

# CATEGORICAL DATA: RELATIONSHIPS

# Chapter 10

## Objectives

In this chapter we extend our study of categorical data to several populations. We will

- discuss independence and association for categorical variables.
- describe a chi-square test to assess the independence between two categorical variables.
- consider the conditions under which a chi-square test is valid.
- describe Fisher’s exact test of independence between two categorical variables.
- present McNemar’s test to analyze paired categorical data.
- calculate relative risk, the odds ratio, and its associated confidence interval.

## 10.1 Introduction

In Chapter 9 we considered the analysis of a single sample of categorical data. The basic techniques we employed were estimation of category probabilities and comparison of observed category frequencies with frequencies “expected” according to a null hypothesis. In this chapter we will extend these basic techniques to more complicated situations. To set the stage, here are two examples, the first of which presents an experiment; the second, an observational study.

### Example 10.1.1

**Migraine Headache** Patients who suffered from moderate to severe migraine headache took part in a double-blind clinical trial to assess an experimental surgery. A group of 75 patients were randomly assigned to receive either the real surgery on migraine trigger sites ( $n = 49$ ) or a sham surgery ( $n = 26$ ) in which an incision was made but no further procedure was performed. The surgeons hoped that patients would experience “a substantial reduction\* in migraine headaches,” which we will label as “success.” Table 10.1.1 shows the results of the experiment.<sup>1</sup>

		Surgery	
		Real	Sham
Substantial reduction in migraine headaches?	Success	41	15
	No success	8	11
	Total	49	26

\*“Substantial reduction” means at least a 50 percent reduction in migraine headache frequency, intensity, or duration when compared with baseline (presurgery) values.

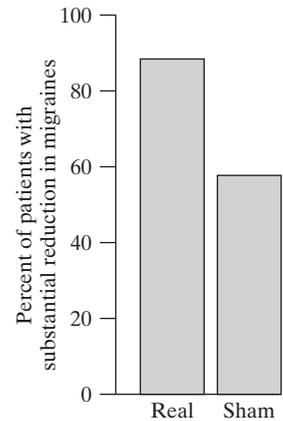
A natural way to express the results is in terms of percentages, as follows:

Of the real surgeries,  $\frac{41}{49}$  or 83.7% were successful.

Of sham surgeries,  $\frac{15}{26}$  or 57.7% were successful.

In this study successful reduction in migraine headache was more common among patients who received the real surgery than among those who received the sham surgery—83.7% versus 57.7%. Table 10.1.2 provides a summary of the data; Figure 10.1.1 is a bar chart showing the percentages of successful surgeries for the two groups. ■

	Surgery	
	Real	Sham
<i>n</i>	49	26
Success	41	15
Percent	83.7%	57.7%



**Figure 10.1.1** Bar chart of migraine surgery data

**Example 10.1.2**

**HIV Testing** A random sample of 120 college students found that 9 of the 61 women in the sample had taken an HIV test, compared to 8 of the 59 men.<sup>2</sup> These data are shown in Table 10.1.3.

	Female	Male
HIV test	9	8
No HIV test	52	51
Total	61	59

Of the women  $\frac{9}{61} = 0.148$  or 14.8% had been tested for HIV. Of the men  $\frac{8}{59} = 0.136$  or 13.6% had been tested for HIV.

These two percentages are nearly identical. ■

Tables such as Tables 10.1.1 and 10.1.3 are called **contingency tables**. The focus of interest in a contingency table is the dependence or association between the column variable and the row variable—for instance, between treatment and response in Tables 10.1.1 and 10.1.3. (The word *contingent* means “dependent.”) In particular, Tables 10.1.1 and 10.1.3 are called **2 × 2** (“two-by-two”) **contingency tables**, because they consist of two rows (excluding the “total” row) and two columns. Each category in the contingency table is called a **cell**; thus, a 2 × 2 contingency table has four cells.

We will consider the analysis and interpretation of  $2 \times 2$  contingency tables before extending the discussion to larger tables.

## 10.2 The Chi-Square Test for the $2 \times 2$ Contingency Table

When analyzing a  $2 \times 2$  contingency table it is natural to think of the probability of an event under either of two conditions being compared. We will find it is useful to extend the language of probability to include a new concept: conditional probability.\*

### Conditional Probability

Recall that the probability of an event predicts how often the event will occur. A **conditional probability** predicts how often an event will occur under specified conditions. The notation for a conditional probability is

$$\Pr\{E|C\}$$

which is read “probability of  $E$ , given  $C$ .” When a conditional probability is estimated from observed data, the estimate is denoted by a hat (“^”); thus,

$$\hat{\Pr}\{E|C\}$$

The following example illustrates these ideas.

#### Example 10.2.1

**Migraine Headache** Consider the migraine headache data from Example 10.1.1. The conditional probabilities of interest are as follows:

$$\begin{aligned} \Pr\{\text{substantial reduction in migraines}|\text{real surgery}\} &= \Pr\{\text{Success}|\text{Real}\} \\ &= \text{probability that a patient will have a substantial reduction in headache} \\ &\quad \text{if given the real surgery} \end{aligned}$$

$$\begin{aligned} \Pr\{\text{substantial reduction in migraines}|\text{sham surgery}\} &= \Pr\{\text{Success}|\text{Sham}\} \\ &= \text{probability that a patient will have a substantial reduction in headache} \\ &\quad \text{if given the sham surgery} \end{aligned}$$

The estimates of these conditional probabilities from the data of Table 10.1.1 are

$$\hat{\Pr}(\text{Success}|\text{Real}) = \frac{41}{49} = 0.837$$

and

$$\hat{\Pr}(\text{Success}|\text{Sham}) = \frac{15}{26} = 0.577$$

The natural hypothesis to test is that the conditional probabilities associated with a  $2 \times 2$  table are equal, which is to say that the probability of the event  $E$  does not depend on whether the first condition,  $C$ , is present or the second condition, “not  $C$ ,” is present.

$$H_0: \Pr\{E|C\} = \Pr\{E|\text{not } C\}$$

The following example illustrates this null hypothesis.

\*Conditional probability is also discussed in optional Section 3.3.

**Example  
10.2.2**

**Migraine Headache** For the migraine study of Example 10.1.1, the null hypothesis is

$$H_0: \Pr\{\text{Success}|\text{Real}\} = \Pr\{\text{Success}|\text{Sham}\}$$

or equivalently

$$H_0: \Pr\{\text{Success}|\text{Real}\} = \Pr\{\text{Success}|\text{not Real}\}$$

## The Chi-Square Statistic

Clearly, a natural way to test the preceding null hypothesis would be to reject  $H_0$  if  $\hat{\Pr}\{E|C\}$  and  $\hat{\Pr}\{E|\text{not } C\}$  are different by a sufficient amount. We describe a test procedure that compares  $\hat{\Pr}\{E|C\}$  and  $\hat{\Pr}\{E|\text{not } C\}$  indirectly, rather than directly. The procedure is a chi-square test, based on the test statistic  $\chi_s^2$  that was introduced in Section 9.4:

$$\chi_s^2 = \sum_{i=1}^4 \frac{(o_i - e_i)^2}{e_i}$$

In the formula, the sum is taken over all four cells in the contingency table. Each  $o$  represents an observed frequency and each  $e$  represents the corresponding expected frequency according to  $H_0$ . We now describe how to calculate the  $e$ 's.

The first step in determining the  $e$ 's for a contingency table is to calculate the row and column total frequencies (these are called the **marginal frequencies**) and also the grand total of all the cell frequencies. The  $e$ 's then follow from a simple rationale, as illustrated in Example 10.2.3.

**Example  
10.2.3**

**Migraine Headache** Table 10.2.1 shows the migraine data of Example 10.1.1, together with the marginal frequencies.

	Surgery		Total
	Real	Sham	
Success	41	15	56
No success	8	11	19
Total	49	26	75

The  $e$ 's should agree exactly with the null hypothesis. Because  $H_0$  asserts that the probability of success does not depend on the treatment, we can generate an estimate of this probability by pooling the two treatment groups; from Table 10.2.1, the pooled estimate, based on the marginal totals, is  $\frac{56}{75}$ . That is, if  $H_0$  is true, then the two columns “Real” and “Sham” are equivalent and we can pool them together. Our best estimate of  $\Pr\{\text{successful outcome}\}$  is then the pooled estimate  $\frac{56}{75}$ . We can then apply this estimate to each treatment group to yield the number of successful outcomes expected according to  $H_0$ , as follows:

$$\text{Real surgery group: } \frac{56}{75} \times 49 = 36.59 \text{ successful outcomes expected}$$

$$\text{Sham surgery group: } \frac{56}{75} \times 26 = 19.41 \text{ successful outcomes expected}$$

Likewise, the pooled estimate of  $\Pr\{\text{a surgery will not be successful}\}$  is  $\frac{19}{75}$ . Applying this probability to the two treatment groups gives

$$\text{Real surgery group: } \frac{19}{75} \times 49 = 12.41 \text{ unsuccessful outcomes expected}$$

$$\text{Sham surgery group: } \frac{19}{75} \times 26 = 6.59 \text{ unsuccessful outcomes expected}$$

The expected frequencies are shown in parentheses in Table 10.2.2. Note that the marginal totals for the  $e$ 's are the same as for the  $o$ 's. ■

	Surgery		Total
	Real	Sham	
Success	41 (36.59)	15 (19.41)	56
No success	8 (12.41)	11 (6.59)	19
Total	49	26	75

In practice, it is not necessary to proceed through a chain of reasoning to obtain the expected frequencies for a contingency table. The procedure for calculating the  $e$ 's can be condensed into a simple formula. The expected frequency for each cell is calculated from the marginal total frequencies for the same row and column, as follows:

**Expected Frequencies in a Contingency Table**

$$e = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Grand total}}$$

The formula produces the same calculation as does the rationale given in Example 10.2.2, as the following example shows.

**Example 10.2.4**

**Migraine Headache** We will apply the preceding formula to the migraine data of Example 10.1.1. The expected frequency of successful outcomes for the real surgery is calculated from the marginal totals as

$$e = \frac{56 \times 49}{75} = 36.59$$

Note that this is the same answer obtained in Example 10.2.2. Proceeding similarly for each cell in the contingency table, we would obtain all the  $e$ 's shown in Table 10.2.2. ■

**Note:** Although the formula for  $\chi^2$  for contingency tables is the same as given for goodness-of-fit tests in Section 9.4, the method of calculating the  $e$ 's is quite different for contingency tables because the null hypothesis is different.

### The Test Procedure

Other than the differences noted previously when computing expected counts, the chi-square test for a contingency table is carried out similarly to the chi-square goodness-of-fit test. Large values of  $\chi^2$  indicate evidence against  $H_0$ . Critical values

are determined from Table 9; the number of degrees of freedom for a  $2 \times 2$  contingency table is

$$df = 1$$

The chi-square test for a  $2 \times 2$  table has 1 degree of freedom because, in a sense, there only is one free cell in the table. Table 10.2.2 has four cells, but once we have determined that the expected cell frequency for the top-left cell is 36.59, the expected frequency for the top-right cell is constrained to be 19.41, since the top row adds across to a total of 56. Likewise, the bottom-left cell is constrained to be 12.41, since the left column adds down to a total of 49. Once these three cells are determined, the remaining cell, on the bottom right, is constrained as well. Thus, there are four cells in the table, but only one of them is “free”; once we have used the null hypothesis to determine the expected frequency for one of the cells, the other cells are constrained.

For a  $2 \times 2$  contingency table, the alternative hypothesis can be directional or nondirectional. Directional alternatives are handled by the familiar two-step procedure, cutting the nondirectional  $P$ -value in half if the data deviate from  $H_0$  in the direction specified by  $H_A$  (or reporting that the  $P$ -value is  $>0.50$  if the data deviate from  $H_0$  in the direction opposite to the direction specified by  $H_A$ ). Note that  $\chi_s^2$  itself does not express directionality; to determine the directionality of the data, one must calculate and compare the estimated probabilities.

The following example illustrates the chi-square test.

### Example 10.2.5

**Migraine Headache** For the migraine experiment of Example 10.1.1, let us apply a chi-square test. Given that the experiment involves cranial surgery, so that a Type I error would be quite serious, a conservative choice of  $\alpha$  is called for; we will use  $\alpha = 0.01$ . We may state the null hypothesis and a directional alternative informally as follows:

$H_0$ : The real surgery is no better than the sham surgery for reducing migraine headache.

$H_A$ : The real surgery is better than the sham surgery for reducing migraine headache.

Using the notation of conditional probability, the statements are

$$H_0: \Pr\{\text{Success} | \text{Real}\} = \Pr\{\text{Success} | \text{Sham}\}$$

$$H_A: \Pr\{\text{Success} | \text{Real}\} > \Pr\{\text{Success} | \text{Sham}\}$$

To check the directionality of the data, we calculate the estimated probabilities of response:

$$\hat{\Pr}\{\text{Success} | \text{Real}\} = \frac{41}{49} = 0.837$$

$$\hat{\Pr}\{\text{Success} | \text{Sham}\} = \frac{15}{26} = 0.577$$

and we note that

$$\hat{\Pr}\{\text{Success} | \text{Real}\} > \hat{\Pr}\{\text{Success} | \text{Sham}\}$$

Thus, the data do deviate from  $H_0$  in the direction specified by  $H_A$ . We proceed to calculate the chi-square statistic from Table 10.2.2 as

$$\begin{aligned} \chi_s^2 &= \frac{(41 - 36.59)^2}{36.59} + \frac{(15 - 19.41)^2}{19.41} + \frac{(8 - 12.41)^2}{12.41} + \frac{(11 - 6.59)^2}{6.59} \\ &= 6.06 \end{aligned}$$

From Table 9 with  $df = 1$ , we find that  $\chi_{1,0.02}^2 = 5.41$  and  $\chi_{1,0.01}^2 = 6.63$ , and so we have  $0.005 < P\text{-value} < 0.01$ . Thus, we reject  $H_0$  and find that the data provide

sufficient evidence to conclude that the real surgery is better than the sham surgery for reducing migraine headache.

Note that, even though  $\hat{\Pr}\{\text{Reduction}|\text{Real}\}$  and  $\hat{\Pr}\{\text{Reduction}|\text{Sham}\}$  do not enter into the calculation of  $\chi_s^2$ , the calculation of  $\hat{\Pr}\{\text{Reduction}|\text{Real}\}$  and  $\hat{\Pr}\{\text{Reduction}|\text{Sham}\}$  is an important part of the test procedure; the information provided by the quantities  $\hat{\Pr}\{\text{Success}|\text{Real}\}$  and  $\hat{\Pr}\{\text{Success}|\text{Sham}\}$  is essential for meaningful interpretation of the results.\*

**Computational Notes** The following tips are helpful in analyzing a  $2 \times 2$  contingency table:

1. The contingency table format is convenient for computations. For presenting the data in a report, however, it is usually better to use a more readable form of display such as Table 10.1.2; some additional examples are shown in the exercises.
2. For calculating  $\chi_s^2$ , the observed frequencies ( $o$ 's) must be *absolute*, rather than relative, frequencies; also, *the table must contain all four cells*, so that the sum of the  $o$ 's is equal to the total number of observations.

### Illustration of the Null Hypothesis

The chi-square statistic measures discrepancy between the data and the null hypothesis in an indirect way; the sample conditional probabilities are involved indirectly in the calculation of the expected frequencies. If sample conditional probabilities are equal, then the value of  $\chi_s^2$  is zero. Here is an example.

**Example 10.2.6**

**Fictitious Migraine Study** Table 10.2.3 shows fictitious data for a migraine study similar to that described in Example 10.1.1.

For the data of Table 10.2.3, the estimated probabilities of successful surgery are *equal*:

$$\hat{\Pr}\{\text{Success}|\text{Real}\} = \frac{30}{150} = 0.20$$

$$\hat{\Pr}\{\text{Success}|\text{Sham}\} = \frac{20}{100} = 0.20$$

You can easily verify that, for Table 10.2.3, the expected frequencies are equal to observed frequencies, so that the value of  $\chi_s^2$  is zero. Also notice that the columns of the table are proportional to each other:

$$\frac{30}{120} = \frac{20}{80}$$

	Surgery		Total
	Real	Sham	
Success	30	20	50
No success	120	80	200
Total	150	100	250

\*It is natural to wonder why we do not use a more direct comparison of  $\hat{\Pr}\{E|C\}$  and  $\hat{\Pr}\{E|\text{not } C\}$ . In fact, there is a test procedure based on a  $t$ -type statistic, calculated by dividing  $(\hat{\Pr}\{E|C\} - \hat{\Pr}\{E|\text{not } C\})$  by its standard error. This  $t$ -type procedure is equivalent to the chi-square test. We have chosen to present the chi-square test instead, for two reasons: (1) It can be extended to contingency tables larger than  $2 \times 2$ ; (2) in certain applications the chi-square statistic is more natural than the  $t$ -type statistic; some of these applications appear in Section 10.3.

As the preceding example suggests, an “eyeball” analysis of a contingency table is based on checking for proportionality of the columns. If the columns are nearly proportional, then the data agree fairly well with  $H_0$ ; if they are highly nonproportional, then the data disagree with  $H_0$ . The following example shows a case in which the data agree quite well with the expected frequencies under  $H_0$ .

**Example**  
**10.2.7**

**HIV Testing** The data from Example 10.1.2 show similar percentages of men and women who had been tested for HIV. The natural null hypothesis is that  $\Pr\{\text{HIV test}|\text{Female}\} = \Pr\{\text{HIV test}|\text{Male}\}$  and that the sample proportions differ only due to chance error in the sampling process. The expected frequencies are shown in parentheses in Table 10.2.4. The chi-square test statistic is  $\chi_s^2 = 0.035$ . From Table 8 with  $df = 1$ , we find that  $\chi_{1,0.20}^2 = 1.64$ . Thus, the  $P$ -value is greater than 0.20 (using a computer yields  $P$ -value = 0.85) and we do not reject the null hypothesis. Our conclusion is that the data provide no significant evidence that there is a difference in the rates which which men and women (at the college where the study was conducted) have been tested for HIV. ■

**Table 10.2.4** Observed and expected frequencies for HIV study

	Female	Male	Total
HIV test	9 (8.64)	8 (8.36)	17
No HIV test	52 (52.36)	51 (50.64)	103
Total	61	59	120

Note that the actual value of  $\chi_s^2$  depends on the sample sizes as well as the degree of nonproportionality; as discussed in Section 9.4, the value of  $\chi_s^2$  varies directly with the number of observations if the percentage composition of the data is kept fixed and the number of observations is varied. This reflects the fact that a given percentage deviation from  $H_0$  is less likely to occur by chance with a larger number of observations.

## Exercises 10.2.1–10.2.14

**10.2.1** The accompanying partially complete contingency table shows the responses to two treatments:

		TREATMENT	
		1	2
Response	Success	70	
	Failure		
Total		100	200

- (a) Invent a fictitious data set that agrees with the table and for which  $\chi_s^2 = 0$ .
- (b) Calculate the estimated probabilities of success ( $\Pr(\text{Success}|\text{Treatment 1})$  and  $\Pr(\text{Success}|\text{Treatment 2})$ ) a for your data set. Are they equal?

**10.2.2** Proceed as in Exercise 10.2.1 for the following contingency table:

		TREATMENT	
		1	2
Response	Success	30	
	Failure		
	Total	300	100

**10.2.3** Proceed as in Exercise 10.2.1 for the following contingency table:

		TREATMENT	
		1	2
Response	Success	5	20
	Failure	10	

**10.2.4** Most salamanders of the species *P. cinereus* are red striped, but some individuals are all red. The all-red form is thought to be a mimic of the salamander *N. viridescens*, which is toxic to birds. In order to test whether the mimic form actually survives more successfully, 163 striped and 41 red individuals of *P. cinereus* were exposed to predation by a natural bird population. After two hours, 65 of the striped and 23 of the red individuals were still alive.<sup>3</sup> Use a chi-square test to assess the evidence that the mimic form survives more successfully. Use a directional alternative and let  $\alpha = 0.05$ .

- (a) State the null hypothesis in words.
- (b) State the null hypothesis in symbols.
- (c) Compute the sample survival proportions for each group and display the values in a table similar to Table 10.1.2.
- (d) Find the value of the test statistic and the  $P$ -value.
- (e) State the conclusion of the test in the context of this setting.

**10.2.5** Can attack of a plant by one organism induce resistance to subsequent attack by a different organism? In a study of this question, individually potted cotton (*Gossypium*) plants were randomly allocated to two groups. Each plant in one group received an infestation of spider mites (*Tetranychus*); the other group were kept as controls. After two weeks the mites were removed and all plants were inoculated with *Verticillium*, a fungus that causes wilt disease. The accompanying table shows the numbers of plants that developed symptoms of wilt disease.<sup>4</sup> Do the data provide sufficient evidence to conclude that infestation with mites induces resistance to wilt disease? Use a chi-square test against a directional alternative following the five steps (a–e) in Problem 10.2.4. Let  $\alpha = 0.01$ .

		TREATMENT	
		MITES	NO MITES
Response	Wilt disease	11	17
	No wilt disease	15	4
	Total	26	21

**10.2.6** It has been suspected that prolonged use of a cellular telephone increases the chance of developing brain cancer due to the microwave-frequency signal that is transmitted by the cell phone. According to this theory, if a cell phone is repeatedly held near one side of the head, then brain tumors are more likely to develop on that side of the head. To investigate this, a group of patients were studied who had used cell phones for a least six months prior to developing brain tumors. The patients were asked whether they routinely held the cell phone to a certain ear and, if so, which ear. The 88 responses (from those who preferred one side over the other) are shown in the following table.<sup>5</sup> Do the data provide sufficient

evidence to conclude that use of cellular telephones leads to an increase in brain tumors on that side of the head? Use a chi-square test against a directional alternative following the five steps (a–e) in Problem 10.2.4. Let  $\alpha = 0.05$ .

		PHONE HOLDING SIDE	
		LEFT	RIGHT
Brain tumor side	Left	14	28
	Right	19	27
	Total	33	55

**10.2.7** Phenytoin is a standard anticonvulsant drug which unfortunately has many toxic side effects. A study was undertaken to compare phenytoin with valproate, another drug in the treatment of epilepsy. Patients were randomly allocated to receive either phenytoin or valproate for 12 months. Of 20 patients receiving valproate, 6 were free of seizures for the 12 months while 6 of 17 patients receiving phenytoin were seizure-free.<sup>6</sup>

- (a) Use a chi-square test to compare the seizure-free response rates for the two drugs. Let  $H_A$  be non-directional and  $\alpha = 0.10$  following the five steps (a–e) in Problem 10.2.4.
- (b) Do your conclusions in part (a) provide evidence that valproate and phenytoin are equally effective in preventing seizures? Discuss.

**10.2.8** Estrus synchronization products are used to bring cows into heat at a predictable time so that they can be reliably impregnated by artificial insemination. In a study of two estrus synchronization products, 42 mature cows (aged 4 to 8 years) were randomly allocated to receive either product A or product B, and then all cows were bred by artificial insemination. The table shows how many of the inseminations resulted in pregnancy.<sup>7</sup> Use a chi-square test to compare the effectiveness of the two products in producing pregnancy following the five steps (a–e) in Problem 10.2.4. Use a nondirectional alternative and let  $\alpha = 0.05$ .

	TREATMENT	
	PRODUCT A	PRODUCT B
Total number of cows	21	21
Number of cows pregnant	8	15

**10.2.9** Experimental studies of cancer often use strains of animals that have a naturally high incidence of tumors. In one such experiment, tumor-prone mice were kept in a sterile environment; one group of mice was maintained entirely germ free, while another group was exposed to the intestinal bacterium *Escherichia coli*.

The accompanying table shows the incidence of liver tumors.<sup>8</sup>

TREATMENT	TOTAL NUMBER OF MICE	MICE WITH LIVER TUMORS	
		NUMBER	PERCENT
Germ free	49	19	39%
<i>E. coli</i>	13	8	62%

- (a) How strong is the evidence that tumor incidence is higher in mice exposed to *E. coli*? Use a chi-square test against a directional alternative following the five steps (a–e) in Problem 10.2.4. Let  $\alpha = 0.05$ .
- (b) How would the result of part (a) change if the percentages (39% and 62%) of mice with tumors were the same, but the sample sizes were (i) doubled (98 and 26)? (ii) tripled (147 and 39)? [Hint: Part (b) requires almost no calculation.]

**10.2.10** In a randomized clinical trial to determine the most effective timing of administration of chemotherapeutic drugs to lung cancer patients, 16 patients were given four drugs simultaneously and 11 patients were given the same drugs sequentially. Objective response to the treatment (defined as shrinkage of the tumor by at least 50%) was observed in 11 of the patients treated simultaneously and in 3 of the patients treated sequentially.<sup>9</sup> Do the data provide evidence as to which timing is superior? Use a chi-square test against a nondirectional alternative following the five steps (a–e) in Problem 10.2.4. Let  $\alpha = 0.05$ .

**10.2.11** Physicians conducted an experiment to investigate the effectiveness of external hip protectors in preventing hip fractures in elderly people. They randomly assigned some people to get hip protectors and others to be the control group. They recorded the number of hip fractures in each group.<sup>10</sup> Do the data in the following table provide sufficient evidence to conclude that hip protectors reduce the likelihood of fracture? Use a chi-square test against a directional alternative following the five steps (a–e) in Problem 10.2.4. Let  $\alpha = 0.01$ .

Response		TREATMENT	
		HIP PROTECTOR	CONTROL
Hip fracture		13	67
	No hip fracture	640	1081
Total		653	1148

**10.2.12** A sample of 276 healthy adult volunteers were asked about the variety of social networks that they were in (e.g., relationships with parents, close neighbors, workmates, etc.). They were then given nasal drops containing a rhinovirus and were quarantined for five days. Of the

123 subjects who were in five or fewer types of social relationships 57 (46.3%) developed colds. Of 153 who were in at least six types of social relationships 52 (34.0%) developed colds.<sup>11</sup> Thus, the data suggest that having more types of social relationships helps one develop resistance to the common cold. Determine whether this difference is statistically significant. That is, use a chi-square test to test the null hypothesis that the probability of getting a cold does not depend on the number of social relationships a person is in following the five steps (a–e) in Problem 10.2.4. Use a nondirectional alternative and let  $\alpha = 0.05$ .

**10.2.13** The drug ancrod was tested in a double-blind clinical trial in which subjects who had strokes were randomly assigned to get either ancrod or a placebo. One response variable in the study was whether or not a subject experienced intracranial hemorrhaging.<sup>12</sup> The data are provided in the following table. Use a chi-square test to determine whether the difference in hemorrhaging rates is statistically significant following the five steps (a–e) in Problem 10.2.4. Use a nondirectional alternative and let  $\alpha = 0.05$ .

		TREATMENT	
		ANCROD	PLACEBO
Hemorrhage?	Yes	13	5
	No	235	247
	Total	248	252

**10.2.14** Do women respond to men's solicitations more readily during the fertile phase of their menstrual cycles? In a study of this question each of two hundred 18- to 25-year-old women who were walking alone in a city were approached by an attractive 20-year-old man who solicited the woman's telephone number. Previous research suggested that during the fertile phase of her menstrual cycle a woman would be more receptive to this kind of request than at other times. Of 60 women who were in the fertile phase of their cycles 13 gave out their phone numbers and 47 refused. The corresponding numbers for the 140 women not in the fertile phase of their cycles were 11 and 129.<sup>13</sup> The data are summarized in the following table. Use a chi-square test to determine whether the difference in success rates provides significant evidence in favor of an appropriate directional alternative, using  $\alpha = 0.02$  following the five steps (a–e) in Problem 10.2.4.

		PHASE	
		FERTILE	NOT
Success?	Yes	13	11
	No	47	129
	Total	60	140

## 10.3 Independence and Association in the $2 \times 2$ Contingency Table

The  $2 \times 2$  contingency table is deceptively simple. In this section we explore further the relationships that it can express.

### Two Contexts for Contingency Tables

A  $2 \times 2$  contingency table can arise in two contexts, namely:

1. Two independent samples with a dichotomous observed variable
2. One sample with two dichotomous observed variables

The first context is illustrated by the migraine data of Example 10.1.1, which can be viewed as two independent samples—the real surgery group and the sham surgery group—of sizes  $n_1 = 49$  and  $n_2 = 26$ . The observed variable is success (or failure) of the surgery. Any study involving a dichotomous observed variable and completely randomized allocation to two treatments can be viewed this way. The second context is illustrated by the HIV data of Example 10.1.2, which can be viewed as a single sample of  $n = 120$  students, observed with respect to two dichotomous variables—sex (male or female) and HIV test status (whether or not the student had been tested for HIV).

The two contexts—two samples with one variable or one sample with two variables—are not always sharply differentiated. For instance, the HIV data of Example 10.1.2 could have been collected in two samples—61 women and 59 men—observed with respect to one dichotomous variable (HIV test status).

The arithmetic of the chi-square test is the same in both contexts, but the statement and interpretation of hypotheses and conclusions can be very different.

### Independence and Association

In many contingency tables, the columns of the table play a different role than the rows. For instance, in the migraine data of Example 10.1.1, the columns represent treatments and the rows represent responses. Also, in Example 10.1.2 it seems more natural to define the columnwise conditional probabilities  $\Pr\{\text{HIV test}|\text{F}\}$  and  $\Pr\{\text{HIV test}|\text{M}\}$  rather than the rowwise conditional probabilities  $\Pr\{\text{F}|\text{HIV test}\}$  and  $\Pr\{\text{M}|\text{HIV test}\}$ .

On the other hand, in some cases it is natural to think of the rows and the columns of the contingency table as playing interchangeable roles. In such a case, conditional probabilities may be calculated either rowwise or columnwise, and the null hypothesis for the chi-square test may be expressed either rowwise or columnwise. The following is an example.

#### Example 10.3.1

**Hair Color and Eye Color** To study the relationship between hair color and eye color in a German population, an anthropologist observed a sample of 6,800 men, with the results shown in Table 10.3.1.<sup>14</sup>

		Hair color		Total
		Dark	Light	
Eye Color	Dark	726	131	857
	Light	3,129	2,814	5,943
Total		3,855	2,945	6,800

The data of Table 10.3.1 would be naturally viewed as a single sample of size  $n = 6,800$  with two dichotomous observed variables—hair color and eye color. To describe the data, let us denote dark and light eyes by DE and LE, and dark and light hair by DH and LH. We may calculate estimated columnwise conditional probabilities as follows:

$$\hat{\Pr}\{\text{DE}|\text{DH}\} = \frac{726}{3855} \approx 0.19$$

$$\hat{\Pr}\{\text{DE}|\text{LH}\} = \frac{131}{2945} \approx 0.04$$

A natural way to analyze the data is to compare these values: 0.19 versus 0.04. On the other hand, it is just as natural to calculate and compare estimated rowwise conditional probabilities:

$$\hat{\Pr}\{\text{DH}|\text{DE}\} = \frac{726}{857} \approx 0.85$$

$$\hat{\Pr}\{\text{DH}|\text{LE}\} = \frac{3129}{5943} \approx 0.53$$

Corresponding to these two views of the contingency table, the null hypothesis for the chi-square test can be stated columnwise as

$$H_0: \Pr\{\text{DE}|\text{DH}\} = \Pr\{\text{DE}|\text{LH}\}$$

or rowwise as

$$H_0: \Pr\{\text{DH}|\text{DE}\} = \Pr\{\text{DH}|\text{LE}\}$$

As we shall see, these two hypotheses are equivalent—that is, any population that satisfies one of them also satisfies the other. ■

When a data set is viewed as a single sample with two observed variables, the relationship expressed by  $H_0$  is called **statistical independence** of the row variable and the column variable. Variables that are not independent are called **dependent** or **associated**. Thus, the chi-square test is sometimes called a “test of independence” or a “test for association.”

### Example 10.3.2

**Hair Color and Eye Color** The null hypothesis of Example 10.3.1 can be stated verbally as

$$H_0: \text{Eye color is independent of hair color}$$

or

$$H_0: \text{Hair color is independent of eye color}$$

or, more symmetrically,

$$H_0: \text{Hair color and eye color are independent} \quad \blacksquare$$

The null hypothesis of independence can be stated generically as follows. Two groups,  $G_1$  and  $G_2$ , are to be compared with respect to the probability of a characteristic  $C$ . The null hypothesis is

$$H_0: \Pr\{C|G_1\} = \Pr\{C|G_2\}$$

Note that each of the two statements of  $H_0$  in Example 10.3.1 is of this form.

To further clarify the meaning of the null hypothesis of independence, in the following example we examine a data set that agrees *exactly* with  $H_0$ .

**Example**  
**10.3.3**

**Plant Height and Disease Resistance** Consider a (fictitious) species of plant that can be categorized as short (S) or tall (T) and as resistant (R) or nonresistant (NR) to a certain disease. Consider the following null hypothesis:

$H_0$ : Plant height and disease resistance are independent

Each of the following is a valid statement of  $H_0$ :

1.  $H_0$ :  $\Pr\{R|S\} = \Pr\{R|T\}$
2.  $H_0$ :  $\Pr\{NR|S\} = \Pr\{NR|T\}$
3.  $H_0$ :  $\Pr\{S|R\} = \Pr\{S|NR\}$
4.  $H_0$ :  $\Pr\{T|R\} = \Pr\{T|NR\}$

The following is not a statement of  $H_0$ :

5.  $H_0$ :  $\Pr\{R|S\} = \Pr\{NR|S\}$

Note the difference between statements 5 and 1. Statement 1 compares two groups (short and tall plants) with respect to disease resistance, whereas statement 5 is a statement about the distribution of disease resistance in only *one* group (short plants); statement 5 merely asserts that half (50%) of short plants are resistant and half are nonresistant.

Suppose, now, that we choose a random sample of 100 plants from the population and we obtain the data in Table 10.3.2.

		Height		Total
		S	T	
Resistance	R	12	18	30
	NR	28	42	70
Total		40	60	100

The data in Table 10.3.2 agree exactly with  $H_0$ ; this agreement can be checked in four different ways, corresponding to the four different symbolic statements of  $H_0$ :

1.  $\hat{\Pr}\{R|S\} = \hat{\Pr}\{R|T\}$   

$$\frac{12}{40} = 0.30 = \frac{18}{60}$$
2.  $\hat{\Pr}\{NR|S\} = \hat{\Pr}\{NR|T\}$   

$$\frac{28}{40} = 0.70 = \frac{42}{60}$$
3.  $\hat{\Pr}\{S|R\} = \hat{\Pr}\{S|NR\}$   

$$\frac{12}{30} = 0.40 = \frac{28}{70}$$

$$4. \hat{\Pr}\{T|R\} = \hat{\Pr}\{T|NR\}$$

$$\frac{18}{30} = 0.60 = \frac{42}{70}$$

Note that the data in Table 10.3.2 do *not* agree with statement 5:

$$\hat{\Pr}\{R|S\} = \frac{12}{40} = 0.30 \text{ and } \hat{\Pr}\{NR|S\} = \frac{28}{40} = 0.70$$

$$0.30 \neq 0.70$$

## Facts about Rows and Columns

The data in Table 10.3.2 display independence whether viewed rowwise or columnwise. This is no accident, as the following fact shows.

**Fact 10.3.1** The columns of a  $2 \times 2$  table are proportional if and only if the rows are proportional. Specifically, suppose that  $a$ ,  $b$ ,  $c$ , and  $d$  are any positive numbers, arranged as in Table 10.3.3.

Table 10.3.3 A general $2 \times 2$ contingency table		
		Total
	$a$ $b$	$a + b$
	$c$ $d$	$c + d$
Total	$a + c$ $b + d$	

Then

$$\frac{a}{c} = \frac{b}{d} \text{ if and only if } \frac{a}{b} = \frac{c}{d}$$

Another way to express this is

$$\frac{a}{a+c} = \frac{b}{b+d} \text{ if and only if } \frac{a}{a+b} = \frac{c}{c+d}$$

You can easily show that Fact 10.3.1 is true; just use simple algebra. Because of Fact 10.3.1, the relationship of independence in a  $2 \times 2$  contingency table is the same whether the table is viewed rowwise or columnwise. Note also that the expected frequencies, and therefore the value of  $\chi^2$ , would remain the same if the rows and columns of the contingency table were interchanged. The following fact shows that the *direction* of dependence is also the same whether viewed rowwise or columnwise.

**Fact 10.3.2** Suppose that  $a$ ,  $b$ ,  $c$ , and  $d$  are any positive numbers, arranged as in Table 10.3.3. Then

$$\frac{a}{a+c} > \frac{b}{b+d} \text{ if and only if } \frac{a}{a+b} > \frac{c}{c+d}$$

Also

$$\frac{a}{a+c} < \frac{b}{b+d} \text{ if and only if } \frac{a}{a+b} < \frac{c}{c+d}$$

**Note:** For more discussion of conditional probability and independence, see optional Section 3.3.

## Verbal Description of Association

Ideas of logical implication are expressed in everyday English in subtle ways. The following excerpt is from *Alice in Wonderland*, by Lewis Carroll:

“. . . you should say what you mean,” the March Hare went on.

“I do,” Alice hastily replied; “at least—at least I mean what I say—that’s the same thing, you know.”

“Not the same thing a bit!” said the Hatter. “Why, you might just as well say that ‘I see what I eat’ is the same thing as ‘I eat what I see!’”

. . . “You might just as well say,” added the Dormouse. . . , “That ‘I breathe when I sleep’ is the same thing as ‘I sleep when I breathe!’”

“It *is* the same thing with you,” said the Hatter. . .

We also use ordinary language to express ideas of probability, conditional probability, and association. For instance, consider the following four statements:

Color-blindness is more common among males than among females.

Maleness is more common among color-blind people than femaleness.

Most color-blind people are male.

Most males are color-blind.

The first three statements are all true; they are actually just different ways of saying the same thing. However, the last statement is false.<sup>15</sup>

In interpreting contingency tables, it is often necessary to describe probabilistic relationships in words. This can be quite a challenge. If you become fluent in such description, then you can always “say what you mean” *and* “mean what you say.” The following two examples illustrate some of the issues.

### Example 10.3.4

**Plant Height and Disease Resistance** For the plant height and disease resistance study of Example 10.3.3, we considered the null hypothesis

$H_0$ : Height and resistance are independent.

This hypothesis could also be expressed verbally in various other ways, such as

$H_0$ : Short and tall plants are equally likely to be resistant.

$H_0$ : Resistant and nonresistant plants are equally likely to be tall.

$H_0$ : Resistance is equally common among short and tall plants. ■

### Example 10.3.5

**Hair Color and Eye Color** Let us consider the interpretation of Table 10.3.1. The chi-square statistic is  $\chi_s^2 = 314$ ; from Table 9 we see that the  $P$ -value is tiny, so the null hypothesis of independence is overwhelmingly rejected. We might state our

conclusion in various ways. For instance, suppose we focus on the incidence of dark eyes. From the data we found that

$$\hat{\Pr}\{\text{DE}|\text{DH}\} > \hat{\Pr}\{\text{DE}|\text{LH}\}$$

that is,

$$\frac{726}{3855} = 0.19 > \frac{131}{2945} = 0.04$$

A natural conclusion from this comparison would be

Conclusion 1: There is sufficient evidence to conclude that dark-haired men have a greater tendency to be dark-eyed than do light-haired men.

This statement is carefully phrased, because the statement

“Dark-haired men have a greater tendency to be dark-eyed.”

is ambiguous by itself; it could mean

“Dark-haired men have a greater tendency to be dark-eyed than do light-haired men.”

or

“Dark-haired men have a greater tendency to be dark-eyed than to be light-eyed.”

The first of these statements says that

$$\hat{\Pr}\{\text{DE}|\text{DH}\} > \hat{\Pr}\{\text{DE}|\text{LH}\}$$

whereas the second says that

$$\hat{\Pr}\{\text{DE}|\text{DH}\} > \hat{\Pr}\{\text{LE}|\text{DH}\}$$

The second statement asserts that more than half of dark-haired men have dark eyes. Note that the data do not support this assertion; of the 3,855 dark-haired men, only 19% have dark eyes.

Conclusion 1 is only one of several possible wordings of the conclusion from the contingency table analysis. For instance, one might focus on dark hair and find

Conclusion 2: There is sufficient evidence to conclude that dark-eyed men have a greater tendency to be dark-haired than do light-eyed men.

A more symmetrical phrasing would be

Conclusion 3: There is sufficient evidence to conclude that dark hair is associated with dark eyes.

However, the phrasing in conclusion 3 is easily misinterpreted; it may suggest something like

“There is sufficient evidence to conclude that most dark-haired men are dark-eyed.”

which is not a correct interpretation. ■

We emphasize once again the principle that we stated in Section 10.2: *The calculation and comparison of appropriate conditional probabilities or  $\hat{\Pr}$ 's is an essential part of the chi-square test.* Example 10.3.5 provides ample illustration of this point.

Exercises 10.3.1–10.3.12

**10.3.1** Consider a fictitious population of mice. Each animal’s coat is either black (B) or grey (G) in color and is either wavy (W) or smooth (S) in texture. Express each of the following relationships in terms of probabilities or conditional probabilities relating to the population of animals.

- (a) Smooth coats are more common among black mice than among grey mice.
- (b) Smooth coats are more common among black mice than wavy coats are.
- (c) Smooth coats are more often black than are wavy coats.
- (d) Smooth coats are more often black than grey.
- (e) Smooth coats are more common than wavy coats.

**10.3.2** Consider a fictitious population of mice in which each animal’s coat is either black (B) or grey (G) in color and is either wavy (W) or smooth (S) in texture (as in Exercise 10.3.1). Suppose a random sample of mice is selected from the population and the coat color and texture are observed; consider the accompanying partially complete contingency table for the data.

		HEIGHT	
		B	G
Texture	W		50
	S		
Total		60	150

- (a) Invent fictitious data sets that agree with the table and for which
  - (i)  $\hat{\Pr}\{W|B\} > \hat{\Pr}\{W|G\}$ ; (ii)  $\hat{\Pr}\{W|B\} = \hat{\Pr}\{W|G\}$
 In each case, verify your answer by calculating the estimated conditional probabilities.
- (b) For each of the two data sets you invented in part (a), calculate  $\hat{\Pr}\{B|W\}$  and  $\hat{\Pr}\{B|S\}$ .
- (c) Which of the data sets of part (a) has  $\hat{\Pr}\{B|W\} > \hat{\Pr}\{B|S\}$ ? Can you invent a data set for which

$$\hat{\Pr}\{W|B\} > \hat{\Pr}\{W|G\} \text{ but } \hat{\Pr}\{B|W\} < \hat{\Pr}\{B|S\}$$

If so, do it. If not, explain why not.

**10.3.3** Men with prostate cancer were randomly assigned to undergo surgery ( $n = 347$ ) or “watchful waiting” (no surgery,  $n = 348$ ). Over the next several years there were 83 deaths in the first group and 106 deaths in the second group. The results are given in the table.<sup>16</sup>

		TREATMENT		TOTAL
		SURGERY	WW	
Survival	Died	83	106	189
	Alive	264	242	506
Total		347	348	695

- (a) Let D and A represent died and alive, respectively, and let S and WW represent surgery and watchful waiting. Estimate  $\Pr\{D|S\}$  and  $\Pr\{D|WW\}$ .
- (b) The value of the contingency-table chi-square statistic for these data is  $\chi_s^2 = 3.75$ . Test for a relationship between the treatment and survival. Use a nondirectional alternative and let  $\alpha = 0.05$ .

**10.3.4** In a study of behavioral asymmetries, 2,391 women were asked which hand they preferred to use (for instance, to write) and which foot they preferred to use (for instance, to kick a ball). The results are reported in the table.<sup>17</sup>

PREFERRED HAND	PREFERRED FOOT	NUMBER OF WOMEN
Right	Right	2,012
Right	Left	142
Left	Right	121
Left	Left	116
Total		2,391

- (a) Estimate the conditional probability that a woman is right-footed, given that she is right-handed.
- (b) Estimate the conditional probability that a woman is right-footed, given that she is left-handed.
- (c) Suppose we want to test the null hypothesis that hand preference and foot preference are independent. Calculate the chi-square statistic for this hypothesis.
- (d) Suppose we want to test the null hypothesis that right-handed women are equally likely to be right-footed or left-footed. Calculate the chi-square statistic for this hypothesis.

**10.3.5** Consider a study to investigate a certain suspected disease-causing agent. One thousand people are to be chosen at random from the population; each individual is to be classified as diseased or not diseased and as exposed or not exposed to the agent. The results are to be cast in the following contingency table:

		EXPOSURE	
		YES	NO
Disease	Yes		
	No		

Let EY and EN denote exposure and nonexposure and let DY and DN denote presence and absence of the disease. Express each of the following statements in terms of conditional probabilities. (Note that “a majority” means “more than half.”)

- (a) The disease is more common among exposed than among nonexposed people.
- (b) Exposure is more common among diseased people than among nondiseased people.
- (c) Exposure is more common among diseased people than is nonexposure.
- (d) A majority of diseased people are exposed.
- (e) A majority of exposed people are diseased.
- (f) Exposed people are more likely to be diseased than are nonexposed people.
- (g) Exposed people are more likely to be diseased than to be nondiseased.

**10.3.6** Refer to Exercise 10.3.5. Which of the statements express the assertion that occurrence of the disease is associated with exposure to the agent? (There may be more than one.)

**10.3.7** Refer to Exercise 10.3.5. Invent fictitious data sets as specified, and verify your answer by calculating appropriate estimated conditional probabilities. (Your data need not be statistically significant.)

- (a) Invent a data set for which

$$\hat{\Pr}\{DY|EY\} > \hat{\Pr}\{DY|EN\} \text{ but } \hat{\Pr}\{EY|DY\} < \hat{\Pr}\{EN|DY\}$$

or explain why it is not possible.

- (b) Invent a data set that agrees with statement (a) of Exercise 10.3.5 but with neither (d) nor (e); or, explain why it is not possible.
- (c) Invent a data set for which

$$\hat{\Pr}\{DY|EY\} > \hat{\Pr}\{DY|EN\} \text{ but } \hat{\Pr}\{EY|DY\} < \hat{\Pr}\{EY|DN\}$$

or explain why it is not possible.

**10.3.8** An ecologist studied the spatial distribution of tree species in a wooded area. From a total area of 21 acres, he randomly selected 144 quadrats (plots), each 38 feet square, and noted the presence or absence of maples and hickories in each quadrat. The results are shown in the table.<sup>18</sup>

		MAPLES	
		PRESENT	ABSENT
Hickories	Present	26	63
	Absent	29	26

The value of the chi-square statistic for this contingency table is  $\chi_s^2 = 7.96$ . Test the null hypothesis that the two species are distributed independently of each other. Use a nondirectional alternative and let  $\alpha = 0.01$ . In stating your conclusion, indicate whether the data suggest attraction between the species or repulsion. Support your interpretation with estimated conditional probabilities from the data.

**10.3.9** Refer to Exercise 10.3.8. Suppose the data for fictitious tree species, A and B, were as presented in the accompanying table. The value of the chi-square statistic for this contingency table is  $\chi_s^2 = 9.07$ . As in Exercise 10.3.9, test the null hypothesis of independence and interpret your conclusion in terms of attraction or repulsion between the species.

		SPECIES A	
		PRESENT	ABSENT
Species B	Present	30	10
	Absent	49	55

**10.3.10** A randomized experiment was conducted in which patients with coronary artery disease either had angioplasty or bypass surgery. The accompanying table shows the incidence of angina (chest pain) among the patients five years after treatment.<sup>19</sup>

		TREATMENT		
		ANGIOPLASTY	BYPASS	TOTAL
Angina?	Yes	111	74	185
	No	402	441	843
Total		513	515	1,028

Let A represent angioplasty and B represent bypass.

- (a) Calculate  $\hat{\Pr}\{\text{Yes}|A\}$  and  $\hat{\Pr}\{\text{Yes}|B\}$ .
- (b) Calculate  $\hat{\Pr}\{A|\text{Yes}\}$  and  $\hat{\Pr}\{A|\text{No}\}$ .

**10.3.11** Refer to Exercise 10.3.10. Invent a fictitious data set on coronary treatment and angina for 1,000 patients, for which  $\hat{\Pr}\{\text{Yes}|A\}$  is twice as great as  $\hat{\Pr}\{\text{Yes}|B\}$ , but nevertheless the majority of patients who have angina also had bypass surgery (as opposed to angioplasty).

**10.3.12** Suppose pairs of fraternal twins are examined and the handedness of each twin is determined; assume that all the twins are brother–sister pairs. Suppose data are collected for 1,000 twin pairs, with the results shown in the following table.<sup>20</sup> State whether each of the following statements is true or false.

- (a) Most of the brothers have the same handedness as their sisters.
- (b) Most of the sisters have the same handedness as their brothers.
- (c) Most of the twin pairs are either both right-handed or both left-handed.
- (d) Handedness of twin sister is independent of handedness of twin brother.
- (e) Most left-handed sisters have right-handed brothers.

		SISTER		TOTAL
		LEFT	RIGHT	
Brother	Left	15	85	100
	Right	135	765	900
Total		150	850	1,000

## 10.4 Fisher's Exact Test (Optional)

In this optional section we consider an alternative to the chi-square test for  $2 \times 2$  contingency tables. This procedure, known as **Fisher's exact test**, is particularly appropriate when dealing with small samples. Example 10.4.1 presents a situation in which Fisher's exact test can be used.

### Example 10.4.1

**ECMO** Extracorporeal membrane oxygenation (ECMO) is a potentially life-saving procedure that is used to treat newborn babies who suffer from severe respiratory failure. An experiment was conducted in which 29 babies were treated with ECMO and 10 babies were treated with conventional medical therapy (CMT). The data are shown in Table 10.4.1.<sup>21</sup>

		Treatment		Total
		CMT	ECMO	
Outcome	Die	4	1	5
	Live	6	28	34
Total		10	29	39

The data in Table 10.4.1 show that 34 of the 39 babies survived, but 5 of them died. The death rate was 40% for those given CMT and was 3.4% for those given ECMO. However, the sample sizes here are quite small. Is it possible that the difference in death rates happened simply by chance?

The null hypothesis of interest is that outcome (live or die) is independent of treatment (CMT or ECMO). If the null hypothesis is true, then we can think of the data in the following way: The two column headings of “CMT” and “ECMO” are arbitrary labels. Five of the babies would have died no matter which treatment group they were in; 4 of these babies ended up in the CMT group by chance.

The alternative hypothesis asserts that probability of death depends on treatment group. This means that there is a real difference between CMT and ECMO survival rates, which accounts for the sample percentages being different.

Thus, a question of interest is this: “If the null hypothesis is true, how likely is it to get a table of data like Table 10.4.1?” In conducting Fisher's exact test we find the probability that the observed table, Table 10.4.1, would arise by chance, given that the marginal totals—5 deaths and 34 survivors, 10 given CMT and 29 given ECMO—are fixed. To make this more concrete, suppose the null hypothesis is true and another experiment is conducted, with 10 babies given CMT and 29 given ECMO. Further, suppose that 5 of these 39 babies are going to die, no matter which group they are in. That is, there are 5 babies who are so seriously ill that neither treatment would be able to save them. What is the probability that 4 of them will be assigned to the CMT group?

To find this probability, we need to determine the following:

1. The number of ways of assigning exactly 4 of the 5 babies who are fated to die to the CMT group
2. The number of ways of assigning exactly 6 of the 34 babies who are going to survive to the CMT group
3. The number of ways of assigning 10 of the 39 babies to the CMT group

The product of (1) and (2), divided by (3), gives the probability in question. ■

## Combinations

In Section 3.6 we presented the binomial distribution formula. Part of that formula is the quantity  ${}_nC_j$  (which in Section 3.6 we called a binomial coefficient). The quantity  ${}_nC_j$  is the number of ways in which  $j$  objects can be chosen out of a set of  $n$  objects. For instance, the number of ways that a group of 4 babies can be chosen out of 5 babies is  ${}_5C_4$ . The numerical value of  ${}_nC_j$  is given by formula 10.4.1:

$${}_nC_j = \frac{n!}{j!(n-j)!} \quad (10.4.1)$$

where  $n!$  (“ $n$  factorial”) is defined for any positive integer as

$$n! = n(n-1)(n-2)\dots(2)(1)$$

and  $0! = 1$ .

For example, if  $j = 1$  then we have  ${}_nC_1 = \frac{n!}{1!(n-1)!} = n$ , which makes sense: There are  $n$  ways to choose 1 object from a set of  $n$  objects. If  $j = n$  then we have  ${}_nC_n = \frac{n!}{n!0!} = 1$ , since there is only one way to choose all  $n$  objects from a set of size  $n$ .

### Example 10.4.2

**ECMO** We can apply formula 10.4.1 as follows.

1. The number of ways of assigning 4 babies to the CMT group from among the 5 who are fated to die is  ${}_5C_4 = \frac{5!}{4!1!} = 5$ .
2. The number of ways of assigning 6 babies to the CMT group from among the 34 who are going to survive is  ${}_{34}C_6 = \frac{34!}{6!28!} = 1,344,904$ .
3. The number of ways of assigning 10 babies to the CMT group from among the 39 total babies is  ${}_{39}C_{10} = \frac{39!}{10!29!} = 635,745,396$ .\*

Thus, the probability of getting the same data as those in Table 10.4.1, given that the marginal totals are fixed, is  $\frac{{}_5C_4 \times {}_{34}C_6}{{}_{39}C_{10}} = \frac{5 \times 1,344,904}{635,745,396} = 0.01058$ . ■

When conducting Fisher’s exact test of a null hypothesis against a directional alternative, we need to find the probabilities of all tables of data (having the same margins as the observed table) that provide evidence as strongly against  $H_0$ , in the direction predicted by  $H_A$ , as the observed table.

### Example 10.4.3

**ECMO** Prior to this experiment described in Example 10.4.1, there was evidence that suggested that ECMO is better than CMT. Hence, a directional alternative hypothesis is appropriate:

$$H_A: \Pr\{\text{death}|\text{ECMO}\} < \Pr\{\text{death}|\text{CMT}\}$$

The data in the observed table, Table 10.4.1, support  $H_A$ . There is one other possible table, shown as Table 10.4.2, that has the same margins as Table 10.4.1 but which is even more extreme in supporting  $H_A$ . Given that 5 of 39 babies died and

\*It is evident from this example that a computer or a graphing calculator is a very handy tool when conducting Fisher’s exact test. This is a statistical procedure that is almost never carried out without the use of technology.

**Table 10.4.2** A more extreme table that could have resulted from the ECMO experiment

		Treatment		Total
		CMT	ECMO	
Outcome	Die	5	0	5
	Live	5	29	34
Total		10	29	39

that 10 babies were assigned to CMT, the most extreme possible result supporting the alternative hypothesis (that ECMO is better than CMT) is the table in which none of the ECMO babies die and all 5 deaths occur in the CMT group.

The probability of Table 10.4.2 occurring, if  $H_0$  is true, is  $\frac{{}_5C_5 \times {}_{34}C_0}{{}_{39}C_{10}} = \frac{1 \times 278256}{635745396} = 0.00044$ . The  $P$ -value is the probability of obtaining data at least as extreme as those observed, if  $H_0$  is true. In this case, the  $P$ -value is the probability of obtaining either the data in Table 10.4.1 or in Table 10.4.2, if  $H_0$  is true. Thus,  $P\text{-value} = 0.01058 + 0.00044 = 0.01102$ . This  $P$ -value is quite small, so the experiment provided strong evidence that  $H_0$  is false and that ECMO really is better than CMT. ■

## Comparison to the Chi-Square Test

The chi-square test presented in Section 10.2 is often used for analyzing  $2 \times 2$  contingency tables. One advantage of the chi-square test is that it can be extended to  $2 \times 3$  tables and other tables of larger dimension, as will be shown in Section 10.6. The  $P$ -value for the chi-square test is based on the chi-square distribution, as the name implies. It can be shown that as the sample size becomes large, this distribution provides a good approximation to the theoretical sampling distribution of the chi-square test statistic  $\chi_s^2$ . If the sample size is small, however, then the approximation can be poor and the  $P$ -value from the chi-square test can be misleading.

Fisher's exact test is called an "exact" test because the  $P$ -value is determined exactly, using calculations such as those shown in Example 10.4.2, rather than being based on an asymptotic approximation. Example 10.4.4 shows how the exact test and the chi-square test compare for the ECMO data.

### Example 10.4.4

**ECMO** Conducting a chi-square test on the ECMO experiment data in Table 10.4.1 gives a test statistic of

$$\begin{aligned}\chi_s^2 &= \frac{(4 - 1.28)^2}{1.28} + \frac{(1 - 3.72)^2}{3.72} + \frac{(6 - 8.72)^2}{8.72} + \frac{(28 - 25.28)^2}{25.28} \\ &= 8.89\end{aligned}$$

The  $P$ -value (using a directional alternative) is 0.0014. This is quite a bit smaller than the  $P$ -value found with the exact test of 0.01102. ■

## Nondirectional Alternatives and the Exact Test

Typically, the difference between a directional and a nondirectional test is that the  $P$ -value for the nondirectional test is twice the  $P$ -value for the directional test (assuming that the data deviate from  $H_0$  in the direction specified by  $H_A$ ).

For Fisher’s exact test this is not true. The  $P$ -value when  $H_A$  is nondirectional is not found by simply doubling the  $P$ -value from the directional test. Rather, a generally accepted procedure is to find the probabilities of all tables that are as likely or less likely than the observed table. These probabilities are added together to get the  $P$ -value for the nondirectional test.\* Example 10.4.5 illustrates this idea.

**Example 10.4.5**

**Flu Shots** A random sample of college students found that 13 of them had gotten a flu shot at the beginning of the winter and 28 had not. Of the 13 who had a flu shot, 3 got the flu during the winter. Of the 28 who did not get a flu shot, 15 got the flu.<sup>22</sup> These data are shown in Table 10.4.3. Consider the null hypothesis that the probability of getting the flu is the same whether or not one gets a flu shot. The probability of the data in Table 10.4.3, given that the margins are fixed, is  $\frac{18C_3 \times 23C_{10}}{41C_{13}} = 0.05298$ .

Table	Probability				
<table border="1"><tr><td>15</td><td>3</td></tr><tr><td>13</td><td>10</td></tr></table>	15	3	13	10	0.05298
15	3				
13	10				
<table border="1"><tr><td>16</td><td>2</td></tr><tr><td>12</td><td>11</td></tr></table>	16	2	12	11	0.01174
16	2				
12	11				
<table border="1"><tr><td>17</td><td>1</td></tr><tr><td>11</td><td>12</td></tr></table>	17	1	11	12	0.00138
17	1				
11	12				
<table border="1"><tr><td>18</td><td>0</td></tr><tr><td>10</td><td>13</td></tr></table>	18	0	10	13	0.00006
18	0				
10	13				

		No shot	Flu shot	Total
Flu?	Yes	15	3	18
	No	13	10	23
Total		28	13	41

**Figure 10.4.1**

Table	Probability				
<table border="1"><tr><td>5</td><td>13</td></tr><tr><td>23</td><td>0</td></tr></table>	5	13	23	0	0.00000
5	13				
23	0				
<table border="1"><tr><td>6</td><td>12</td></tr><tr><td>22</td><td>1</td></tr></table>	6	12	22	1	0.00002
6	12				
22	1				
<table border="1"><tr><td>7</td><td>11</td></tr><tr><td>21</td><td>2</td></tr></table>	7	11	21	2	0.00046
7	11				
21	2				
<table border="1"><tr><td>8</td><td>10</td></tr><tr><td>20</td><td>3</td></tr></table>	8	10	20	3	0.00440
8	10				
20	3				
<table border="1"><tr><td>9</td><td>9</td></tr><tr><td>19</td><td>4</td></tr></table>	9	9	19	4	0.02443
9	9				
19	4				
<table border="1"><tr><td>10</td><td>8</td></tr><tr><td>18</td><td>5</td></tr></table>	10	8	18	5	0.08356
10	8				
18	5				

**Figure 10.4.2**

A natural directional alternative would be that getting a flu shot reduces one’s chance of getting the flu. Figure 10.4.1 shows the obtained data (from Table 10.4.3) along with tables of possible outcomes that more strongly support  $H_A$ . The probability of each table is given in Figure 10.4.1, as well.

The  $P$ -value for the directional test is the sum of the probabilities of these tables:  $P$ -value =  $0.05298 + 0.01174 + 0.00138 + 0.00006 = 0.06616$ .

A nondirectional alternative states that the probability of getting the flu depends on whether or not one gets a flu shot but does not state whether a flu shot increases or decreases the probability. (Some people might get the flu *because* of the shot, so it is plausible that the overall flu rate is higher among people who get the shot than among those who don’t—although public health officials certainly hope otherwise!)

Figure 10.4.2 shows tables of possible outcomes for which the flu rate is higher among those who got the shot than among those who didn’t. The probability of each table is given, as well. The first five tables all have probabilities less than 0.05298, which is the probability of the observed data in Table 10.4.3, but the probability of the sixth table is greater than 0.05298. Thus, the contribution to the  $P$ -value from this set of tables is the sum of the first five probabilities:  $0.00000 + 0.00002 + 0.00046 + 0.00440 + 0.02443 = 0.02931$ . Adding this to the  $P$ -value for the directional test of 0.06616 gives the  $P$ -value for the nondirectional test:  $P$ -value =  $0.06616 + 0.02931 = 0.09547$ .

As this example shows, the calculation of a  $P$ -value for Fisher’s exact test is quite cumbersome, particularly when the alternative is nondirectional. It is highly recommended that statistics software be used to carry out the test. ■

\*There is not universal agreement on this process. The  $P$ -value can be taken to be the sum of the probabilities of all “extreme” tables, but there are several ways to define “extreme.” One alternative to the method presented here is to order tables according to the values of  $\chi^2$  and to count a table as extreme if it has a value of  $\chi^2$  that is at least as large as the  $\chi^2$  found from the observed table. Another approach is to order the tables according to  $|p_1 - p_2|$ . These methods will sometimes lead to a different  $P$ -value than the  $P$ -value being presented here.

## Exercises 10.4.1–10.4.8

**10.4.1** Consider conducting Fisher's exact test with the following fictitious table of data. Let the null hypothesis be that treatment and response are independent and let the alternative be the directional hypothesis that treatment B is better than treatment A. List the tables of possible outcomes that more strongly support  $H_A$ .

		TREATMENT		TOTAL
		A	B	
Outcome	Die	4	2	6
	Live	10	14	24
Total		14	16	30

**10.4.2** Repeat Exercise 10.4.1 with the following table of data.

		TREATMENT		TOTAL
		A	B	
Outcome	Die	5	3	8
	Live	12	13	25
Total		17	16	33

**10.4.3** In a randomized, double-blind clinical trial, 156 subjects were given an antidepressant medication to help them stop smoking; a second group of 153 subjects were given a placebo. Insomnia was more common in the antidepressant group than in the placebo group; Fisher's exact test of the insomnia data gave a  $P$ -value of 0.008.<sup>23</sup> Interpret this  $P$ -value in the context of the clinical trial.

**10.4.4 (Computer exercise)** A random sample of 99 students in a Conservatory of Music found that 9 of the 48 women sampled had "perfect pitch" (the ability to identify, without error, the pitch of a musical note), but only 1 of the 51 men sampled had perfect pitch.<sup>24</sup>

Conduct Fisher's exact test of the null hypothesis that having perfect pitch is independent of sex. Use a directional alternative and let  $\alpha = 0.05$ . Do you reject  $H_0$ ? Why or why not?

**10.4.5** Consider the data from Exercise 10.4.4. Conduct a chi-square test and compare the results of the chi-square test to the results of Fisher's exact test.

**10.4.6 (Computer exercise)** The growth factor pleiotrophin is associated with cancer progression in humans. In an attempt to monitor the growth of tumors, doctors measured serum pleiotrophin levels in patients with pancreatic cancer and in a control group of patients. They found that only 2 of 28 control patients had serum levels more than two standard deviations above the control group mean, whereas 20 of 41 cancer patients had serum levels this high.<sup>25</sup> Use Fisher's exact test to determine whether a discrepancy this large (2 of 28 versus 20 of 41) is likely to happen by chance. Use a directional alternative and let  $\alpha = 0.05$ .

**10.4.7 (Computer exercise)** A group of 225 men with benign prostatic hyperplasia were randomly assigned to take saw palmetto extract or a placebo in a double-blind trial. One year into the experiment 45 of the 112 men in the saw palmetto group (40%) thought they were taking saw palmetto, compared to 52 of the 113 men in the placebo group (46%).<sup>26</sup> Is this difference consistent with chance variation? Conduct Fisher's exact test using a nondirectional alternative.

**10.4.8 (Computer exercise)** An experiment involving subjects with schizophrenia compared "personal therapy" to "family therapy." Only 2 out of 23 subjects assigned to the personal therapy group suffered psychotic relapses in the first year of the study, compared to 8 of the 24 subjects assigned to the family therapy group.<sup>27</sup> Is this sufficient evidence to conclude, at the 0.05 level of significance, that the two types of therapies are not equally effective? Conduct Fisher's exact test using a nondirectional alternative.

## 10.5 The $r \times k$ Contingency Table

The ideas of Sections 10.2 and 10.3 extend readily to contingency tables that are larger than  $2 \times 2$ . We now consider a contingency table with  $r$  rows and  $k$  columns, which is termed an  $r \times k$  contingency table. Here is an example.

### Example 10.5.1

**Plover Nesting** Wildlife ecologists monitored the breeding habitats of mountain plovers for three years and made note of where the plovers nested. They found 66 nests on agricultural fields (AF), 67 nests in shortgrass prairie dog habitat (PD), and

20 nests on other grassland (G). The nesting choices varied across the years for these 153 sampled plover broods; Table 10.5.1 shows the data.<sup>28</sup>

**Table 10.5.1** Plover nest locations across three years

Location	Year			Total
	2004	2005	2006	
<b>Agricultural field (AF)</b>	21	19	26	<b>66</b>
<b>Prairie dog habitat (PD)</b>	17	38	12	<b>67</b>
<b>Grassland (G)</b>	5	6	9	<b>20</b>
<b>Total</b>	<b>43</b>	<b>63</b>	<b>47</b>	<b>153</b>

To compare the distributions in the three locations, we can calculate the columnwise percentages, as displayed in Table 10.5.2. (For instance, in the 2004 sample  $\frac{21}{43}$ , or 48.8%, of the nests were on agricultural fields.) Inspection of Table 10.5.2 shows some clear differences among the three percentage distributions (columns), with prairie dog habitat being much more common in 2005 than in the other years. ■

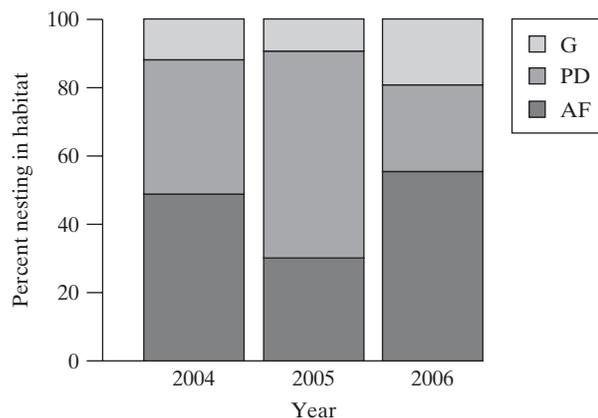
**Table 10.5.2** Percentage distributions of plover nests by year

Location	Year		
	2004	2005	2006
<b>Agricultural field (AF)</b>	48.8	30.2	55.3
<b>Prairie dog habitat (PD)</b>	39.5	60.3	25.5
<b>Grassland (G)</b>	11.6	9.5	19.1
<b>Total</b>	<b>99.9*</b>	<b>100.0</b>	<b>99.9*</b>

\*The sums of the 2004 and 2006 percentages differ from 100% due to rounding.

Figure 10.5.1 is a bar chart of the data, which gives a visual impression of the distributions.

**Figure 10.5.1** Stacked relative frequency (percentage) chart of plover nest locations by year



## The Chi-Square Test for the $r \times k$ Table

The goal of statistical analysis of an  $r \times k$  contingency table is to investigate the relationship between the row variable and the column variable. Such an investigation can begin with an inspection of the columnwise or rowwise percentages, as in Table 10.5.2. One route to further analysis is to ask whether the discrepancies in percentages are too large to be explained as sampling error. This question can be answered by a chi-square test. The chi-square statistic is calculated from the familiar formula

$$\sum_{\text{all cells}} \frac{(o_i - e_i)^2}{e_i}$$

where the sum is over all  $I = r \times k$  cells of the contingency table, and the expected frequencies ( $e$ 's) are calculated as

$$e = \frac{(\text{Row total}) \times (\text{Column total})}{\text{Grand total}}$$

This method of calculating the  $e$ 's can be justified by a simple extension of the rationale given in Section 10.2. Critical values for the chi-square test are obtained from Table 9 with

$$\text{df} = (r - 1)(k - 1)$$

The following example illustrates the chi-square test.

### Example 10.5.2

**Plover Nesting** Let us apply the chi-square test to the plover nesting data of Example 10.5.1. The null hypothesis is

$H_0$ : The population distributions of nest locations are the same in the three years.

This hypothesis can be stated symbolically in conditional probability notation as follows:

$$H_0: \left\{ \begin{array}{l} \Pr \{AF|2004\} = \Pr \{AF|2005\} = \Pr \{AF|2006\} \\ \Pr \{PD|2004\} = \Pr \{PD|2005\} = \Pr \{PD|2006\} \\ \Pr \{G|2004\} = \Pr \{G|2005\} = \Pr \{G|2006\} \end{array} \right\}$$

Note that the percentages in Table 10.5.2 are the estimated conditional probabilities; that is,

$$\hat{\Pr}\{AF|2004\} = 0.488$$

$$\hat{\Pr}\{AF|2005\} = 0.302$$

and so on. We test  $H_0$  against the nondirectional alternative hypothesis

$H_A$ : The population distributions of nest locations are not the same in all three years.

Table 10.5.3 shows the observed and expected frequencies.

**Table 10.5.3** Observed and expected frequencies of plover nests

Location	Year			Total
	2004	2005	2006	
<b>Agricultural field (AF)</b>	21 (18.55)	19 (21.18)	26 (20.27)	<b>66</b>
<b>Prairie dog habitat (PD)</b>	17 (18.83)	38 (27.59)	12 (20.58)	<b>67</b>
<b>Grassland (G)</b>	5 (5.62)	6 (8.24)	9 (6.14)	<b>20</b>
<b>Total</b>	<b>43</b>	<b>63</b>	<b>47</b>	<b>153</b>

From Table 10.5.3, we can calculate the test statistic as

$$\begin{aligned}\chi_s^2 &= \frac{(21 - 18.55)^2}{18.55} + \frac{(19 - 21.18)^2}{21.18} + \dots + \frac{(9 - 6.14)^2}{6.14} \\ &= 14.09\end{aligned}$$

For these data,  $r = 3$  and  $k = 3$ , so that

$$\text{df} = (3 - 1)(3 - 1) = 4$$

From Table 9 with  $\text{df} = 4$ , we find that  $\chi_{4,0.01}^2 = 13.28$  and  $\chi_{4,0.001}^2 = 18.47$ , and so we have  $0.001 < P\text{-value} < 0.01$ . Thus, the chi-square test shows that there is significant evidence that the nesting location preferences differed across the three years. ■

Note that  $H_0$  in Example 10.5.2 is a compound null hypothesis in the sense defined in Section 9.4—that is,  $H_0$  contains more than one independent assertion. This will always be true for contingency tables larger than  $2 \times 2$ , and consequently for such tables the alternative hypothesis for the chi-square test will always be nondirectional and the conclusion, if  $H_0$  is rejected, will be nondirectional. Thus, the chi-square test will often not represent a complete analysis of an  $r \times k$  contingency table.

## Two Contexts for $r \times k$ Contingency Tables

We noted in Section 10.3 that a  $2 \times 2$  contingency table can arise in two different contexts. Similarly, an  $r \times k$  contingency table can arise in the following two contexts:

1.  $k$  independent samples; a categorical observed variable with  $r$  categories
2. One sample; two categorical observed variables—one with  $k$  categories and one with  $r$  categories

As with the  $2 \times 2$  table, the calculation of the chi-square statistic is the same for both contexts, but the statement of hypotheses and conclusions can differ. The following example illustrates the second context.

### Example 10.5.3

**Hair Color and Eye Color** Table 10.5.4 shows the relationship between hair color and eye color for 6,800 German men.<sup>29</sup> (This is the same study as in Example 10.3.2.)

Let us use a chi square test to test the hypothesis

$$H_0: \text{Hair color and eye color are independent.}$$

		Hair color			
		Brown	Black	Fair	Red
Eye Color	Brown	438	288	115	16
	Grey or Green	1,387	746	946	53
	Blue	807	189	1,768	47

For the data of Table 10.5.4, one can calculate  $\chi_s^2 = 1,074$ . The degrees of freedom for the test are  $df = (3 - 1)(4 - 1) = 6$ . From Table 9 we find  $\chi_{6,0.0001}^2 = 27.86$ . Thus,  $H_0$  is overwhelmingly rejected and we conclude that there is extremely strong evidence that hair color and eye color are associated. ■

### Exercises 10.5.1–10.5.8

**10.5.1** Patients with painful knee osteoarthritis were randomly assigned in a clinical trial to one of five treatments: glucosamine, chondroitin, both, placebo, or Celebrex, the standard therapy. One outcome recorded was whether or not each patient experienced substantial improvement in pain or in ability to function. The data are given in the following table.<sup>30</sup>

TREATMENT	SUCCESSFUL OUTCOME		
	SAMPLE SIZE	NUMBER	PERCENT
Glucosamine	317	192	60.6
Chondroitin	318	202	63.5
Both	317	208	65.6
Placebo	313	178	56.9
Celebrex	318	214	67.3

- (a) Use a chi-square test to compare the success rates at  $\alpha = 0.05$ . (The value of the chi-square statistic is  $\chi_s^2 = 9.29$ .)  
 (b) Verify the value of  $\chi_s^2$  given in part (a).

**10.5.2** For a study of free-living populations of the fruitfly *Drosophila subobscura*, researchers placed baited traps in two woodland sites and one open-ground area. The numbers of male and female flies trapped in a single day are given in the table.<sup>31</sup>

	WOODLAND	WOODLAND	OPEN GROUND
	SITE I	SITE II	
Males	89	34	74
Females	31	20	136
Total	120	54	210

- (a) Use a chi-square test to compare the sex ratios at the three sites. Let  $\alpha = 0.05$ .  
 (b) Construct a table that displays the data in a more readable format, such as the one in Exercise 10.5.1.

**10.5.3** In a classic study of peptic ulcer, blood types were determined for 1,655 ulcer patients. The accompanying table shows the data for these patients and for an independently chosen group of 10,000 healthy controls from the same city.<sup>32</sup>

BLOOD TYPE	ULCER PATIENTS	CONTROLS
O	911	4,578
A	579	4,219
B	124	890
AB	41	313
Total	1,655	10,000

- (a) The value of the chi-square statistic for this contingency table is  $\chi_s^2 = 49.0$ . Carry out the chi-square test at  $\alpha = 0.01$ .  
 (b) Construct a table showing the percentage distributions of blood type for patients and for controls.  
 (c) Verify the value of  $\chi_s^2$  given in part (a).

**10.5.4** The two claws of the lobster (*Homarus americanus*) are identical in the juvenile stages. By adulthood, however, the two claws normally have differentiated into a stout claw called a “crusher” and a slender claw called a “cutter.” In a study of the differentiation process, 26 juvenile animals were reared in smooth plastic trays and 18 were reared in trays containing oyster chips (which they could use to exercise their claws). Another 23 animals were reared in trays containing only one oyster chip. The claw configurations of all the animals as adults are summarized in the table.<sup>33</sup>

TREATMENT	CLAW CONFIGURATION		
	RIGHT CRUSHER, LEFT CUTTER	RIGHT CUTTER, LEFT CRUSHER	RIGHT CUTTER, LEFT CUTTER
Oyster chips	8	9	1
Smooth plastic	2	4	20
One oyster chip	7	9	7

- (a) The value of the contingency-table chi-square statistic for these data is  $\chi_s^2 = 24.35$ . Carry out the chi-square test at  $\alpha = 0.01$ .
- (b) Verify the value of  $\chi_s^2$  given in part (a).
- (c) Construct a table showing the percentage distribution of claw configurations for each of the three treatments.
- (d) Interpret the table from part (c): In what way is claw configuration related to treatment? (For example, if you wanted a lobster with two cutter claws, which treatment would you choose and why?)

**10.5.5** A randomized, double-blind, placebo-controlled experiment was conducted in which patients with Alzheimer’s disease were given either extract of Ginkgo biloba (EGb) or a placebo for one year. The change in each patient’s Alzheimer’s Disease Assessment Scale—Cognitive subscale (ADAS-Cog) score was measured. The results are given in the table.<sup>34</sup> (Note: If the ADAS-Cog went down, then the patient improved.)

	CHANGE IN ADAS-Cog SCORE				
	-4 OR BETTER	-2 TO -3	-1 TO +1	+2 TO +3	+4 OR WORSE
EGb	22	18	12	7	16
Placebo	10	11	19	11	24

- (a) Use a chi-square test to compare the prevalence rates at  $\alpha = 0.05$ . (The value of the chi-square statistic is  $\chi_s^2 = 10.26$ .)
- (b) Verify the value of  $\chi_s^2$  given in part (a).

**10.5.6** Marine biologists have noticed that the color of the outermost growth band on a clam tends to be related to the time of the year in which the clam dies. A biologist conducted a small investigation of whether this is true for the species *Protothaca staminea*. She collected a sample of 78 clam shells from this species and cross-classified them according to (1) month when the clam died and (2) color of the outermost growth band. The data are shown in the following table.<sup>35</sup>

	COLOR		
	CLEAR	DARK	UNREADABLE
February	9	26	9
March	6	25	3
Total	15	51	12

Use a chi-square test to compare the color distributions for the two months. Let  $\alpha = 0.10$ .

**10.5.7** A group of patients with a binge-eating disorder were randomly assigned to take either the experimental drug fluvoxamine or a placebo in a nine-week long double-blind clinical trial. At the end of the trial the condition of each patient was classified into one of four categories: no response, moderate response, marked response, or remission. The following table shows a cross classification of the data.<sup>36</sup> Is there statistically significant evidence, at the 0.10 level, to conclude that there is an association between treatment group (fluvoxamine versus placebo) and condition?

	NO RESPONSE	MODERATE RESPONSE	MARKED RESPONSE	REMISSION	TOTAL
	Fluvoxamine	15	7	3	
Placebo	22	7	3	11	43
Total	37	14	6	26	

**10.5.8** Patients with coronary artery disease were randomly assigned to either receive angioplasty plus medical therapy ( $n = 1149$ ) or medical therapy alone ( $n = 1138$ ) in a clinical trial. Over the next several years 85 angioplasty and 95 medical therapy patients died, with cause

of death categorized as cardiac, other, or unknown. The following table shows a cross classification of the data.<sup>37</sup> Is there statistically significant evidence, at the 0.10 level, to conclude that there is an association between treatment group (angioplasty versus medical therapy) and outcome?

	CARDIAC DEATH	OTHER DEATH	UNKNOWN CAUSE OF DEATH	ALIVE	TOTAL
Angioplasty	23	45	17	1,064	1,149
Medical therapy	25	51	19	1,043	1,138
Total	48	96	36	2,107	2,287

## 10.6 Applicability of Methods

In this section we discuss guidelines for deciding when to use a chi-square test.

### Conditions for Validity

A chi-square test is valid under the following conditions:

1. *Design conditions* For the contingency-table chi-square test, it must be appropriate to view the data in one of the following ways:
  - (a) As two or more independent random samples, observed with respect to a categorical variable; or
  - (b) As one random sample, observed with respect to two categorical variables.

For either type of chi-square test, the observations within a sample must be independent of each other.

2. *Sample size conditions* The sample size must be large enough. The critical values given in Table 9 are only approximately correct for determining the  $P$ -value associated with  $\chi^2_s$ . As a rule of thumb, the approximation is considered adequate if each expected frequency ( $e$ ) is at least equal to 5.\* (If the expected frequencies are small and the data form a  $2 \times 2$  contingency table, then Fisher's exact test might be appropriate—see optional Section 10.4.)
3. *Form of  $H_0$*  A generic form of the null hypothesis for the contingency-table chi-square test may be stated as follows:

$H_0$ : The row variable and the column variable are independent.

4. *Scope of inference* As with other statistical tests, if the data arise from an experiment with random assignment of treatments, as in Example 10.1.1, then we can draw a causal inference; if the experimental units were drawn at random from a population, then we can extend the causal inference to that population. However, if the data arise from an observational study, as in Example 10.1.2, then a small  $P$ -value only allows us to infer that the observed association is not due to chance, but we cannot rule out other explanations.

\*For an  $r \times k$  table with more than 2 rows and columns, the approximation is adequate if the average expected frequency is at least 5, even if some of the cell counts are smaller.

## Verification of Design Conditions

To verify the design conditions, we need to identify a population from which the data may be viewed as a random sample. If the data consist of several samples [situation 1(a)], then the samples are required to be independent of each other. Failure to observe this restriction may result in a loss of power. If the design includes any pairing or matching of experimental units, then the samples would not be independent. A method of analysis for dependent samples is described in Section 10.8.

As always, bias in the sampling procedure must be ruled out. Moreover, chi-square methods are not appropriate when complex random sampling schemes such as cluster sampling or stratified random sampling are used. Finally, there must be no dependency or hierarchical structure in the design. Failure to observe this restriction can result in a vastly inflated chance of Type I error (which is usually much more serious than a loss of power). The following examples show the relevance of checking for dependency in the observations.

### Example 10.6.1

**Food Choice by Insect Larvae** In a behavioral study of the clover root curculio *Sitona hispidulus*, 20 larvae were released into each of six petri dishes. Each dish contained nodulated and nonnodulated alfalfa roots, arranged in a symmetric pattern. (This experiment was more fully described in Example 1.1.5.) After 24 hours the location of each larva was noted, with the results shown in Table 10.6.1.<sup>38</sup>

Dish	Number of larvae		
	Nodulated Roots	Nonnodulated Roots	Other (died, lost, etc.)
1	5	3	12
2	9	1	10
3	6	3	11
4	7	1	12
5	5	1	14
6	14	3	3
Total	46	12	62

Suppose the following analysis is proposed. A total of 58 larvae made a choice; the observed frequencies of choosing nodulated and nonnodulated roots were 46 and 12, and the corresponding expected frequencies (assuming random choice) would be 29 and 29; these data yield  $\chi_s^2 = 19.93$ , from which (using a directional alternative) we find from Table 9 that  $P\text{-value} < 0.00005$ . The validity of this proposed analysis is highly doubtful because it depends on the assumption that all the observations in a given dish are independent of each other; this assumption would certainly be false if (as is biologically plausible) the larvae tend to follow each other in their search for food.

How, then, should the data be analyzed? One approach is to make the reasonable assumption that the observations in one dish are independent of those in another dish. Under this assumption one could use a paired analysis on the six dishes ( $n_d = 6$ ); a paired  $t$  test yields  $P\text{-value} \approx 0.005$  and a sign test yields  $P\text{-value} \approx 0.02$ . Note that the questionable assumption of independence within dishes led to a  $P\text{-value}$  that was much too small. ■

**Example**  
**10.6.2**

**Pollination of Flowers** A study was conducted to determine the adaptive significance of flower color in the scarlet gilia (*Ipomopsis aggregata*). Six red-flowered plants and six white-flowered plants were chosen for observation in field conditions; hummingbirds were permitted to visit the flowers, but the other major pollinator, a moth, was excluded by covering the plants at night. Table 10.6.2 shows, for each plant, the total number of flowers at the end of the season and the number that had set fruit.<sup>39</sup>

**Table 10.6.2** Fruit set in scarlet gilia flowers

	Red-flowered plants			White-flowered plants		
	Number of flowers	Number setting fruit	Percent setting fruit	Number of flowers	Number setting fruit	Percent setting fruit
	140	26	19	125	21	17
	116	11	9	134	17	13
	34	0	0	273	81	30
	79	9	11	146	38	26
	185	28	15	103	17	17
	<u>106</u>	<u>11</u>	10	<u>82</u>	<u>24</u>	29
Sum	660	85		863	198	

The question of interest is whether the percentage of fruit set is different for red-flowered than for white-flowered plants. Suppose this question is approached by regarding the individual flower as the observational unit; then the data could be cast in the contingency table format of Table 10.6.3.

**Table 10.6.3** Fruit set in scarlet gilia flowers

		Flower color	
		Red	White
Fruit set	Yes	85	198
	No	575	665
Total		660	863
	Percent setting fruit	13	23

Table 10.6.3 yields  $\chi_s^2 = 25.0$ , for which Table 9 gives  $P$ -value  $< 0.0001$ . However, this analysis is not correct, because the observations on flowers on the same plant are not independent of each other; they are dependent because the pollinator (the hummingbird) tends to visit flowers in groups, and perhaps also because the flowers on the same plant are physiologically and genetically related. The chi-square test is invalidated by the hierarchical structure in the data.

A better approach would be to treat the entire plant as the observational unit. For instance, one could take the “Percentage Setting Fruit” column of Table 10.6.2 as the basic observations; applying a  $t$  test to the values yields  $t_s = 2.88$  (with  $0.01 < P$ -value  $< 0.02$ ), and applying a Wilcoxon-Mann-Whitney test yields  $U_s = 32$  (with  $0.02 < P$ -value  $< 0.05$ ). Thus, the  $P$ -value from the inappropriate chi-square analysis is much too small. ■

## Power Considerations

In many studies the chi-square test is valid but is not as powerful as a more appropriate test. Specifically, consider a situation in which the rows or the columns (or both) of the contingency table correspond to a *rankable* categorical variable with more than two categories. The following is an example.

### Example 10.6.3

**Physiotherapy** A randomized clinical trial was conducted to determine whether the addition of Saturday physiotherapy sessions (the “treatment”) to the usual Monday–Friday sessions (the “control”) would benefit patients undergoing rehabilitation in a hospital. One outcome measure was the destination of a patient upon being discharged, with the categories being home, low-level residential care (LLRC), high-level residential care (HLRC), or acute hospital transfer (AHT). The results are shown in Table 10.6.4.<sup>40</sup>

		Group	
		Treatment	Control
Discharge destination	Home	107	103
	LLRC	10	15
	HLRC	6	1
	AHT	7	13
	Total	130	132

A contingency-table chi-square test would be valid to compare treatment and control, but the test would lack power because it does not use the information contained in the *ordering* of the discharge destination categories (home is preferred to LLRC, which is preferred to HLRC, which is preferred to AHT). A related weakness of the chi-square test is that, even if  $H_0$  is rejected, the test does not yield a directional conclusion such as “the treatment leads to better discharge destinations than does the control.”

Methods are available to analyze contingency tables with rankable row and/or column variables; such methods, however, are beyond the scope of this book.

## Exercises 10.6.1–10.6.3

**10.6.1** Refer to the chemotherapy data of Exercise 10.2.10. Are the sample sizes large enough for the approximate validity of the chi-square test?

**10.6.2** In a study of prenatal influences on susceptibility to seizures in mice, pregnant females were randomly allocated to a control group or a “handled” group. Handled

mice were given sham injections three times during gestation, while control mice were not touched. The offspring were tested for their susceptibility to seizures induced by a loud noise. The investigators noted that the response varied considerably from litter to litter. The accompanying table summarizes the results.<sup>41</sup>

TREATMENT	NUMBER OF LITTERS	NUMBER OF MICE	RESPONSE TO LOUD NOISE		
			NO RESPONSE	WILD RUNNING	SEIZURE
Handled	19	104	23	10	71
Control	20	120	47	13	60

If these data are analyzed as a  $2 \times 3$  contingency table, the chi-square statistic is  $\chi_s^2 = 8.45$  and Table 9 gives  $0.01 < P\text{-value} < 0.02$ . Is this an appropriate analysis for this experiment? Explain. (*Hint*: Does the design meet the conditions for validity of the chi-square test?)

**10.6.3** In control of diabetes it is important to know how blood glucose levels change after eating various foods. Ten volunteers participated in a study to compare the effects of two foods—a sugar and a starch. A blood specimen was drawn before each volunteer consumed a measured amount of food; then additional blood specimens were drawn at 11 times during the next four hours. Each volunteer repeated the entire test on another occasion with the other food. Of particular concern were blood glucose levels that dropped below the initial level; the accompanying table shows the number of such values.<sup>42</sup>

FOOD	NO. OF VALUES LESS THAN INITIAL VALUE	TOTAL NUMBER OF OBSERVATIONS
Sugar	26	110
Starch	14	110

Suppose we analyze the given data as a contingency table. The test statistic would be

$$\chi_s^2 = \frac{(26 - 20)^2}{20} + \frac{(14 - 20)^2}{20} + \frac{(84 - 90)^2}{90} + \frac{(96 - 90)^2}{90} = 4.40$$

At  $\alpha = 0.05$  we would reject  $H_0$  and find that there is sufficient evidence to conclude that blood glucose values below the initial value occur more often after ingestion of sugar than after ingestion of starch. This analysis contains two flaws. What are they? (*Hint*: Are the conditions for validity of the test satisfied?)

## 10.7 Confidence Interval for Difference between Probabilities

The chi-square test for a  $2 \times 2$  contingency table answers only a limited question: Do the estimated probabilities—call them  $\hat{p}_1$  and  $\hat{p}_2$ —differ enough to conclude that the true probabilities—call them  $p_1$  and  $p_2$ —are not equal? A complementary mode of analysis is to construct a confidence interval for the magnitude of the difference,  $(p_1 - p_2)$ .

When we discussed constructing a confidence interval for a single proportion,  $p$ , in Section 9.2, we defined an estimate  $\tilde{p}$ , based on the idea of “adding 2 successes and 2 failures to the data.” Making this adjustment to the data resulted in a confidence interval procedure that has good coverage properties. Likewise, when constructing a confidence interval for the difference in two proportions, we will define new estimates that are based on the idea of adding 1 observation to each cell of the table (so that a *total* of 2 successes and 2 failures are added to the data).

Consider a  $2 \times 2$  contingency table that can be viewed as a comparison of two samples, of sizes  $n_1$  and  $n_2$ , with respect to a dichotomous response variable. Let the  $2 \times 2$  table be given as

Sample 1	Sample 2
$y_1$	$y_2$
$n_1 - y_1$	$n_2 - y_2$
$n_1$	$n_2$

We define

$$\tilde{p}_1 = \frac{y_1 + 1}{n_1 + 2}$$

and

$$\tilde{p}_2 = \frac{y_2 + 1}{n_2 + 2}$$

We will use the difference in the new values,  $(\tilde{p}_1 - \tilde{p}_2)$ , to construct a confidence interval for  $(p_1 - p_2)$ . Like all quantities calculated from samples, the quantity  $(\tilde{P}_1 - \tilde{P}_2)$  is subject to sampling error. The magnitude of the sampling error can be expressed by the standard error of  $(\tilde{P}_1 - \tilde{P}_2)$ , which is calculated from the following formula:

$$SE_{(\tilde{P}_1 - \tilde{P}_2)} = \sqrt{\frac{\tilde{p}_1(1 - \tilde{p}_1)}{n_1 + 2} + \frac{\tilde{p}_2(1 - \tilde{p}_2)}{n_2 + 2}}$$

Note that  $SE_{(\tilde{P}_1 - \tilde{P}_2)}$  is analogous to  $SE_{(\bar{Y}_1 - \bar{Y}_2)}$  as described in Section 6.6.

An approximate confidence interval can be based on  $SE_{(\tilde{P}_1 - \tilde{P}_2)}$ ; for instance, a 95% confidence interval is

$$(\tilde{p}_1 - \tilde{p}_2) \pm (1.96) SE_{(\tilde{P}_1 - \tilde{P}_2)}$$

Confidence intervals constructed this way have good coverage properties (i.e., approximately 95% of all 95% confidence intervals cover the true difference  $p_1 - p_2$ ) for almost any sample sizes  $n_1$  and  $n_2$ .<sup>43</sup> The following example illustrates the construction of the confidence interval.\*

### Example 10.7.1

**Migraine Headache** For the migraine headache data of Example 10.1.1, the sample sizes are  $n_1 = 49$  and  $n_2 = 26$ , and the estimated probabilities of substantial reduction in migraines are

$$\tilde{p}_1 = \frac{42}{51} = 0.824$$

$$\tilde{p}_2 = \frac{16}{28} = 0.571$$

The difference between these is

$$\begin{aligned} \tilde{p}_1 - \tilde{p}_2 &= 0.824 - 0.571 \\ &= 0.253 \\ &\approx 0.25 \end{aligned}$$

Thus, we estimate that the real surgery increases the probability of substantial reduction in migraines by 0.25, compared to the sham surgery. To set confidence limits on this estimate, we calculate the standard error as

$$\begin{aligned} SE_{(\tilde{P}_1 - \tilde{P}_2)} &= \sqrt{\frac{0.824(0.176)}{51} + \frac{0.571(0.429)}{28}} \\ &= 0.1077 \end{aligned}$$

\*In Section 9.3 we presented a general version of the “add 2 successes and 2 failures” idea, in which the formula for  $\tilde{p}$  depends on the confidence level (95%, 90%, etc.). When constructing a confidence interval for a difference in two proportions, the coverage properties of the interval are best when 1 is added to each cell in the  $2 \times 2$  table, no matter what confidence level is being used.<sup>44</sup>

The 95% confidence interval is

$$0.253 \pm (1.96)(0.1077)$$

$$0.253 \pm 0.211$$

$$0.042 < p_1 - p_2 < 0.464$$

We are 95% confident that the probability of substantial reduction in migraines is between 0.042 and 0.464 higher with the real surgery than with the sham surgery. ■

**Relationship to Test** The chi-square test for a  $2 \times 2$  contingency table (Section 10.2) is approximately, but not exactly, equivalent to checking whether a confidence interval for  $(p_1 - p_2)$  includes zero. [Recall from Section 7.3 that there is an exact equivalence between a  $t$  test and a confidence interval for  $(\mu_1 - \mu_2)$ .]

## Exercises 10.7.1–10.7.6

**10.7.1** Elderly patients who had suffered hip fractures were randomly assigned to receive either a placebo ( $n = 1,062$ ) or zolendronic acid ( $n = 1,065$ ) in a double-blind clinical trial. During the trial 139 placebo patients and 92 zolendronic acid patients had new fractures.<sup>45</sup> Let  $p_1$  and  $p_2$  represent the probabilities of fracture on placebo and zolendronic acid, respectively. Construct a 95% confidence interval for  $(p_1 - p_2)$ .

**10.7.2** Refer to the liver tumor data of Exercise 10.2.9.

- Construct a 95% confidence interval for  $(\Pr\{\text{liver tumor} \mid \text{germ-free}\} - \Pr\{\text{liver tumor} \mid E. coli\})$ .
- Interpret the confidence interval from part (a). That is, explain what the interval tells you about tumor probabilities.

**10.7.3** For women who are pregnant with twins, complete bed rest in late pregnancy is commonly prescribed in order to reduce the risk of premature delivery. To test the value of this practice, 212 women with twin pregnancies were randomly allocated to a bed-rest group or a control group. The accompanying table shows the incidence of preterm delivery (less than 37 weeks of gestation).<sup>46</sup>

	BED REST	CONTROLS
No. of preterm deliveries	32	20
No. of women	105	107

Construct a 95% confidence interval for  $(\Pr\{\text{preterm} \mid \text{bed rest}\} - \Pr\{\text{preterm} \mid \text{contro}\})$ . Does the confidence interval suggest that bed rest is beneficial?

**10.7.4** Refer to Exercise 10.7.3. The numbers of infants with low birthweight (2,500 gm or less) born to the women are shown in the table.

	BED REST	CONTROLS
No. of low-birthweight babies	76	92
Total no. of babies	210	214

Let  $p_1$  and  $p_2$  represent the probabilities of a low-birthweight baby in the two conditions. Explain why the above information is not sufficient to construct a confidence interval for  $(p_1 - p_2)$ .

**10.7.5** Refer to the blood type data of Exercise 10.5.3. Let  $p_1$  and  $p_2$  represent the probabilities of type O blood in the patient population and the control population, respectively.

- Construct a 95% confidence interval for  $(p_1 - p_2)$ .
- Interpret the confidence interval from part (a). That is, explain what the interval tells you about the difference in probabilities of type O blood.

**10.7.6** In an experiment to treat patients with “generalized anxiety disorder,” the drug hydroxyzine was given to 71 patients and 30 of them improved. A group of 70 patients were given a placebo and 20 of them improved.<sup>47</sup> Let  $p_1$  and  $p_2$  represent the probabilities of improvement using hydroxyzine and the placebo, respectively. Construct a 95% confidence interval for  $(p_1 - p_2)$ .

## 10.8 Paired Data and $2 \times 2$ Tables (Optional)

In Chapter 8 we considered paired data when the response variable is continuous. In this section we consider the analysis of paired categorical data.

### Example 10.8.1

**HIV Transmission to Children** A study was conducted to determine a woman's risk of transmitting HIV to her unborn child. A sample of 114 HIV-infected women who gave birth to two children found that HIV infection occurred in 19 of the 114 older siblings and in 20 of the 114 younger siblings.<sup>48</sup> These data are shown in Table 10.8.1.

		Older sibling	Younger sibling
HIV?	Yes	19	20
	No	95	94
Total		114	114

At first glance, it might appear that a regular chi-square test could be used to test the null hypothesis that the probability of HIV infection is the same for older siblings as for younger siblings. However, as we stated in Section 10.6, for the chi-square test to be valid the two samples—of 114 older siblings and of 114 younger siblings—must be independent of each other. In this case the samples are clearly dependent. Indeed, these are paired data, with a family generating the pair (older sibling, younger sibling).

Table 10.8.2 presents the data in a different format. This format helps focus attention on the relevant part of the data.\*

		Younger sibling HIV?	
		Yes	No
Older sibling	Yes	2	17
HIV?	No	18	77

From Table 10.8.2 we can see that there are 79 pairs in which both siblings have the same HIV status: Two are “yes/yes” pairs and 77 are “no/no” pairs. These 79 pairs, which are called **concordant pairs**, do not help us determine whether HIV infection is more likely for younger siblings than for older siblings. The remaining 35 pairs—17 “yes/no” pairs and 18 “no/yes” pairs—do provide information on the relative likelihood of HIV infection for older and younger siblings. These pairs are called **discordant pairs**; we will focus on these 35 pairs in our analysis.

\*Note that Table 10.8.2 cannot be derived from Table 10.8.1.

If the chance of HIV infection is the same for older siblings as it is for younger siblings, then the two kinds of pairs—“yes/no” and “no/yes”—are equally likely. Thus, the null hypothesis

$H_0$ : the probability of HIV infection is the same for older siblings as it is for younger siblings

is equivalent to

$$H_0: \text{among discordant pairs, } \Pr(\text{“yes/no”}) = \Pr(\text{“no/yes”}) = \frac{1}{2} \quad \blacksquare$$

### McNemar’s Test

The hypothesis that discordant pairs are equally likely to be “yes/no” or “no/yes” can be tested with the chi-square goodness-of-fit test developed in Section 9.4. This application of the goodness-of-fit test is known as **McNemar’s test** and has a particularly simple form.\* Let  $n_{11}$  denote the number of “yes/yes” pairs,  $n_{12}$  the number of “yes/no” pairs,  $n_{21}$  the number of “no/yes” pairs, and  $n_{22}$  the number of “no/no” pairs, as shown in Table 10.8.3. If  $H_0$  is true, the expected number of “yes/no” pairs is  $\frac{n_{12} + n_{21}}{2}$ , as is the expected number of “no/yes” pairs. Thus, the test statistic is

$$\chi_s^2 = \frac{\left(n_{12} - \frac{(n_{12} + n_{21})}{2}\right)^2}{\frac{(n_{12} + n_{21})}{2}} + \frac{\left(n_{21} - \frac{(n_{12} + n_{21})}{2}\right)^2}{\frac{(n_{12} + n_{21})}{2}}$$

which simplifies to

$$\chi_s^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

The distribution of  $\chi_s^2$  under the null hypothesis is approximately a  $\chi^2$  distribution with 1 degree of freedom.

<b>Table 10.8.3</b> A general table of paired proportion data		
	Yes	No
Yes	$n_{11}$	$n_{12}$
No	$n_{21}$	$n_{22}$

\*The null hypothesis tested by McNemar’s test can also be tested by using the binomial distribution. The null hypothesis states that among discordant pairs,  $\Pr(\text{“yes/no”}) = \Pr(\text{“no/yes”}) = 1/2$ . Thus, under the null hypothesis, the number of “yes/no” pairs has a binomial distribution with  $n$  = the number of discordant pairs and  $p = 0.5$ .

**Example 10.8.2** **HIV Transmission to Children** For the data given in Example 10.8.1,  $n_{12} = 17$  and  $n_{21} = 18$ . Thus,

$$\chi_s^2 = \frac{(17 - 18)^2}{17 + 18} = 0.0286$$

From Table 9 we see that the  $P$ -value is greater than 0.20. (Using a computer gives  $P$ -value = 0.87.) The data are very much consistent with the null hypothesis that the probability of HIV infection is the same for older siblings as it is for younger siblings. ■

### Exercises 10.8.1–10.8.4

**10.8.1** As part of a study of risk factors for stroke, 155 women who had experienced a hemorrhagic stroke (cases) were interviewed. For each case, a control was chosen who had not experienced a stroke; the control was matched to the case by neighborhood of residence, age, and race. Each woman was asked whether she used oral contraceptives. The data for the 155 pairs are displayed in the table. “Yes” and “No” refer to use of oral contraceptives.<sup>49</sup>

		CASE	
		NO	YES
Control	No	107	30
	Yes	13	5

To test for association between oral contraceptive use and stroke, consider only the 43 discordant pairs (pairs who answered differently) and test the hypothesis that a discordant pair is equally likely to be “yes/no” or “no/yes.” Use McNemar’s test to test the hypothesis that having a stroke is independent of use of oral contraceptives against a nondirectional alternative at  $\alpha = 0.05$ .

**10.8.2** Example 10.8.1 referred to a sample of HIV-infected women who gave birth to two children. One of the outcomes that was studied was whether the gestational age of the child was less than 38 weeks; this information was recorded for 106 of the families. The data for this variable are shown in the following table. Analyze these data using McNemar’s test. Use a nondirectional alternative and let  $\alpha = 0.10$ .

		YOUNGER SIBLING < 38 WEEKS?	
		YES	NO
Older sibling	Yes	26	5
< 38 weeks?	No	21	54

**10.8.3** A study of 85 patients with Hodgkin’s disease found that 41 had had their tonsils removed. Each patient was matched with a sibling of the same sex. Only 33 of the siblings had undergone tonsillectomy. The data are shown in the following table.<sup>50</sup> Use McNemar’s test to test the hypothesis that “yes/no” and “no/yes” pairs are equally likely. Previous research had suggested that having a tonsillectomy is associated with an increased risk of Hodgkin’s disease; thus, use a directional alternative. Let  $\alpha = 0.05$ .

		SIBLING TONSILLECTOMY?	
		YES	NO
Hodgkin’s patient tonsillectomy	Yes	26	15
	No	7	37

**10.8.4** In a study of the mating behavior of *Gryllus campestris*, pairs of female crickets were placed in a plexiglass arena with a single male cricket. There were 54 cases in which the females fought; these resulted in 42 cases in which the winning female copulated with the male, 8 cases in which the losing female copulated with the male, and 4 cases that ended with no copulation. The data are summarized in the following table.<sup>51</sup> Use McNemar’s test to test the hypothesis that winners and losers are equally likely to copulate. Use an appropriate directional alternative and let  $\alpha = 0.05$ .

		WINNERS	
		YES	NO
Losers	COPULATE		
	Yes	0	8
	No	42	4

## 10.9 Relative Risk and the Odds Ratio (Optional)

It is quite common to test the null hypothesis that two population proportions,  $p_1$  and  $p_2$ , are equal. A chi-square test, based on a  $2 \times 2$  table, is often used for this purpose. A confidence interval for  $(p_1 - p_2)$  provides information about the magnitude of the difference between  $p_1$  and  $p_2$ . In this section we consider two other measures of dependence: the relative risk and the odds ratio.

### Relative Risk

Sometimes researchers prefer to compare probabilities in terms of their *ratio*, rather than their difference. When the outcome event is deleterious (such as having a heart attack or getting cancer) the ratio of probabilities is called the **relative risk**, or the risk ratio. The relative risk is defined as  $p_1/p_2$ . This measure is widely used in studies of human health. The following is an example.

#### Example 10.9.1

**Smoking and Lung Cancer** The health histories of 11,900 middle-aged men were tracked over many years. During the study 126 of the men developed lung cancer, including 89 men who were smokers and 37 men who were former smokers. Table 10.9.1 shows the data.<sup>52</sup>

		Smoking history	
		Smoker	Former smoker
Lung cancer?	Yes	89	37
	No	6,063	5,711
	Total	6,152	5,748

The probabilities of primary interest are the columnwise conditional probabilities:

$$p_1 = \Pr\{\text{lung cancer} \mid \text{smoker}\}$$

$$p_2 = \Pr\{\text{lung cancer} \mid \text{former smoker}\}$$

The estimates of these from the data are

$$\hat{p}_1 = \frac{89}{6152} = 0.01447 \approx 0.014$$

$$\hat{p}_2 = \frac{37}{5748} = 0.00644 \approx 0.006$$

The estimated relative risk is

$$\frac{\hat{p}_1}{\hat{p}_2} = \frac{0.01447}{0.00644} = 2.247 \approx 2.2$$

Thus, we estimate that the risk (i.e., the conditional probability) of developing lung cancer is about 2.2 times as great for smokers as for former smokers. (Of course, because this is an observational study, we would not be justified in concluding that smoking *causes* lung cancer.)

## The Odds Ratio

Another way to compare two probabilities is in terms of **odds**. The odds of an event  $E$  is defined to be the ratio of the probability that  $E$  occurs to the probability that  $E$  does not occur:

$$\text{odds of } E = \frac{\Pr\{E\}}{1 - \Pr\{E\}}$$

For instance, if the probability of an event is  $1/4$ , then the odds of the event are  $\frac{1/4}{3/4} = 1/3$  or 1:3. As another example, if the probability of an event is  $1/2$ , then the odds of the event are  $\frac{1/2}{1/2} = 1$  or 1:1.

The **odds ratio** is simply the ratio of odds under two conditions. Specifically, suppose that  $p_1$  and  $p_2$  are the conditional probabilities of an event under two different conditions. Then the odds ratio, which we will denote by  $\theta$  (“theta”), is defined as follows:

$$\theta = \frac{\frac{p_1}{1 - p_1}}{\frac{p_2}{1 - p_2}}$$

If the estimated probabilities  $\hat{p}_1$  and  $\hat{p}_2$  are calculated from a  $2 \times 2$  contingency table, the corresponding estimated odds ratio, denoted  $\hat{\theta}$ , is calculated as

$$\hat{\theta} = \frac{\frac{\hat{p}_1}{1 - \hat{p}_1}}{\frac{\hat{p}_2}{1 - \hat{p}_2}}$$

We illustrate with an example.

### Example 10.9.2

**Smoking and Lung Cancer** From the data of Example 10.9.1, we estimate the odds of developing lung cancer as follows:

$$\hat{\text{odds}} = \frac{0.01447}{1 - 0.01447} = 0.01468 \text{ among smokers}$$

$$\hat{\text{odds}} = \frac{0.00644}{1 - 0.00644} = 0.00648 \text{ among former smokers}$$

The estimated odds ratio is

$$\hat{\theta} = \frac{0.01468}{0.00648} = 2.265 \approx 2.3$$

Thus, we estimate that the odds of developing lung cancer are about 2.3 times as great for smokers as for former smokers. ■

## Odds Ratio and Relative Risk

The odds ratio measures association in an unfamiliar way; the relative risk is a more natural measure. Fortunately, in many applications the two measures are approximately equal. In general the relationship between the odds ratio and the relative risk is given by

$$\text{odds ratio} = \text{relative risk} \times \frac{1 - p_2}{1 - p_1}$$

Notice that if  $p_1$  and  $p_2$  are small, then the relative risk is approximately equal to the odds ratio. We illustrate with the smoking and lung cancer data.

**Example 10.9.3**

**Smoking and Lung Cancer** For the data in Table 10.9.1 we found that the estimated relative risk of lung cancer is

$$\text{estimated relative risk} = 2.247$$

and the estimated odds ratio is

$$\hat{\theta} = 2.265$$

These are approximately equal because the outcome of interest (developing lung cancer) is rare, so that  $\hat{p}_1$  and  $\hat{p}_2$  are small. ■

### Advantage of the Odds Ratio

Both the relative risk  $p_1/p_2$  and the difference  $(p_1 - p_2)$  are easier to interpret than the odds ratio. Why, then, is the odds ratio used at all? One important advantage of the odds ratio is that, in certain kinds of studies, the odds ratio can be estimated even though  $p_1$  and  $p_2$  cannot be estimated. To explain this property, we must first discuss the question of estimability of conditional probabilities in contingency tables.

In a  $2 \times 2$  contingency table, the conditional probabilities can be defined by rows or by columns. Whether these probabilities can be estimated from the observed data depends on the study design. The following example illustrates this point.

**Example 10.9.4**

**Smoking and Lung Cancer** In studying the relationship between smoking and lung cancer, the conditional probabilities of primary interest are

$$p_1 = \Pr\{\text{lung cancer} \mid \text{smoker}\}$$

and

$$p_2 = \Pr\{\text{lung cancer} \mid \text{former smoker}\}$$

These are columnwise probabilities in a table like Table 10.9.1. One could, however, also consider the following rowwise conditional probabilities:

$$p_1^* = \Pr\{\text{smoker} \mid \text{lung cancer}\}$$

and

$$p_2^* = \Pr\{\text{smoker} \mid \text{no lung cancer}\}$$

(Of course,  $p_1^*$  and  $p_2^*$  are not particularly meaningful biologically.) From the study described in Example 10.9.1, that is, a single sample of size  $n = 11,900$  observed with respect to smoking status and lung cancer, one can estimate not only  $p_1$  and  $p_2$  but also  $p_1^*$  and  $p_2^*$ . However, there are other important study designs that do not provide enough information to estimate all these conditional probabilities. For example, suppose that a study is conducted by choosing a group of 500 smokers and a group of 500 former smokers and then observing how many of them develop lung cancer. This kind of study is called a prospective study

or **cohort study**. Such a study might produce the fictitious but realistic data of Table 10.9.2.

		Smoking history	
		Smoker	Former smoker
Lung cancer?	Yes	7	3
	No	473	497
	Total	500	500

The data of Table 10.9.2 can be viewed as two independent samples. From the data we can estimate the conditional probabilities of lung cancer in the two populations (smokers and former smokers):

$$\hat{p}_1 = \frac{7}{500} = 0.014 \quad \hat{p}_2 = \frac{3}{500} = 0.006$$

By contrast, the rowwise probabilities  $p_1^*$  and  $p_2^*$  cannot be estimated from Table 10.9.2. Because the relative numbers of smokers and former smokers were predetermined by the design of the study ( $n_1 = 500$  and  $n_2 = 500$ ), the data contain no information about the prevalence of smoking, and therefore no information about the population values of

$\Pr\{\text{smoker} \mid \text{lung cancer}\}$  and  $\Pr\{\text{smoker} \mid \text{no lung cancer}\}$

Table 10.9.2 was generated by fixing the column totals and observing the row variable. Consider now the reverse sort of design. Suppose we choose 500 men who died from lung cancer and 500 men who did not die from lung cancer and we then determine the smoking histories of the men. This design is called a **case-control design**. Such a design might generate the fictitious but realistic data of Table 10.9.3.

		Smoking history		Total
		Smoker	Former smoker	
Lung cancer?	Yes	273	227	500
	No	173	327	500

From Table 10.9.3 we can estimate the rowwise conditional probabilities

$$\hat{p}_1^* = \frac{273}{500} = 0.546 \approx 0.55$$

$$\hat{p}_2^* = \frac{173}{500} = 0.346 \approx 0.35$$

However, from the data in Table 10.9.3 we cannot estimate the columnwise conditional probabilities  $p_1$  and  $p_2$ : Because the row totals were predetermined by design, the data contain no information about  $\Pr\{\text{lung cancer} \mid \text{smoker}\}$  and  $\Pr\{\text{lung cancer} \mid \text{former smoker}\}$ . ■

The preceding example shows that, depending on the design, a study may not permit estimation of both columnwise probabilities  $p_1$  and  $p_2$  and rowwise probabilities  $p_1^*$  and  $p_2^*$ . Fortunately, the odds ratio is the same whether it is determined columnwise or rowwise. Specifically,

$$\theta = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \frac{\frac{p_1^*}{1-p_1^*}}{\frac{p_2^*}{1-p_2^*}}$$

Because of this relationship, the odds ratio  $\theta$  can be estimated by estimating  $p_1$  and  $p_2$  or by estimating  $p_1^*$  and  $p_2^*$ . This fact has important applications, especially for case-control studies, as illustrated by the following example.

**Example 10.9.5**

**Smoking and Lung Cancer** To characterize the relationship between smoking and lung cancer mortality, the columnwise probabilities  $p_1$  and  $p_2$  are more biologically meaningful than the rowwise probabilities  $p_1^*$  and  $p_2^*$ . If we investigate the relationship using a case-control design, neither  $p_1$  nor  $p_2$  can be estimated from the data. (See Example 10.9.4.) However, the odds ratio *can* be estimated from the data. For instance, from Table 10.9.3 we obtain

$$\begin{aligned} \hat{\theta} &= \frac{\frac{\hat{p}_1^*}{1-\hat{p}_1^*}}{\frac{\hat{p}_2^*}{1-\hat{p}_2^*}} \\ &= \frac{\frac{0.546}{1-0.546}}{\frac{0.346}{1-0.346}} = 2.265 \approx 2.27 \end{aligned}$$

We can interpret this odds ratio as follows: We know that the outcome event—developing lung cancer—is rare, and so we know that the odds ratio is approximately equal to the relative risk,  $p_1/p_2$ . We therefore estimate that the risk of lung cancer is about 2.3 times as great for smokers as for former smokers. ■

There is an easier way to compute the odds ratio for a  $2 \times 2$  contingency table. For a general  $2 \times 2$  table, let  $n_{11}$  denote the number of observations in the first row and the first column. Likewise, let  $n_{12}$  be the number of observations in the first row and second column, and so on. The general  $2 \times 2$  table then has the form

$n_{11}$	$n_{12}$
$n_{21}$	$n_{22}$

The estimated odds ratio from the table is

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

**Example 10.9.6**

**Smoking and Lung Cancer** From the data in Table 10.9.1, we can calculate the estimated odds ratio as

$$\hat{\theta} = \frac{89 \times 5,711}{37 \times 6,063} = 2.265 \approx 2.27$$

The case-control design is often the most efficient design for investigating rare outcome events, such as rare diseases. Although Table 10.9.3 was constructed assuming that the two samples, cases and controls, were chosen independently, a more common design is to incorporate matching of cases and controls with respect to potential confounding factors (for example, age). As we have seen, by taking advantage of the odds ratio, one can estimate the relative risk from a case-control study of a rare event even though one cannot estimate the risks  $p_1$  and  $p_2$  separately.

If the odds ratio (or the relative risk) is equal to 1.0, then the odds (or the risk) are the same for both of the groups being compared. In the smoking and lung cancer data of Table 10.9.1 the calculated odds ratio was *greater* than 1.0, indicating that the odds of lung cancer are greater for smokers than for former smokers. Notice that we could have focused attention on the odds of not getting lung cancer. In this case, the odds ratio would be *less* than 1.0, as shown in Example 10.9.7.

**Example 10.9.7**

**Smoking and Lung Cancer** Suppose we rearrange the data in Table 10.9.1 by putting lung cancer in the second row and not getting lung cancer in the first row:

		Smoking History	
		Smoker	Former Smoker
Lung cancer?	No	6,063	5,711
	Yes	89	37
Total		6,152	5,748

In this case the odds ratio is the odds of not getting lung cancer for a smoker divided by the odds of not getting lung cancer for a former smoker. We can calculate the estimated odds ratio as

$$\hat{\theta} = \frac{6,063 \times 37}{5,711 \times 89} = 0.44$$

This is the reciprocal of the odds ratio calculated in Example 10.9.6:  $\frac{1}{2.27} = 0.44$ . The fact that the odds ratio is less than 1.0 means that the event (being free of lung cancer) is less likely for smokers than for former smokers.

## Confidence Interval for the Odds Ratio

In Chapter 6 we discussed confidence intervals for proportions, which are of the form  $\tilde{p} \pm z_{\alpha/2} \text{SE}_{\tilde{p}}$ , where  $\tilde{p} = \frac{y+2}{n+4}$ . In particular, a 95% confidence interval for  $p$  is given by  $\tilde{p} \pm z_{0.025} \text{SE}_{\tilde{p}}$ . Such confidence intervals are based on the fact that for large samples the sampling distribution of  $\tilde{p}$  is approximately normal (according to the Central Limit Theorem).

In a similar way, we can construct a confidence interval for an odds ratio. One problem is that the sampling distribution of  $\hat{\theta}$  is not normal. However, if we take the natural logarithm of  $\hat{\theta}$ , then we have a distribution that is approximately normal. Hence, we construct a confidence interval for  $\theta$  by first finding a confidence interval for  $\ln(\theta)$  and then transforming the endpoints back to the original scale.

In order to construct a confidence interval for  $\ln(\theta)$ , we need the standard error of  $\ln(\hat{\theta})$ . The formula for the standard error of  $\ln(\hat{\theta})$  is given in the box.

#### Standard Error of $\ln(\hat{\theta})$

$$SE_{\ln(\hat{\theta})} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

A 95% confidence interval for  $\ln(\theta)$  is given by  $\ln(\hat{\theta}) \pm (1.96)SE_{\ln(\hat{\theta})}$ . We then exponentiate the two endpoints of the interval to get a 95% confidence interval for  $\theta$ . Intervals with other confidence coefficients are constructed analogously; for instance, for a 90% confidence interval one would use  $z_{0.05}$  (1.645) instead of  $z_{0.025}$  (1.960). The process for finding a confidence interval for  $\theta$  is summarized in the following box.\*

#### Confidence Interval for $\theta$

To construct a 95% confidence interval for  $\theta$ ,

1. Calculate  $\ln(\hat{\theta})$ .
2. Construct a confidence interval for  $\ln(\theta)$  using the formula  $\ln(\hat{\theta}) \pm (1.96)SE_{\ln(\hat{\theta})}$ .
3. Exponentiate the endpoints to get a confidence interval for  $\theta$ .

This process is illustrated in the following examples.

#### Example 10.9.8

**Smoking and Lung Cancer** From the data in Table 10.9.1, the estimated odds ratio is

$$\hat{\theta} = \frac{89 \times 5711}{37 \times 6063} = 2.27$$

Thus,  $\ln(\hat{\theta}) = \ln(2.27) = 0.820$ .

The standard error is given by  $SE_{\ln(\hat{\theta})} = \sqrt{\frac{1}{89} + \frac{1}{37} + \frac{1}{6063} + \frac{1}{5711}} = 0.1965$ .

A 95% confidence interval for  $\ln(\theta)$  is  $0.820 \pm (1.96)(0.1965)$  or  $0.820 \pm 0.385$ . This interval is (0.435, 1.205).

To get a 95% confidence interval for  $\theta$ , we evaluate  $e^{0.435} = 1.54$  and  $e^{1.205} = 3.24$ . Thus, we are 95% confident that the population value of the odds ratio is between 1.54 and 3.24. ■

\*A confidence interval for the relative risk can be found in a suitably modified manner for those situations in which the relative risk can be estimated from the data.

**Example 10.9.9**

**Heart Attacks and Aspirin** During the Physician’s Health Study, 11,037 physicians were randomly assigned to take 325 mg of aspirin every other day; 104 of them had heart attacks during the study. Another 11,034 physicians were randomly assigned to take a placebo; 189 of them had heart attacks. These data are shown in Table 10.9.4.<sup>53</sup> The odds ratio for comparing the heart attack rate on aspirin to the heart attack rate on placebo is

$$\hat{\theta} = \frac{189 \times 10,933}{104 \times 10,845} = 1.832$$

Thus,  $\ln(\hat{\theta}) = \ln(1.832) = 0.605$ .

The standard error is

$$SE_{\ln(\hat{\theta})} = \sqrt{\frac{1}{189} + \frac{1}{104} + \frac{1}{10,845} + \frac{1}{10,933}} = 0.123.$$

A 95% confidence interval for  $\ln(\theta)$  is  $0.605 \pm (1.96)(0.123)$  or  $0.605 \pm 0.241$ . This interval is (0.364, 0.846).

To get a 95% confidence interval for  $\theta$ , we evaluate  $e^{0.364} = 1.44$  and  $e^{0.846} = 2.33$ . Thus, we are 95% confident that the population value of the odds ratio is between 1.44 and 2.33. Because heart attacks are relatively rare in this data set, the relative risk is nearly equal to the odds ratio. Thus, we can say that we are 95% confident that the probability of a heart attack is about 1.44 to 2.33 times greater when taking the placebo than when taking aspirin. ■

**Table 10.9.4** Heart attacks on placebo and on aspirin

	Placebo	Aspirin
Heart attack	189	104
No heart attack	10,845	10,933
Total	11,034	11,037

**Exercises 10.9.1–10.9.8**

**10.9.1** For each of the following tables, calculate (i) the relative risk and (ii) the odds ratio.

(a)

25	23
492	614

(b)

12	8
93	84

**10.9.2** For each of the following tables, calculate (i) the relative risk and (ii) the odds ratio.

(a)

14	16
322	412

(b)

15	7
338	82

**10.9.3** Hip dysplasia is a hip socket abnormality that affects many large breed dogs. A review of medical records of dogs seen at 27 veterinary medical teaching hospitals found that hip dysplasia was more common in Golden Retrievers than in Border Collies; the data are shown in the following table.<sup>54</sup> Calculate the relative risk of hip dysplasia for Golden Retrievers compared to Border Collies.

		GOLDEN RETRIEVER	BORDER COLLIE
Hip dysplasia?	Yes	3,995	221
	No	42,946	5,007
Total		46,941	5,228

**10.9.4** Consider the data from Exercise 10.9.3.

- (a) Calculate the sample value of the odds ratio.
- (b) Construct a 95% confidence interval for the population value of the odds ratio.
- (c) Interpret the confidence interval from part (b) in the context of this setting.

**10.9.5** As part of the National Health Interview Survey, occupational injury data were collected on thousands of American workers. The table below summarizes part of these data.<sup>55</sup>

		SELF-EMPLOYED	EMPLOYED BY OTHERS
Injured?	Yes	210	4,391
	No	33,724	421,502
	Total	33,934	425,893

- (a) Calculate the sample value of the odds ratio.
- (b) According to the odds ratio, are self-employed workers more likely, or less likely, to be injured than persons who work for others?
- (c) Construct a 95% confidence interval for the population value of the odds ratio.
- (d) Interpret the confidence interval from part (b) in the context of this setting.

**10.9.6** Many over-the-counter decongestants and appetite suppressants contain the ingredient phenylpropranolamine. A study was conducted to investigate whether this ingredient is associated with strokes. The study found that 6 of 702 stroke victims had used an appetite suppressant containing phenylpropranolamine, compared to only 1 of 1,376 subjects in a control group. The following table summarizes these data.<sup>56</sup>

		STROKE	NO STROKE
Appetite suppressant?	Yes	6	1
	No	696	1,375
Total		702	1,376

- (a) Calculate the sample value of the odds ratio.
- (b) Construct a 95% confidence interval for the population value of the odds ratio.
- (c) Upon hearing of these data, some scientists called the study “inconclusive” because the numbers of users of appetite suppressants containing phenylpropranolamine (7 total: 6 in one group and 1 in the other) are so small. What is your response to these scientists?

**10.9.7** Two treatments, heparin and enoxaparin, were compared in a double-blind, randomized clinical trial of patients with coronary artery disease. The subjects can be classified as having a positive or negative response to treatment; the data are given in the following table.<sup>57</sup>

		HEPARIN	ENOXAPARIN
Outcome	Negative	309	266
	Positive	1,255	1,341
Total		1,564	1,607

- (a) Calculate the sample value of the odds ratio.
- (b) Construct a 95% confidence interval for the population value of the odds ratio.
- (c) Interpret the confidence interval from part (b) in the context of this setting.

**10.9.8** Consider the data from Exercise 10.7.1. Given that there were 139 hip fractures in 1,062 placebo patients and 92 hip fractures in 1,065 zolendronic acid patients, construct a 95% confidence interval for the population value of the odds ratio.<sup>45</sup>

## 10.10 Summary of Chi-Square Test

The chi-square test is often applied to contingency tables; it is summarized here.

### Summary of Chi-Square Test for a Contingency Table

*Null hypothesis:*

$H_0$ : Row variable and column variable are independent

*Calculation of expected frequencies:*

$$e_i = \frac{(\text{Row Total}) \times (\text{Column Total})}{\text{Grand Total}}$$

Test statistic:

$$\chi_s^2 = \sum_{\text{all cells}} \frac{(o_i - e_i)^2}{e_i}$$

Null distribution (approximate):

$$\chi^2 \text{ distribution with } df = (r - 1)(k - 1)$$

where  $r$  is the number of rows and  $k$  is the number of columns in the contingency table. This approximation is adequate if  $e_i \geq 5$  for every cell. If  $r$  and  $k$  are large, the condition that  $e_i \geq 5$  is less critical and the  $\chi^2$  approximation is adequate if the average expected frequency is at least 5, even if some of the cell counts are smaller.

The observations must be independent of one another. If paired data are collected for a  $2 \times 2$  table, then McNemar's test is appropriate (Section 10.8).

## Supplementary Exercises 10.S.1–10.S.19

(Note: Exercises preceded by an asterisk refer to optional sections.)

**10.S.1** In the Women's Health Initiative Dietary Modification Trial women were randomly assigned to an intervention or a control group. The intervention included counseling sessions designed to reduce fat intake and to increase consumption of fruits and vegetables. Over six years data were collected on coronary heart disease (CHD); results are shown in the table.<sup>58</sup> Do the data provide evidence that the intervention makes a difference? The value of the chi-square statistic for this contingency table is  $\chi_s^2 = 0.69$ . Carry out the chi-square test against a nondirectional alternative at  $\alpha = 0.10$ .

		GROUP		
		INTERVENTION	CONTROL	
CHD?	Yes	1,000	1,549	2,549
	No	18,541	27,745	46,286
	Total	19,541	29,294	48,835

**10.S.2** Use the data from Exercise 10.S.1 to construct a 95% confidence interval for  $(\Pr\{\text{CHD}|\text{intervention}\} - \Pr\{\text{CHD}|\text{control}\})$ .

**10.S.3** As part of a study of environmental influences on sex determination in the fish *Menidia*, eggs from a single mating were divided into two groups and raised in either

a warm or a cold environment. It was found that 73 of 141 offspring in the warm environment and 107 of 169 offspring in the cold environment were females.<sup>59</sup> In each of the following chi-square tests, use a nondirectional alternative and let  $\alpha = 0.05$ .

- Test the hypothesis that the population sex ratio is 1:1 in the warm environment.
- Test the hypothesis that the population sex ratio is 1:1 in the cold environment.
- Test the hypothesis that the population sex ratio is the same in the warm as in the cold environment.
- Define the population to which the conclusions reached in parts (a)–(c) apply. (Is it the entire genus *Menidia*?)

**10.S.4** The cilia are hairlike structures that line the nose and help to protect the respiratory tract from dust and foreign particles. A medical team obtained specimens of nasal tissue from nursery school children who had viral upper respiratory infections, and also from healthy children in the same classroom. The tissue was sectioned and the cilia were examined with a microscope for specific defects, with the results shown in the accompanying table.<sup>60</sup> The data show that the percentage of defective cilia was much higher in the tissue from infected children (15.7% versus 3.1%). Would it be valid to apply a chi-square test to compare these percentages? If so, do it. If not, explain why not.

	NUMBER OF CHILDREN	TOTAL NUMBER OF CILIA COUNTED	CILIA WITH DEFECTS	
			NUMBER	PERCENT
Control	7	556	17	3.1
Respiratory infection	22	1,493	235	15.7

**10.S.5** A group of mountain climbers participated in a trial to investigate the usefulness of the drug acetazolamide in preventing altitude sickness. The climbers were randomly assigned to receive either drug or placebo during an ascent of Mt. Rainier. The experiment was supposed to be double-blind, but the question arose whether some of the climbers might have received clues (perhaps from the presence or absence of side effects or from a perceived therapeutic effect or lack of it) as to which treatment they were receiving. To investigate this possibility, the climbers were asked (after the trial was over) to guess which treatment they had received.<sup>61</sup> The results can be cast in the following contingency table, for which  $\chi_s^2 = 5.07$ :

		TREATMENT RECEIVED	
		DRUG	PLACEBO
Guess	Correct	20	12
	Incorrect	11	21

Alternatively the same results can be rearranged in the following contingency table, for which  $\chi_s^2 = 0.01$ :

		TREATMENT RECEIVED	
		DRUG	PLACEBO
Guess	Drug	20	21
	Placebo	11	12

Consider the null hypothesis

$H_0$ : The blinding was perfect (the climbers received no clues).

Carry out the chi-square test of  $H_0$  against the alternative that the climbers did receive clues. Let  $\alpha = 0.05$ . (You must decide which contingency table is relevant to this question.) (*Hint*: To clarify the issue for yourself, try inventing a fictitious data set in which most of the climbers *have* received strong clues, so that we would expect a large value of  $\chi_s^2$ ; then arrange your fictitious data in each of the two contingency table formats and note which table would yield a larger value of  $\chi_s^2$ .)

**\*10.S.6** Desert lizards (*Dipsosaurus dorsalis*) regulate their body temperature by basking in the sun or moving into the shade, as required. Normally the lizards will maintain a daytime temperature of about 38 °C. When they are sick, however, they maintain a temperature about 2° to 4° higher—that is, a “fever.” In an experiment to see whether this fever might be beneficial, lizards were given a bacterial infection; then 36 of the animals were prevented from developing a fever by keeping them in a 38° enclosure, while 12 animals were kept

at a temperature of 40°. The following table describes the mortality after 24 hours.<sup>62</sup> How strongly do these results support the hypothesis that fever has survival value? Use Fisher’s exact test against a directional alternative. Let  $\alpha = 0.05$ .

	38°	40°
Died	18	2
Survived	18	10
Total	36	12

**10.S.7** Consider the data from Exercise 10.S.6. Analyze these data with a chi-square test. Let  $\alpha = 0.05$ .

**10.S.8** In a randomized clinical trial, 154 women with breast cancer were assigned to receive chemotherapy. Another 164 women were assigned to receive chemotherapy combined with radiation therapy. Survival data after 15 years are given in the following table.<sup>63</sup> Use these data to conduct a test of the null hypothesis that type of treatment does not affect survival rate. Let  $\alpha = 0.05$ .

	CHEMOTHERAPY ONLY	CHEMOTHERAPY AND RADIATION THERAPY
Died	78	66
Survived	76	98
Total	154	164

**\*10.S.9** Refer to the data in Exercise 10.S.8.

- Calculate the sample odds ratio.
- Find a 95% confidence interval for the population value of the odds ratio.

**10.S.10** Two drugs, zidovudine and didanosine, were tested for their effectiveness in preventing progression of HIV disease in children. In a double-blind clinical trial, 276 children with HIV were given zidovudine, 281 were given didanosine, and 274 were given zidovudine plus didanosine. The following table shows the survival data for the three groups.<sup>64</sup> Use these data to conduct a test of the null hypothesis that survival and treatment are independent. Let  $\alpha = 0.10$ .

	ZIDOVUDINE	DIDANOSINE	ZIDOVUDINE AND DIDANOSINE
Died	17	7	10
Survived	259	274	264
Total	276	281	274

**10.S.11** The blood types of malaria patients at a clinic in India were compared with those obtained in a sample of visitors to a nearby hospital. The data are shown in the following table.<sup>65</sup> Use these data to conduct a test of the null hypothesis that blood type is independent of contracting malaria. Let  $\alpha = 0.05$ .

	A	B	O	AB	TOTAL
Malaria cases	138	199	106	33	476
Controls	229	535	428	96	1,300

**10.S.12** The habitat selection behavior of the fruitfly *Drosophila subobscura* was studied by capturing flies from two different habitat sites. The flies were marked with colored fluorescent dust to indicate the site of capture and then released at a point midway between the original sites. On the following two days, flies were recaptured at the two sites. The results are summarized in the table.<sup>66</sup> The value of the chi-square statistic for this contingency table is  $\chi_s^2 = 10.44$ . Test the null hypothesis of independence against the alternative that the flies preferentially tend to return to their site of capture. Let  $\alpha = 0.01$ .

	SITE OF RECAPTURE	
	I	II
Site of original capture	I	II
	78	56
	33	58

**10.S.13** In the garden pea *Pisum sativum*, seed color can be yellow (Y) or green (G), and seed shape can be round (R) or wrinkled (W). Consider the following three hypotheses describing a population of plants:

$$H_0^{(1)}: \Pr\{Y\} = \frac{3}{4}$$

$$H_0^{(2)}: \Pr\{R\} = \frac{3}{4}$$

$$H_0^{(3)}: \Pr\{R|Y\} = \Pr\{R|G\}$$

The first hypothesis asserts that yellow and green plants occur in a 3:1 ratio; the second hypothesis asserts that round and wrinkled plants occur in a 3:1 ratio, and the third hypothesis asserts that color and shape are independent. (In fact, for a population of plants produced by a certain cross—the dihybrid cross—all three hypotheses are known to be true.)

Suppose a random sample of 1,600 plants is to be observed, with the data to be arranged in the following contingency table:

		COLOR	
		Y	G
Shape	R		
	W		
			1,600

Invent fictitious data sets as specified, and verify each answer by calculating the estimated conditional probabilities. (*Hint:* In each case, begin with the marginal frequencies.)

- (a) A data set that agrees perfectly with  $H_0^{(1)}$ ,  $H_0^{(2)}$ , and  $H_0^{(3)}$ .
- (b) A data set that agrees perfectly with  $H_0^{(1)}$  and  $H_0^{(2)}$  but not with  $H_0^{(3)}$ .
- (c) A data set that agrees perfectly with  $H_0^{(3)}$  but not with  $H_0^{(1)}$  or  $H_0^{(2)}$ .

**\*10.S.14** A study of 36,080 persons who had heart attacks found that men were more likely to survive than were women. The following table shows some of the data collected in the study.<sup>67</sup>

		MEN	WOMEN
Survived at least 24 hours?	Yes	25,339	8,914
	No	1,141	686
Total		26,480	9,600

- (a) Calculate the odds ratio for comparing survival of men to survival of women.
- (b) Calculate a 95% confidence interval for the population value of the odds ratio.
- (c) Does the odds ratio give a good approximation to the relative risk for these data? Why or why not?

**\*10.S.15** In the study described in Exercise 10.9.6, one of the variables measured was whether the subjects had used any products containing phenylpropanolamine. The odds ratio was calculated to be 1.49, with stroke victims more likely than the control subjects to have used a product containing phenylpropanolamine.<sup>56</sup> A 95% confidence interval for the population value of the odds ratio is (0.84, 2.64). Interpret this confidence interval in the context of this setting.

**10.S.16 (Computer exercise)** In a study of the effects of smoking cigarettes during pregnancy, researchers examined the placenta from each of 58 women after childbirth. They noted the presence or absence (P or A) of a particular placental abnormality—atrophyed villi. In addition, each woman was categorized as a nonsmoker (N), moderate smoker (M), or heavy smoker (H). The following table shows, for each woman, an ID number (#) and the results for smoking (S) and atrophyed villi (V).<sup>68</sup>

#	S	V	#	S	V	#	S	V	#	S	V
1	N	A	16	H	P	31	M	A	46	M	A
2	M	A	17	H	P	32	M	A	47	H	P
3	N	A	18	N	A	33	N	A	48	H	P
4	M	A	19	M	P	34	N	A	49	H	A
5	M	A	20	N	P	35	N	A	50	N	P
6	M	P	21	M	A	36	H	P	51	N	A
7	H	P	22	H	A	37	N	A	52	M	P
8	N	A	23	M	P	38	H	P	53	M	A
9	N	A	24	N	A	39	H	P	54	H	P
10	M	P	25	N	P	40	N	A	55	H	A
11	N	A	26	N	A	41	M	A	56	M	P
12	N	P	27	N	A	42	N	A	57	H	P
13	H	P	28	M	P	43	H	A	58	H	P
14	M	A	29	N	A	44	M	A			
15	M	P	30	N	A	45	M	P			

- (a) Test for a relationship between smoking status and atrophied villi. Use a chi-square test at  $\alpha = 0.05$ .
- (b) Prepare a table that shows the total number of women in each smoking category, and the number and percentage in each category who had atrophied villi.
- (c) What pattern appears in the table of part (b) that is not used by the test of part (a)?

**\*10.S.17** Researchers studied the cellular telephone records of 699 persons who had automobile accidents. They determined that 170 of the 699 had made a cellular telephone call during the 10-minute period prior to their accident; this period is called the hazard interval. There were 37 persons who had made a call during a corresponding 10-minute period on the day before their accident; this period is called the control interval. Finally, there were 13 who made calls both during the hazard interval and the control interval.<sup>69</sup> Do these data indicate that use of a cellular telephone is associated with an increase in accident rate? Analyze these data using McNemar's test. Use a directional alternative and let  $\alpha = 0.01$ .

		CALL DURING CONTROL INTERVAL?	
		YES	NO
Call during hazard interval?	Yes	13	157
	No	24	505

**10.S.18** Prior to an influenza season subjects were randomly assigned to receive either a flu vaccine or a placebo. During that season there were 28 cases of the flu among 813 vaccine recipients and 35 cases of the flu among the 325 subjects who were given the placebo.<sup>70</sup> Do these data indicate that the vaccine was effective? Conduct an appropriate test using a directional alternative with  $\alpha = 0.05$ .

**\*10.S.19** Refer to the data in Exercise 10.S.18.

- (a) Calculate the sample odds ratio.
- (b) Find a 95% confidence interval for the population value of the odds ratio.

**10.S.20** Consider Exercise 9.S.18. The romantic partners of the 36 men discussed in Exercise 9.S.18 were also tested, in the same manner as the men (i.e., they were blindfolded and asked to identify their partner by touching the backs of the hands of three men, one of whom was their partner). Among the women, 25 were successful and 11 were not. Are these data significant evidence for the hypothesis that men and women differ in their ability to identify their partners?

Conduct a test, using  $\alpha = 0.05$ ; use a nondirectional alternative.