 0 and 4, n_{112} can range between 0 and 4, and n_{113} can range between 0 and 2. The total $\sum_k n_{11k}$ can take values of the three cases. The observed value of $\sum_k n_{11k}$ equals 0, and the P-value is the null probability of this outcome, which software reveals to equal .026. A two-sided P-value, based on summing the probabilities of all tables having probabilities no greater than the observed table, equals .056. There is some evidence that promotion is related to race.

Table 3.4 Promotion Decisions by Race and by Month

| Race | July Promotions | | August Promotions | | September Promotions | |
|-------|-----------------|----|-------------------|----|----------------------|----|
| | Yes | No | Yes | No | Yes | No |
| Black | 0 | 7 | 0 | 7 | 0 | 8 |
| White | 4 | 16 | 4 | 13 | 2 | 13 |

Source: J. Gastwirth, *Statistical Reasoning in Law and Public Policy* (San Diego: Academic Press, 1988, p. 266).

PROBLEMS

- 3.1. In murder trials in 20 Florida counties during 1976 and 1977, the death penalty was given in 19 out of 151 cases in which a white killed a white, in 0 out of 9 cases in which a white killed a black, in 11 out of 63 cases in which a black killed a white, and in 6 out of 103 cases in which a black killed a black (M. Radelet, *Amer. Sociol. Rev.*, 46: 918–927 (1981)).
- Exhibit the data as a three-way contingency table.
 - Construct the partial tables needed to study the conditional association between defendant's race and the death penalty verdict. Compute and interpret the sample conditional odds ratios, adding 0.5 to each cell to reduce the impact of the 0 cell count.
 - Compute and interpret the sample marginal odds ratio between defendant's race and the death penalty verdict. Do these data exhibit Simpson's paradox? Explain.
- 3.2. For all trials in Florida involving homicides between 1976 and 1987, M. Radelet and G. Pierce (*Florida Law Review*, 43: 1–34 (1991)) reported the following results: The death penalty was given in 227 out of 4645 cases in which a white killed a white, in 92 out of 731 cases in which a black killed a white, in 9 out of 264 cases in which a white killed a black, and in 36 out of 4428 cases in which a black killed a black. Compute and interpret the sample conditional odds ratios between defendant's race and the death penalty verdict. Do the conditional associations seem to be homogeneous? Explain.
- 3.3. Smith and Jones are baseball players. Smith had a higher batting average than Jones in 1994 and 1995. Is it possible that for the combined data for these two years, Jones had the higher batting average? Explain, and illustrate using data.
- 3.4. Give a "real world" example of three variables X , Y , and Z , for which you expect X and Y to be marginally associated but conditionally independent, controlling for Z .
- 3.5. Based on 1987 murder rates in the United States, the Associated Press reported that the probability a newborn child has of eventually being a murder victim is 0.0263 for nonwhite males, 0.0049 for white males, 0.0072 for nonwhite females, and 0.0023 for white females.
- Find the conditional odds ratios between race and whether a murder victim, given gender. Interpret, and discuss whether these variables exhibit homogeneous association.
 - If half the newborns are of each gender, for each race, find the marginal odds ratio between race and whether a murder victim.
- 3.6. Using graphs or tables to illustrate, explain what is meant by "no interaction" in modeling a response Y and explanatory variables X and Z , when (a) all variables are continuous (multiple regression), (b) Y and X are continuous, Z is categorical (analysis of covariance), (c) Y is continuous, X and Z are categorical (two-way ANOVA), (d) all variables are categorical ("no interaction" = "homogeneous association").

3.3.3 Exact Confidence Interval for Common Odds Ratio

As discussed in Section 2.6.3, for small samples the discreteness implies that exact tests are conservative. When H_0 is true, for instance, the P -value may fall below .05 less than 5% of the time. One can alleviate conservativeness by using the mid- P value. It sums half the probability of the observed result with the probability of "more extreme" tables.

One can also construct "exact" confidence intervals for an assumed common value θ of $\{\theta_{X^*Y^*}\}$. Because of discreteness, these are also conservative. For a 95% "exact" confidence interval, the true confidence level is at least as large as .95, but is unknown. A more useful 95% confidence interval is the one containing the values θ_0 having mid P -values exceeding .05 in tests of $H_0: \theta = \theta_0$. Though mid P -based confidence intervals do not guarantee that the true confidence level is at least as large as the nominal one, they are narrower and usually more closely match the nominal level than either the exact or large-sample intervals. Both approaches are computationally complex and require software.

We illustrate with Table 3.4. Because the sample result is the most extreme possible, the Mantel-Haenszel estimator (3.2.2) of an assumed common odds ratio equals $\hat{\theta}_{MH} = 0.0$. StatXact reports an "exact" 95% confidence interval for a common odds ratio of (0, 1.01). We can be at least 95% confident that the true odds ratio falls in this interval. A 95% confidence interval based on correspondence with tests using the mid P -value is (0, 0.78). We can be approximately 95% confident that the odds of promotion for blacks are no more than 78% of the odds for whites.

3.3.4 Exact Test of Homogeneity of Odds Ratios

The Breslow-Day test of homogeneity of odds ratios (Section 3.2.4) is also a large-sample test. It applies when the sample size is relatively large in each partial table. When the total sample size is small, or when the total sample size is large but the number K of partial tables is large and individual tables have small sample sizes, this test is not valid. An exact test of homogeneity of odds ratios, sometimes called Zelen's exact test, handles such cases. The exact distribution is calculated using the set of all $2 \times 2 \times K$ tables that have the same two-way marginal totals as the observed table. The P -value is the sum of probabilities of all $2 \times 2 \times K$ tables that are no more likely than the observed table.

For Table 3.4, the values $\{\hat{\mu}_{i,jk}\}$ that yield the Mantel-Haenszel estimate $\hat{\theta}_{MH} = 0.0$ of the common odds ratio are identical to the observed counts in each partial table. The Breslow-Day statistic contains terms of the form 0/0, and it is undefined. In addition, no table other than the observed table has all the two-way marginal totals of that table. Therefore, Zelen's exact test is degenerate, giving $P = 1.0$. For this pattern of data, one cannot obtain any meaningful information about whether the true odds ratios differ.

A disadvantage of exact inference is that the small-sample conditional distribution is often highly discrete. In some cases, such as this one, it concentrates at only a single point. One can do only so much with small sample sizes!