CHAPTER 23

Contingency Tables

Enumeration data may be collected simultaneously for two nominal-scale variables. These data may be displayed in what is known as a *contingency table*, where the r rows of the table represent the r categories of one variable and the c columns indicate the c categories of the other variable; thus, there are rc "cells" in the table. (This presentation of data is also known as a *cross tabulation* or *cross classification*.)

Example 23.1a is of a contingency table of two rows and four columns, and may be referred to as a 2×4 ("two by four") table having $(2)(4) = 8$ cells. A sample of 300 people has been obtained from a specified population (let's say members of an actors' professional association), and the variables tabulated are each person's sex and each person's hair color. In this 2×4 table, the number of people in the sample with each of the eight combinations of sex and hair color is recorded in one of the eight cells of the table. These eight data could also be recorded in a 4×2 contingency table, with the four hair colors appearing as rows and the two sexes as columns, and that would not change the statistical hypothesis tests or the conclusions that result from them. As with previous statistical tests, the total number of data in the sample is designated as n.

EXAMPLE 23.1 A 2 × 4 Contingency Table for Testing the Independence of Hair Color and Sex in Humans

(a) H_0: Human hair color is independent of sex in the population sampled.
H_A: Human hair color is not independent of sex in the population sampled.

$\alpha = 0.05$

	Hair color				
Sex	*Black*	*Brown*	*Blond*	*Red*	**Total**
Male	32	43	16	9	100 $(= R_1)$
Female	55	65	64	16	200 $(= R_2)$
Total	87 $(= C_1)$	108 $(= C_2)$	80 $(= C_3)$	25 $(= C_4)$	300 $(= n)$

(b) The observed frequency, f_{ij}, in each cell is shown, with the frequency expected if H_0 is true (i.e., $\hat{f}_{ij}$) in parentheses.

	Hair color				
Sex	*Black*	*Brown*	*Blond*	*Red*	**Total**
Male	32 *(29.0000)*	43 *(36.0000)*	16 *(26.6667)*	9 *(8.3333)*	100 (= R_1)
Female	55 *(58.0000)*	65 *(72.0000)*	64 *(53.3333)*	16 *(16.6667)*	200 (= R_2)
Total	87 (= C_1)	108 (= C_2)	80 (= C_3)	25 (= C_4)	300 (= n)

$$\chi^2 = \sum\sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}$$

$$= \frac{(32 - 29.0000)^2}{29.0000} + \frac{(43 - 36.0000)^2}{36.0000} + \frac{(16 - 26.6667)^2}{26.6667}$$

$$+ \frac{(9 - 8.3333)^2}{8.3333} + \frac{(55 - 58.0000)^2}{58.0000} + \frac{(65 - 72.0000)^2}{72.0000}$$

$$+ \frac{(64 - 53.3333)^2}{53.3333} + \frac{(16 - 16.6667)^2}{16.6667}$$

$$= 0.3103 + 1.3611 + 4.2667 + 0.0533 + 0.1552 + 0.6806 + 2.1333$$
$$+ 0.0267 = 8.987$$

$$\nu = (r - 1)(c - 1) = (2 - 1)(4 - 1) = 3$$

$$\chi^2_{0.05,3} = 7.815$$

Therefore, reject H_0.

$$0.025 < P < 0.05 \quad [P = 0.029]$$

The hypotheses to be tested in this example may be stated in any of these three ways:

H_0: In the sampled population, a person's hair color is independent of that person's sex (that is, a person's hair color is not associated with the person's sex), and

H_A: In the sampled population, a person's hair color is not independent of that person's sex (that is, a person's hair color is associated with the person's sex), *or*

H_0: In the sampled population, the ratio of males to females is the same for people having each of the four hair colors, and

H_A: In the sampled population, the ratio of males to females is not the same for people having each of the four hair colors; *or*

H_0: In the sampled population, the proportions of people with the four hair colors is the same for both sexes, and

H_A: In the sampled population, the proportions of people with the four hair colors is not the same for both sexes.

In order to test the stated hypotheses, the sample of data in this example could have been collected in a variety of ways:

- It could have been stipulated, in advance of collecting the data, that a specified number of males would be taken at random from all the males in the population and a specified number of females would be taken at random from all the females in the population. Then the hair color of the people in the sample would be recorded for each sex. That is what was done for Example 23.1a, where it was decided, before the data were collected, that the sample would consist of 100 males and 200 females.
- It could have been stipulated, in advance of collecting the data, that a specified number of people with each hair color would be taken at random from all persons in the population with that hair color. Then the sex of the people in the sample would be recorded for each hair color.
- It could have been stipulated, in advance of collecting, that a sample of n people would be taken at random from the population, without specifying how many of each sex would be in the sample or how many of each hair color would be in the sample. Then the sex and hair color of each person would be recorded.

For most contingency-table situations, the same statistical testing procedure applies to any one of these three methods of obtaining the sample of n people, and the same result is obtained. However, when dealing with the smallest possible contingency table, namely one with only two rows and two columns (Section 23.3), an additional sampling strategy may be encountered that calls for a different statistical procedure.

Section 23.8 will introduce procedures for analyzing contingency tables of more than two dimensions, where frequencies are tabulated simultaneously for more than two variables.

23.1 CHI-SQUARE ANALYSIS OF CONTINGENCY TABLES

The most common procedure for analyzing contingency table data uses the chi-square statistic.* Recall that for the computation of chi-square one utilizes observed and expected frequencies (and never proportions or percentages). For the goodness-of-fit analysis introduced in Section 22.1, f_i denoted the frequency observed in category i of the variable under study. In a contingency table, we have two variables under consideration, and we denote an observed frequency as f_{ij}. Using the double subscript, f_{ij} refers to the frequency observed in row i and column j of the contingency table. In Example 23.1, the value in row 1 column 1 is denoted as f_{11}, that in row 2 column 3 as f_{23}, and so on. Thus, $f_{11} = 32, f_{12} = 43, f_{13} = 16, \ldots, f_{23} = 64$, and $f_{24} = 16$.

The total frequency in row i of the table is denoted as R_i and is obtained as $R_i = \sum_{j=1}^{c} f_{ij}$. Thus, $R_1 = f_{11} + f_{12} + f_{13} + f_{14} = 100$, which is the total number of males in the sample, and $R_2 = f_{21} + f_{22} + f_{23} + f_{24} = 200$, which is the total number of females in the sample. The column totals, C_j, are obtained by analogous

*The early development of chi-square analysis of contingency tables is credited to Karl Pearson (1904) and R. A. Fisher (1922). In 1904, Pearson was the first to use the term "contingency table" (David, 1995).

and it is in this way that the $\hat{f}_{ij}$ values in Example 23.1b were obtained. Note that we can check for arithmetic errors in our calculations by observing that $R_i = \sum_{j=1}^{c} f_{ij} = \sum_{j=1}^{c} \hat{f}_{ij}$ and $C_j = \sum_{i=1}^{r} f_{ij} = \sum_{i=1}^{r} \hat{f}_{ij}$. That is, the row totals of the expected frequencies equal the row totals of the observed frequencies, and the column totals of the expected frequencies equal the column totals of the observed frequencies.

Once χ^2 has been calculated, its significance can be ascertained from Appendix Table B.1, but to do so we must determine the degrees of freedom of the contingency table.

The degrees of freedom for a chi-square calculated from contingency-table data are*

$$\nu = (r - 1)(c - 1). \tag{23.5}$$

In Example 23.1, which is a 2×4 table, $\nu = (2 - 1)(4 - 1) = 3$. The calculated statistic is 9.987 and the critical value is $\chi^2_{0.05,3} = 7.815$, so the null hypothesis is rejected.

It is good to calculate expected frequencies and other intermediate results to at least four decimal places and to round to three decimal places after arriving at the value of χ^2. Barnett and Lewis (1994: 431–440) and Simonoff (2003: 228–234) discuss outliers in contingency-table data.

(a) Comparing Proportions. Hypotheses for data in a contingency table with only two rows (or only two columns) often refer to ratios or proportions. In Example 23.1, the null hypothesis could have been stated as, "In the sampled population, the sex ratio is the same for each hair color" or as "In the sampled population, the proportion of males is the same for each hair color." The comparison of two proportions is discussed in Sections 23.3b and 24.10; and the comparison of more than two proportions is further discussed in Sections 24.13–24.15.

23.2 VISUALIZING CONTINGENCY-TABLE DATA

Among the ways to present contingency-table data in graphical form is a method known as a *mosaic* display.†

In Chapter 1, nominal-scale data were presented in a bar graph in Figure 1.2. The categories of the nominal-scale variable appear on one axis of the graph (typically the horizontal axis, as in Figure 1.2), and the number of observations is on the other

*In the early days of contingency-table analysis, K. Pearson and R. A. Fisher disagreed vehemently over the appropriate degrees of freedom to employ; Fisher's (1922) view has prevailed (Agresti, 2002: 622; Savage, 1976), as has his use of the term *degrees of freedom*.

†The current use of mosaic displays is attributed to Hartigan and Kleiner (1981). In an historical review of rectangular presentations of data, Friendly (2002) credits the English astronomer Edmond (a.k.a. Edmund) Halley (1656–1742), famous for his 1682 observation of the comet that bears his name, with the first use of rectangular areas in the data representation for two independent variables (which, however, were not variables for a contingency table). Further developments in the visual use of rectangular areas took place in France and Germany in the early 1780s; a forerunner of mosaic graphs was introduced in 1844 by French civil engineer Charles Joseph Minard (1791–1870), and what resembled the modern mosaic presentation was first used in 1877 by German statistician Georg von Mayr (1841–1925). In 1977, French cartographer Jacques Bertin (1918–) used graphs very similar to the mosaics of Hartigan and Kleiner.

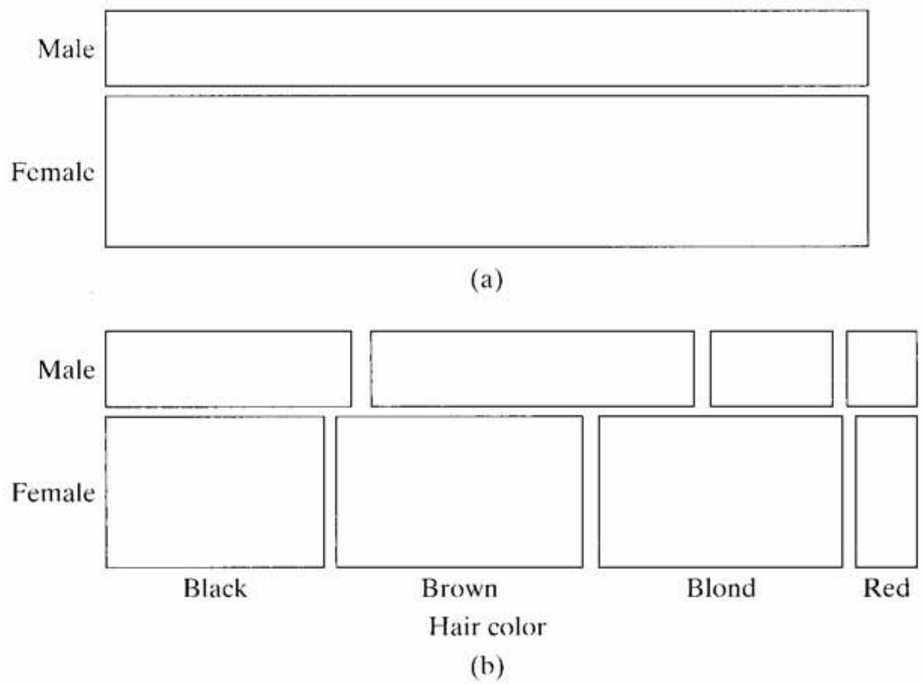

FIGURE 23.1: A mosaic display for the contingency-table data of Example 23.1. (a) The first step displays two horizontal bars of equal width with the height of one of them representing the number of males and the height of the other representing the number of females in the sample. (b) The second step divides each of the two horizontal bars into four tiles, with the width of each tile depicting the frequency in the sample of a hair color among the individuals of one of the sexes.

axis. The lengths of the bars in the graph are representations of the frequencies of occurrence of observations in the data categories; and, when bars are of equal width, the areas of the bars also depict those frequencies.

Figure 23.1 demonstrates visualizing the data in Example 23.1 and shows the two-step process of preparing a mosaic display. The first step is to prepare Figure 23.1a, which is a graph reflecting the numbers of males and females in the sample of data described in Example 23.1. Of the 300 data, 100 are males and 200 are females, so the bar for females is two times as high as the bar for males. (The bars are graphed horizontally to reflect the rows in the Example 23.1 contingency table, but they could have been drawn vertically instead.) The bars are drawn with equal widths, so their areas also express visually the proportion of the 300 data in each sex category, with the lower (female) bar having two times the area of the upper (male) bar.

The second step, shown in Figure 23.1b, is to divide each sex's horizontal bar into four segments representing the relative frequencies of the four hair colors within that sex. For example, black hair among males was exhibited by $32/100 = 0.32$ of the males in the sample, so black is depicted by a bar segment that is 32% of the width of the male bar; and $16/200 = 0.08$ of the sample's females had red hair, so the red-hair segment for females is 8% of the width of the female bar. These bar segments are often referred to as *tiles*, and there will be a tile for each of the $r \times c$ cells in the contingency table. Mosaic displays are usually, but not necessarily, drawn with small gaps between adjacent tiles.

If the boundaries of the tiles for the two bars were perfectly aligned vertically, that would indicate that the $r \times c$ frequencies were in perfect agreement with the

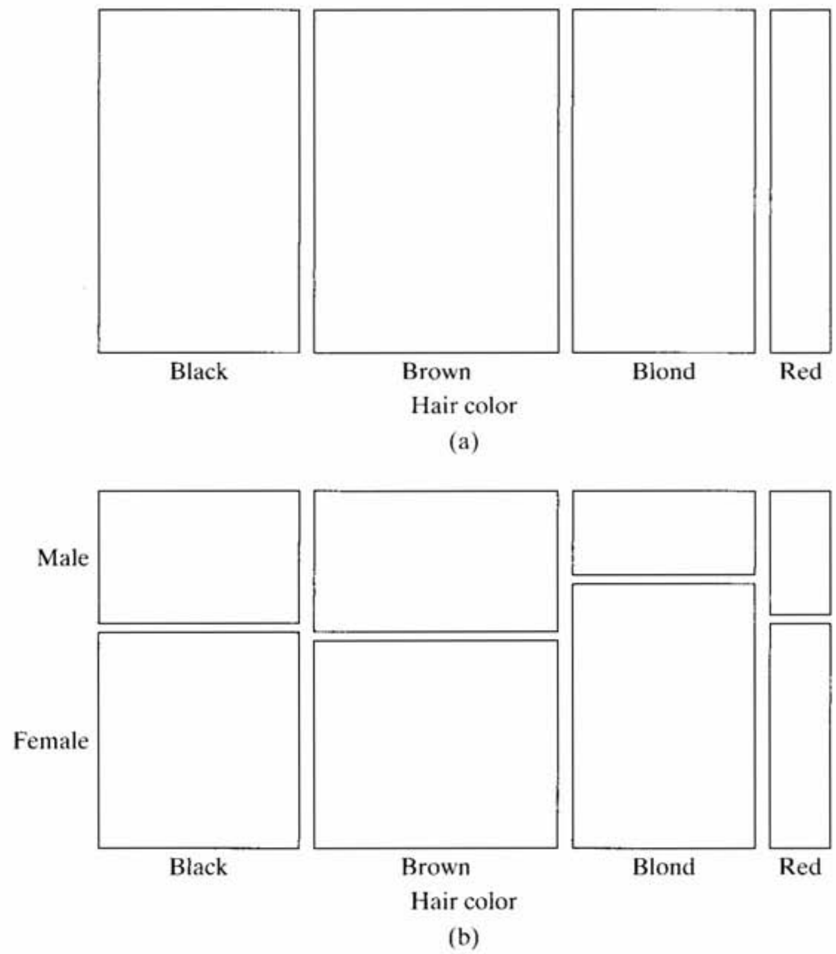

FIGURE 23.2: A mosaic display for the contingency-table data of Example 23.1. (a) The first step displays four vertical bars of equal height, one for each of the hair colors in the sample, with the width of the bars expressing the relative frequencies of the hair colors. (b) The second step divides each of the four vertical bars into two tiles, with the length of each tile depicting the frequency of a members of a sex among the individuals of one of the hair colors.

null hypothesis. The more out of alignment the tiles are, the less likely it is that the sampled population conforms to that specified in H_0.

In Figure 23.1, the data of Example 23.1 were displayed graphically by showing the frequency of each hair color within each sex. Alternatively, the data could be presented as the frequency of each sex for each hair color. This is shown in Figure 23.2. In Figure 23.2a, the widths of the four vertical bars represent the relative frequencies of the four hair colors in the sample, and Figure 23.2b divides each of those four bars into two segments (tiles) with sizes reflecting the proportions of males and females with each hair color.

Either graphical depiction (Figure 23.1b or 23.2b) is legitimate, with the choice depending upon the visual emphasis the researcher wants to give to each of the variables.

Mosaic displays may also be presented for contingency tables having more than two rows *and* more than two columns (such as Exercise 23.4 at the end of this chapter). Friendly (1994, 1995, 1999, 2002) described how the interpretation of mosaic graphs can be enhanced by shading or coloring the tiles to emphasize the degree to which observed frequencies differ from expected frequencies in the cells of the contingency table; and mosaic presentations are also used for contingency tables with more than two dimensions (which are introduced in Section 23.8).

2 × 2 CONTINGENCY TABLES

The smallest possible contingency table is that consisting of two rows and two columns. It is referred to as a 2 × 2 ("two by two") table or a fourfold table, and it is often encountered in biological research. By Equation 23.5, the degrees of freedom for 2 × 2 tables is $(2 - 1)(2 - 1) = 1$.

The information in a 2 × 2 contingency table may be displayed as

$$\begin{array}{|cc|c}
\hline
f_{11} & f_{12} & R_1 \\
f_{21} & f_{22} & R_2 \\
\hline
C_1 & C_2 & n
\end{array}$$

where f_{ij} denotes the frequency observed in row i and column j, R_i is the sum of the two frequencies in row i, C_j is the sum of the two frequencies in column j, and n is the total number of data in the sample. (The sample size, n, is the sum of all four of the f_{ij}'s, is the sum of the two row totals, and is the sum of the two column totals.) The row totals, R_1 and R_2, are said to occupy one margin of the table, and the column totals, C_1 and C_2, are said to occupy an adjacent margin of the table.

There are different experimental designs that result in data that can be arranged in contingency tables, depending upon the nature of the populations from which the samples come. As described by Barnard (1947) and others, these can be categorized on the basis of whether the marginal totals are set by the experimenter before the data are collected.

(a) No Margin Fixed. There are situations where only the size of the sample (n) is declared in advance of data collection, and neither the row totals nor the column totals are prescribed.* In Example 23.2a, the experimenter decided that the total number of data in the sample would be $n = 70$, but there was no specification prior to the data collection of what the total number of boys, of girls, of right-handed children, or of left-handed children would be. A sample of 70 was taken at random from a population of children (perhaps of a particular age of interest), and then the numbers of right-handed boys, right-handed girls, left-handed boys, and left-handed girls were recorded as shown in this example. The statistical analysis shown in Example 23.2b will be discussed in Section 23.3d.

EXAMPLE 23.2 A 2 × 2 Contingency Table with No Fixed Margins

(a) H_0: In the sampled population, handedness is independent of sex.
H_A: In the sampled population, handedness is not independent of sex.

$\alpha = 0.05$

	Boys	*Girls*	*Total*
Left-handed	6	12	18
Right-handed	28	24	52
Total	34	36	70

*This kind of experimental design is sometimes referred to as a double dichotomy or as representing a multinomial sampling distribution, and the resulting test as a test of association or test of independence.

(b) Using Equation 23.6 (Equation 23.1 could also be used, with the same result),

$$\chi^2 = \frac{n(f_{11}f_{22} - f_{12}f_{21})^2}{R_1R_2C_1C_2}$$

$$= \frac{70\left[(6)(24) - (12)(28)\right]^2}{(18)(52)(34)(36)}$$

$$= 2.2524.$$

$\nu = 1$; $\chi^2_{0.05,1} = 3.841$

Therefore, do not reject H_0.

$0.10 < P < 0.25$ $[P = 0.22]$

(b) One Margin Fixed. Some experimental designs not only specify the sample size, n, but also indicate—prior to collecting data—how many data in the sample will be in each row (or how many will be in each column).* Thus, in Example 23.2a, it could have been declared, before counting how many children were in each of the four categories, how many boys would be taken at random from all the boys in the population and how many girls would be taken at random from the girls in the population. Or the column totals might have been fixed, stating how many right-handed children and how many left-handed children would be selected from their respective populations.

Another example of a contingency table with one pair of marginal totals fixed is shown in Example 23.3a. In this study, it was decided to collect, at random, 24 mice of species 1 and 25 of species 2, and the researcher recorded the number of mice of each species that were infected with a parasite of interest.

EXAMPLE 23.3 A 2 × 2 Contingency Table with One Fixed Margin

(a) H_0: The proportion of the population infected with an intestinal parasite is the same in two species of mouse.

H_A: The proportion of the population infected with an intestinal parasite is not the same in two species of mouse.

$\alpha = 0.05$

	Species 1	*Species 2*	*Total*
With parasite	18	10	28
Without parasite	6	15	21
Total	24	25	49

*This experimental design is often called a comparative trial, the resulting test a test of homogeneity, and the underlying distributions binomial distributions (which will be discussed further in Chapter 24).

summations: $C_j = \sum_{i=1}^{r} f_{ij}$. For example, the total number of blonds in the sample data is $C_3 = \sum_{i=1}^{2} f_{i3} = f_{13} + f_{23} = 80$, the total number of redheads is $C_4 = \sum_{i=1}^{2} f_{i4} = 25$, and so on. The total number of observations in all cells of the table is called the grand total and is $\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij} = f_{11} + f_{12} + f_{13} + \cdots + f_{23} + f_{24} = 300$, which is n, the size of our sample. The computation of the grand total may be written in several other notations: $\sum_i \sum_j f_{ij}$ or $\sum_{i,j} f_{ij}$, or simply $\sum\sum f_{ij}$. When no indices are given on the summation signs, we assume that the summation of all values in the sample is desired.

The most common calculation of chi-square analysis of contingency tables is

$$\chi^2 = \sum\sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}. \tag{23.1}$$

In this formula, similar to Equation 22.1 for chi-square goodness of fit, $\hat{f}_{ij}$ refers to the frequency expected in a row i column j if the null hypothesis is true.* If, in Example 23.1a, hair color is in fact independent of sex, then $\frac{100}{300} = \frac{1}{3}$ of all black-haired people would be expected to be males and $\frac{200}{300} = \frac{2}{3}$ would be expected to be females. That is, $\hat{f}_{11} = \frac{100}{300}(87) = 29$ (the expected number of black-haired males), $\hat{f}_{21} = \frac{200}{300}(87) = 58$ (the expected number of black-haired females), $\hat{f}_{12} = \frac{100}{300}(108) = 36$ (the expected number of brown-haired males), and so on.

This may also be explained by the probability rule introduced in Section 5.7: The probability of two independent events occurring at once is the product of the probabilities of the two events. Thus, if having black hair is independent of being male, then the probability of a person being both black-haired and male is the probability of a person being black-haired multiplied by the probability of a person being male, namely $\left(\frac{87}{300}\right) \times \left(\frac{100}{300}\right)$, which is 0.0966667. This means that the expected number of black-haired males in a sample of 300 is $(0.0966667)(300) = 29.0000$. In general, the frequency expected in a cell of a contingency table is

$$\hat{f}_{ij} = \left(\frac{R_i}{n}\right)\left(\frac{C_j}{n}\right)n, \tag{23.3}$$

which reduces to the commonly encountered formula,

$$\hat{f}_{ij} = \frac{(R_i)(C_j)}{n}, \tag{23.4}$$

*Just as Equation 22.2 is equivalent to Equation 22.1 for chi-square goodness of fit, the following are mathematically equivalent to Equation 23.1 for contingency tables:

$$\chi^2 = \sum\sum \frac{f_{ij}^2}{\hat{f}_{ij}} - n \tag{23.2}$$

and

$$\chi^2 = n\left(\sum\sum \frac{f_{ij}^2}{R_i C_j} - 1\right). \tag{23.2a}$$

These formulas are computationally simpler than Equation 23.1, the latter not even requiring the calculation of expected frequencies; however, they do not allow for the examination of the contributions to the computed chi-square, the utility of which will be seen in Section 23.6.

(b) Using Equation 23.6 (Equation 23.1 could also be used, with the same result),

$$\chi^2 = \frac{n(f_{11}f_{22} - f_{12}f_{21})^2}{R_1R_2C_1C_2}$$

$$= \frac{49\left[(18)(15) - (10)(6)\right]^2}{(28)(21)(24)(25)}$$

$$= 6.1250.$$

$$0.01 < P < 0.025 \quad [P = 0.013]$$

If one margin is fixed, hypotheses might be expressed in terms of proportions. For Example 23.3a, the null hypothesis could be stated as H_0: In the sampled population, the proportion of infected mice is the same in species 1 and species 2. The statistical analysis shown in Example 23.3b will be discussed in Section 23.3d. Additional statistical procedures for dealing with proportions are discussed in Chapter 24.

(c) Both Margins Fixed. In some cases (which are very uncommon), both margins in the contingency table are fixed.* That is, R_1, R_2, C_1, C_2, and n are all set before the collection of data.

Data for such a 2 × 2 table are shown in Example 23.4a, where an ecologist wanted to compare the ability of two species of snails to tolerate the current of a stream and adhere to the stream's substrate. The researcher labeled 30 snails that were clinging to the bottom of the stream, 19 of them selected at random from a population of snails of one species and 11 selected at random from a population of snails of a second species. These 30 individuals were then observed as the current washed over them, and it was decided before the experiment began that data collection would end when more than half of the 30 (that is, 16) yielded to the current and were swept downstream.

EXAMPLE 23.4 A 2 × 2 Contingency Table with Two Fixed Margins

(a) H_0: The ability of snails to resist the current is no different between the two species.

H_A: The ability of snails to resist the current is different between the two species.

$\alpha = 0.05$

The four marginal totals are set before performing the experiment, and the four cell frequencies are collected from the experiment.

	Resisted	*Yielded*	
Species 1	12	7	19
Species 2	2	9	11
	14	16	30

*The sampling in this experimental design comes from what is known as a hypergeometric distribution, about which more will be said in Sections 24.2 and 24.16, and the experimental design is sometimes called an independence trial.

(b) Using Equation 23.7 (Equation 23.1 could also be used, with the same result, if $f_{ij} - \hat{f}_{ij}$ is replaced by $|f_{ij} - \hat{f}_{ij}| - 0.5$), the chi-square with the Yates correction for continuity is

$$\chi_c^2 = \frac{n\left(|f_{11}f_{22} - f_{12}f_{21}| - \frac{n}{2}\right)^2}{R_1 R_2 C_1 C_2}$$

$$= \frac{30\left[|(12)(9) - (7)(2)| - \frac{30}{2}\right]^2}{(19)(11)(14)(16)}$$

$$= 3.999.$$

$\nu = 1$

$\chi^2_{0.05,1} = 3.841.$

Therefore, reject H_0.

$0.025 < P < 0.05 \quad [P = 0.046]$

(c) Using Equation 23.7b, the chi-square with the Cochran-Haber correction for continuity is calculated as follows:

$m_1 = R_2 = 11,\ m_2 = C_1 = 14$

$\hat{f} = m_1 m_2 / n = (11)(14)/30 = 5.13$

$f = f_{21} = 2;\ d = |f - \hat{f}| = |2 - 5.13| = 3.13$

$2\hat{f} = 2(5.13) = 10.26;$

As $f < 2\hat{f}, D = 3.0$

$$\chi_H^2 = \frac{n^3 D^2}{R_1 R_2 C_1 C_2}$$

$$= \frac{(30)^3 (3.0)^2}{(19)(11)(14)(16)}$$

$$= 5.191.$$

As $\chi^2_{0.05,1} = 3.841$, reject H_0.

$0.01 < P < 0.025 \quad [P = 0.023]$

Thus, prior to collecting the data, the number of snails of each species was decided upon (as 19 and 11), *and* the total numbers of snails dislodged by the current (16 and 14) were specified. Other illustrations of 2×2 tables with both margins fixed are provided in Examples 24.20 and 24.21 and Exercises 24.20 and 24.21. The statistical analysis demonstrated in Example 23.4b will be discussed in Section 23.3d.

(d) Analysis of 2 × 2 Contingency Tables. Contingency-table hypotheses may be examined by chi-square, as shown in Section 23.1, calculating χ^2 with Equation 23.1 with the expected frequencies ($\hat{f}_{ij}$) obtained via Equation 23.4. However, for a 2×2

table, the following is a simpler computation,* for it does not require that the expected frequencies be determined, and it avoids rounding error associated with calculating $\hat{f}_{ij}$ and $f_{ij} - \hat{f}_{ij}$:

$$\chi^2 = \frac{n(f_{11}f_{22} - f_{12}f_{21})^2}{R_1R_2C_1C_2}. \tag{23.6}$$

As with goodness of fit (Section 22.1), chi-square values that are calculated come from a discrete distribution, but they are to be compared (such as by Appendix Table B.1) to chi-square values from a continuous distribution. Thus, statisticians may recommend that a correction for continuity be applied when $\nu = 1$ (which is the case when dealing with a 2×2 contingency table). More than 20 continuity corrections have been proposed; the most commonly considered is the Yates (1934) correction† (as was used in Section 22.2 for goodness of fit), which is the modification of Equation 23.1 by substituting $|f_{ij} - \hat{f}_{ij}| - 0.5$ for $f_{ij} - \hat{f}_{ij}$ or, equivalently, using the following instead of Equation 23.6:

$$\chi_c^2 = \frac{n\left(|f_{11}f_{22} - f_{12}f_{21}| - \frac{n}{2}\right)^2}{R_1R_2C_1C_2}. \tag{23.7}$$

This is the calculation employed in Example 23.4b, and its use approximates the two-tailed Fisher exact test discussed in Section 24.16b.

Haber (1980) showed that there are other correction methods that often perform better than that of Yates, which tends to be conservative (in that it has a probability less than α of a Type I error and has lower power than a nonconservative test). He proposed using a procedure based on a principle expounded by Cochran (1942, 1952). In the Cochran-Haber method (demonstrated in Example 23.4c), the smallest of the four expected frequencies is determined; using Equation 23.4, this frequency is

$$\hat{f} = \frac{m_1m_2}{n}, \tag{23.7a}$$

where m_1 is the smallest of the four marginal totals and m_2 is the smaller of the two totals in the other margin. In Example 23.4, the smallest marginal total, m_1, is 11, which is a row total; and m_2 is, therefore, the smaller of the two column totals, namely 14. Then the absolute difference between this expected frequency ($\hat{f}$) and its corresponding observed frequency (f) is $d = |f - \hat{f}|$; and

- If $f \le 2\hat{f}$, then define D = the largest multiple of 0.5 that is $<d$; and
- If $f > 2\hat{f}$, then define $D = d - 0.5$.

The chi-square with the Cochran-Haber correction is

$$\chi_H^2 = \frac{n^3D^2}{R_1R_2C_1C_2}. \tag{23.7b}$$

*Richardson (1994) attributed Equation 23.6 to Fisher (1922). Upton (1982) reported a "slight" improvement if $n - 1$ is employed in place of n.

†Pearson (1947) points out that Yates's use of this correction for chi-square analysis was employed as early as 1921 for other statistical purposes. The continuity correction for 2×2 tables should *not* be used in the very rare instances that its inclusion increases, instead of decreases, the numerator (that is, when $|f_{11} - f_{22}| < n/2$).

If $f > 2\hat{f}$, then the Cochran-Haber-corrected chi-square (χ^2_H) is the same as the chi-square with the Yates correction (χ^2_c). Also, if either $C_1 = C_2$ or $R_a = R_2$, then $\chi^2_H = \chi^2_c$.

A great deal has been written about 2 × 2 contingency-table testing.* For example, it has been reported that the power of chi-square testing increases with larger n or with more similarity between the two totals in a margin, and that the difference between results using chi-square and a continuity-corrected chi-square is less for large n.

In addition, many authors have reported that, for 2 × 2 tables having no fixed margin or only one fixed margin, χ^2_c provides a test that is very, very conservative (that is, the probability of a Type I error is far less than that indicated by referring to the theoretical chi-square distribution—such as in Appendix Table B.1), with relatively low power; and they recommend that it should not be used for such sets of data. The use of χ^2 instead of χ^2_c will occasionally result in a test that is somewhat liberal (i.e., the probability of a Type I error is a little greater than that indicated by the chi-square distribution, though it will typically be closer to the latter distribution than χ^2_c will be); this liberalism is more pronounced when the two row totals are very different or the two column totals are very different.

For many decades there has been debate and disagreement over the appropriate statistical procedure for each of the aforementioned three sampling models for data in a 2 × 2 contingency table, with arguments presented on both theoretical and empirical grounds. There is still no consensus, and some believe there never will be,† but there is significant agreement on the following:

- If the 2 × 2 table has no margin fixed or only one margin fixed, then use χ^2. This is demonstrated in Examples 23.2b and 23.3b.
- If the 2 × 2 table has both margins fixed, then use χ^2_c or χ^2_H, as demonstrated in Example 23.4, or use the Fisher exact test of Section 24.16. As noted after Equation 23.7b, there are situations in which χ^2_c and χ^2_H are equal; otherwise, χ^2_H is routinely a better approximation of the Fisher exact test and is preferred to χ^2_c.

Computer software may present χ^2 or a continuity-corrected χ^2_c, or both, and the user must decide which one of these two test statistics to use (such as by the guidelines just given).

*This paragraph and the next are a summary of the findings in many publications, such as those cited in the footnote that follows this one.

†Those promoting the analysis of any of the three models by using chi-square with the Yates correction for continuity (χ^2_c), or the Fisher exact test of Section 24.16, include Camilli (1990), Cox (1984), Fisher (1935), Kendall and Stuart (1979), Martín Andrés (1991), Mehta and Hilton (1993), Upton (1992), and Yates (1984). Among those concluding that χ^2_c should not be employed for all three models are Barnard (1947, 1979); Berkson (1978); Camilli and Hopkins (1978); Conover (1974); D'Agostino, Chase, and Belanger (1988); Garside and Mack (1976); Grizzle (1967); Haber (1980, 1982, 1987, 1990); Haviland (1990); Kempthorne (1979); Kroll (1989); Liddell (1976); Parshall and Kromrey (1996); Pearson (1947); Plackett (1964); Richardson (1990, 1994); Starmer, Grizzle, and Sen (1974); Storer and Kim (1990); and Upton (1982). Other procedures for testing 2 × 2 tables have been proposed (e.g., see Martín Andrés and Silva Mato (1994); Martín Andrés and Tapia Garcia (2004); and Overall, Rhoades, and Starbuck (1987)).

(e) One-Tailed Testing. The preceding hypotheses are two-tailed, which is the typical situation. However, one-tailed hypotheses (where a one-tailed hypothesis is specified before data are collected) are possible for data in 2×2 tables. In Example 23.2, the hypotheses could have been stated as follows:

H_0: In the sampled population, the proportion of left-handed children is the same or greater for boys compared to girls.

H_A: In the sampled population, the proportion of left-handed children is less for boys than for girls.

If the direction of the difference in the sample is that indicated in the null hypothesis (i.e., if $f_{11}/C_1 \geq f_{12}/C_2$), then H_0 cannot be rejected and the one-tailed analysis proceeds no further. However, if the direction of the difference in the sample is not in the direction of the null hypothesis (as in Example 23.2, where $6/34 < 12/36$), then it can be asked whether that difference is likely to indicate a difference in that direction in the population. In this example, one-tailed hypotheses could also have been stated as follows:

H_0: In the sampled population, the proportion of boys is the same or less for left-handed compared to right-handed children.

H_A: In the sampled population, the proportion of boys is greater for left-handed than for right-handed children.

This would ask whether the sample proportion f_{11}/R_1 (namely, 6/18) resulted from a population proportion less than or equal to the population proportion estimated by f_{21}/R_2 (i.e., 28/52).

Consistent with the preceding recommendations for two-tailed hypothesis testing, the following can be advised for one-tailed testing: For 2×2 tables in which no margin or only one margin is fixed, test by using one-half of the chi-square probability (for example, employing the critical value $\chi^2_{0.10,1}$ for testing at $\alpha = 0.05$), by dividing the resultant P by 2, or by using one-tailed values for Z in the normal approximation of Section 24.10. For tables with two fixed margins, the Fisher exact test of Section 24.16 is the preferred method of analysis, though if $R_1 = R_2$ or $C_1 = C_2$, we may calculate χ^2_c or, preferably, χ^2_H, and proceed as indicated previously for situations with one fixed margin. If neither $R_1 = R_2$ nor $C_1 = C_2$, using χ^2_H or χ^2_c yields a very poor approximation to the one-tailed Fisher exact test and is not recommended.

CONTINGENCY TABLES WITH SMALL FREQUENCIES

Section 22.5 discussed bias in chi-square goodness-of-fit testing when expected frequencies are "too small." As with goodness-of-fit testing, for a long time many statisticians (e.g., Fisher, 1925b) advised that chi-square analysis of contingency tables be employed only if each of the expected frequencies was at least 5.0—even after there was evidence that such analyses worked well with smaller frequencies (e.g., Cochran, 1952, 1954). The review and empirical analysis of Roscoe and Byars (1971) offer more useful guidelines. Although smaller sample sizes are likely to work well, a secure practice is to have the mean expected frequency be at least 6.0 when testing with α as small as 0.05, and at least 10.0 for $\alpha = 0.01$. Requiring an average expected frequency of at least 6 is typically less restrictive than stipulating that each $\hat{f}_{ij}$ be at least 5. Since the mean expected frequency is n/rc, the minimum sample size for testing at the 0.05 significance level should be at least $n = 6rc = 6(2)(4) = 48$ for a 2×4 contingency table (such as in Example 23.1) and at least $6(2)(2) = 24$ for a 2×2 table (as in Exercises 23.2, 23.3, 23.4, and 23.5).

If any of the expected frequencies are smaller than recommended, then one or more rows or columns containing an offensively low $\hat{f}_{ij}$ might be discarded, or rows or columns might be combined to result in $\hat{f}_{ij}$'s of sufficient magnitude. However, such practices are not routinely advised, for they disregard information that can be important to the hypothesis testing. When possible, it is better to repeat the experiment with a sufficiently large n to ensure large enough expected frequencies. Some propose employing the log-likelihood ratio of Section 23.7 as a test less affected than chi-square by low frequencies, but this is not universally suggested. If both margins are fixed in a 2×2 contingency table, then the Fisher exact test of Section 24.16 is highly recommended when frequencies are small.

23.5 HETEROGENEITY TESTING OF 2 × 2 TABLES

Testing for heterogeneity of replicate samples in goodness-of-fit analysis was discussed in Section 22.6. An analogous procedure may be used with contingency-table data, as demonstrated in Example 23.5. Here, data set 1 is the data from Example 23.2, and each of the three other sets of data is a sample obtained by the same data-collection procedure for the purpose of testing the same hypothesis. Heterogeneity testing asks whether all four of the data sets are likely to have come from the same population of data. In this example, a calculation of χ^2 was done, as in Section 23.3a, for each of the four contingency tables; and H_0 was not rejected for any of the data sets. This failure to reject H_0 might reflect low power of the test due to small sample sizes, so it would be helpful to use the heterogeneity test to conclude whether it would be reasonable to combine the four sets of data and perform a more powerful test of H_0 with the pooled number of data.

EXAMPLE 23.5 A Heterogeneity Chi-Square Analysis of Four 2 × 2 Contingency Tables, Where Data Set 1 Is That of Example 23.2

(a) H_0: In the sampled population, handedness is independent of sex.
H_A: In the sampled population, handedness is not independent of sex.
$\alpha = 0.05$

Data Set 1

From the data of Example 23.2, $\chi^2 = 2.2523$, DF = 1, $0.10 < P < 0.25$.

Data Set 2

	Boys	*Girls*	*Total*	
Left-handed	4	7	11	
Right-handed	25	13	38	
Total	29	20	49	$\chi^2 = 3.0578$, DF = 1, $0.05 < P < 0.10$

Data Set 3

	Boys	*Girls*	*Total*	
Left-handed	7	10	17	
Right-handed	27	18	45	
Total	34	28	62	$\chi^2 = 1.7653$, DF = 1, $0.10 < P < 0.25$

Data Set 4

	Boys	*Girls*	*Total*	
Left-handed	4	7	11	
Right-handed	22	14	36	
Total	26	21	47	$\chi^2 = 2.0877$, DF = 1, $0.10 < P < 0.25$

(b) H_0: The four samples are homogeneous.
H_A: The four samples are heterogeneous.

Data Sets 1–4 Pooled

	Boys	*Girls*	*Total*	
Left-handed	21	36	57	$\chi^2 = 8.9505$
Right-handed	102	69	171	DF = 1
Total	123	105	228	

χ^2 for Data Set 1:	2.2523	DF = 1
χ^2 for Data Set 2:	3.0578	DF = 1
χ^2 for Data Set 3:	1.7653	DF = 1
χ^2 for Data Set 4:	2.0877	DF = 1
Total chi-square:	9.1631	DF = 4
Chi-square of pooled data:	8.9505	DF = 1
Heterogeneity chi-square	0.2126	DF = 3

For heterogeneity testing (using $\chi^2 = 0.2126$):
$\chi^2_{0.05,3} = 7.815$.
Therefore, do not reject H_0.

$$0.975 < P < 0.99 \quad [P = 0.98]$$

(c) H_0: In the sampled population, handedness is independent of sex.
H_A: In the sampled population, handedness is not independent of sex.

$\alpha = 0.05$

Data Sets 1–4 Pooled

	Boys	*Girls*	*Total*
Left-handed	21	36	57
Right-handed	102	69	171
Total	123	105	228

$\chi^2_{0.05,1} = 3.841$

$\chi^2 = 8.9505$; therefore, reject H_0.

$0.001 < P < 0.005$ [$P = 0.0028$]

In the test for heterogeneity, chi-square is calculated for each of the samples; these four separate χ^2 values are shown in Example 23.5a, along with the χ^2 for the contingency table formed by the four sets of data combined. The χ^2 values for the four separate contingency tables are them summed (to obtain what may be called a total chi-square, which is 9.1631), and the degrees of freedom for the four tables are also summed (to obtain a total DF, which is 4), as shown in Example 23.5b. The test for heterogeneity employs a chi-square value that is the absolute difference between the total chi-square and the chi-square from the table of combined data, with degrees of freedom that are the difference between the total degrees of freedom and the degrees of freedom from the table of combined data. In the present example, the heterogeneity χ^2 is 0.2126, with 3 degrees of freedom. That chi-square is associated with a probability much greater than 0.05, so H_0 is not rejected and it is concluded that the data of the four samples may be combined.

Example 23.5c considers the contingency table formed by combining the data of all four of the original tables and tests the same hypothesis of independence that was tested for each of the original tables. When the heterogeneity test fails to reject H_0, pooling of the data is generally desirable because it allows contingency-table analysis with a larger n.

Heterogeneity testing with 2×2 tables is performed without the chi-square correction for continuity, except when both margins are fixed, in which case χ^2_c is used for the combined data while χ^2 is used for all other steps in the analysis (Cochran, 1942; Lancaster, 1949). The heterogeneity test may also be performed for contingency tables with more than two rows or columns. To test for heterogeneity, the log-likelihood ratio, G (Section 23.7), may be used instead of χ^2.

23.6 SUBDIVIDING CONTINGENCY TABLES

In Example 23.1, the analysis of a 2×4 contingency table, it was concluded that there was a significant difference in human hair-color frequencies between males and females. Expressing the percent males and percent females in each column, as in Example 23.6a, and examining Figures 23.1 and 23.2 shows that the proportion of males in the blond column is prominently less than in the other columns. (Examining the data in this fashion can be helpful, although frequencies, not proportions, are used for the hypothesis test.)

EXAMPLE 23.6a The Data of Example 23.1, Where for Each Hair Color the Percent Males and Percent Females Are Indicated

	Hair color				
Sex	*Black*	*Brown*	*Blond*	*Red*	**Total**
Male	32 (37%)	43 (40%)	16 (20%)	9 (36%)	100
Female	55 (63%)	65 (60%)	64 (80%)	16 (64%)	200
Total	87	108	80	25	300

In Example 23.1, the null hypothesis that the four hair colors are independent of sex was rejected.

Thus, it might be suspected that the significant χ^2 calculated in Example 23.1 was due largely to the frequencies in column 3 of the table. To pursue that supposition, the data in column 3 may be momentarily ignored and the remaining 2×3 table considered; this is done in Example 23.6b. The nonsignificant χ^2 for this table supports the null hypothesis that these three hair colors are independent of sex in the population from which the sample came. Then, in Example 23.6c, a 2×2 table is formed by considering blond versus all other hair colors combined. For this table, the null hypothesis of independence is rejected.

EXAMPLE 23.6b The 2×3 Contingency Table Formed from Columns 1, 2, and 4 of the Original 2×4 Table. $\hat{f}_{ij}$ Values for the Cells of the 2×3 Table Are Shown in Parentheses

H_0: The occurrence of black, brown, and red hair is independent of sex.
H_A: The occurrence of black, brown, and red hair is not independent of sex.

$\alpha = 0.05$

	Hair color			
Sex	*Black*	*Brown*	*Red*	**Total**
Male	32 *(33.2182)*	43 *(41.2364)*	9 *(9.5455)*	84
Female	55 *(53.7818)*	65 *(66.7636)*	16 *(15.4545)*	136
Total	87	108	25	220

$\chi^2 = 0.245$ with DF = 2

$\chi^2_{0.05,2} = 5.991$

Therefore, do not reject H_0.

$0.75 < P < 0.90 \quad [P = 0.88]$

EXAMPLE 23.6c The 2 × 2 Contingency Table Formed by Combining Columns 1, 2, and 4 of the Original Table

H_0: Occurrence of blond and nonblond hair color is independent of sex.
H_A: Occurrence of blond and nonblond hair color is not independent of sex.

$\alpha = 0.05$

	Hair color		
Sex	*Blond*	*Nonblond*	**Total**
Male	16	84	100
Female	64	136	200
Total	80	220	300

$\chi^2 = 8.727$

DF = 1

$\chi^2_{0.05,1} = 3.841$

Therefore, reject H_0.

$0.001 < P < 0.005 \quad [P = 0.0036]$

By the described series of subdivisions and column combinations of the original contingency table, we see evidence suggesting that, among the four hair colors in the population, blond occurs between the sexes with relative frequencies different from those of the other colors. However, it is not strictly proper to test statistical hypotheses developed after examining the data to be tested. Therefore, the analysis of a subdivided contingency table should be considered only as a guide to developing hypotheses. Hypotheses suggested by this analysis then can be tested by obtaining a new set of data from the population of interest and stating those hypotheses in advance of the testing.

23.7 THE LOG-LIKELIHOOD RATIO FOR CONTINGENCY TABLES

The log-likelihood ratio was introduced in Section 22.7, where the G statistic (sometimes called G^2) was presented as an alternative to chi-square for goodness-of-fit testing. The G test may also be applied to contingency tables (Neyman and Pearson, 1928a, 1928b; Wilks, 1935), where

$$G = 2\left[\sum_i \sum_j f_{ij} \ln\left(\frac{f_{ij}}{\hat{f}_{ij}}\right)\right], \tag{23.8}$$

which, without the necessity of calculating expected frequencies, may readily be computed as

$$G = 2\left[\sum_i \sum_j f_{ij} \ln f_{ij} - \sum_i R_i \ln R_i - \sum_j C_j \ln C_j + n \ln n\right]. \tag{23.9}$$

If common logarithms (denoted by "log") are used instead of natural logarithms (indicated as "ln"), then use 4.60517 instead of 2 prior to the left bracket. Because G is approximately distributed as χ^2, Appendix Table B.1 may be used with $(r-1)(c-1)$ degrees of freedom. In Example 23.7, the contingency table of Example 23.1 is analyzed using the G statistic, with very similar results.

EXAMPLE 23.7 The *G* Test for the Contingency Table Data of Example 23.1

H_0: Hair color is independent of sex.

H_A: Hair color is not independent of sex.

$\alpha = 0.05$

	Hair color				
Sex	*Black*	*Brown*	*Blond*	*Red*	**Total**
Male	32	43	16	9	100
Female	55	65	64	16	200
Total	87	108	80	25	300

$$\begin{aligned}
G &= 4.60517\left[\sum\sum f_{ij}\log f_{ij} - \sum R_i \log R_i - \sum C_j \log C_j + n\log n\right] \\
&= 4.60517[(32)(1.50515) + (43)(1.63347) + (16)(1.20412) + (9)(0.95424) \\
&\quad + (55)(1.74036) + (65)(1.81291) + (64)(1.80618) + (16)(1.20412) \\
&\quad - (100)(2.00000) - (200)(2.30103) - (87)(1.93952) \\
&\quad - (108)(2.03342) - (80)(1.90309) - (25)(1.39794) + (300)(2.47712)] \\
&= 4.60517(2.06518) \\
&= 9.510 \text{ with DF} = 3
\end{aligned}$$

$\chi^2_{0.05,3} = 7.815$

Therefore, reject H_0.

$0.01 < P < 0.025$ $[P = 0.023]$

In the case of a 2×2 table, the Yates correction for continuity (see Sections 23.3c and 23.3d) is applied by making each f_{ij} 0.5 closer to $\hat{f}_{ij}$. This may be accomplished (without calculating expected frequencies) as follows: If $f_{11}f_{22} - f_{12}f_{21}$ is negative, add 0.5 to f_{11} and f_{22} and subtract 0.5 from f_{12} and f_{21}; if $f_{11}f_{22} - f_{12}f_{21}$ is positive, subtract 0.5 from f_{11} and f_{22} and add 0.5 to f_{12} and f_{21}; then Equation 23.8 or 23.9 is applied using these modified values of f_{11}, f_{12}, f_{21}, and f_{22}.

Williams (1976) recommended that G be used in preference to χ^2 whenever $|f_{ij} - \hat{f}_{ij}| \geq \hat{f}_{ij}$ for any cell. Both χ^2 and G commonly result in the same conclusion for the hypothesis test, especially when n is large. When they do not, some statisticians favor employing G, and its use is found in some research reports and computer software. However, many others (e.g., Agresti, 2002: 24, 396; Agresti and Yang, 1987; Berry and Mielke, 1988; Hosmane, 1986; Hutchinson, 1979; Koehler, 1986; Larntz, 1978; Margolin and Light, 1974; Stelzl, 2000; Upton, 1982) have concluded that the χ^2 procedure is preferable to G; and generally it more closely refers to the probability of a Type I error.

23.8 MULTIDIMENSIONAL CONTINGENCY TABLES

Thus far, this chapter has considered two-dimensional contingency tables (tables with rows and columns as the two dimensions), where each of the two dimensions represents a nominal-scale variable. However, categorical data may also be collected and tabulated with respect to three or more nominal-scale variables, resulting in what are called multidimensional contingency tables—that is, tables with three or more dimensions (e.g., see Christensen, 1990; Everitt, 1992: Chapter 4; Fienberg, 1970, 1980; Goodman, 1970; Simonoff, 2003: Chapter 8). An example would be data from a study similar to that in Example 23.1, but where eye color is a third variable—in addition to the variables hair color and sex.

As the number of dimensions increases, so does the complexity of the analysis, and various interactions of variables are potentially of interest. Multidimensional contingency tables may be analyzed by extensions of the χ^2 and G testing discussed earlier in this chapter, as will be indicated in this section. Computer-program libraries often include provision for the analysis of such tables, including by utilizing what are known as *log-linear models*,* a large body of statistical procedures (e.g., see Everitt, 1992: Chapter 5; Fienberg, 1970, 1980; Howell, 2007: Chapter 17; Kennedy, 1992; Knoke and Burke, 1980; Tabachnik and Fidell, 2001: Chapter 7).

Figure 23.3 shows a three-dimensional contingency table. The three "rows" are species, the four "columns" are geographic locations, and the two "tiers" (or "layers") are presence and absence of a disease. If a sample is obtained containing individuals of these species, from these locations, and with and without the disease in question, then observed frequencies can be recorded in the 24 cells of this $3 \times 4 \times 2$ contingency table. We shall refer to the observed frequency in row i, column j, and tier l as f_{ijl}. We shall refer to the number of rows, columns, and tiers as r, c, and t, respectively. The sum of the frequencies in row i will be designated R_i, the sum in column j as C_j, and the sum in tier l as T_l. Friendly (1994, 1999), Hartigan and Kleiner (1981, 1984),

*Log-linear models are mathematical representations that also underlie analysis of variance (Chapters 10, 12, 14, 15, and 16) and multiple regression (Chapter 20). The term *log-linear model* was introduced in 1969 by Y. M. M. Bishop and S. E. Fienberg (David, 1995).

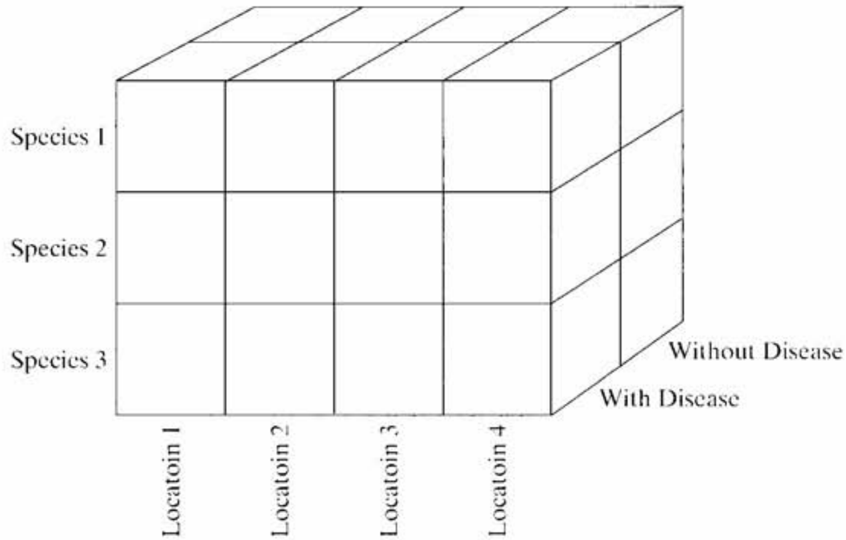

FIGURE 23.3: A three-dimensional contingency table, where the three rows are species, the four columns are locations, and the two tiers are occurrence of a disease. An observed frequency, f_{ijl}, will be recorded in each combination of row, column, and tier.

and Simonoff (2003: 329) discuss mosaic displays for contingency tables with more than two dimensions, and such graphical presentations can make multidimensional contingency table data easier to visualize and interpret than if they are presented only in tabular format.

Example 23.8 presents a $2 \times 2 \times 2$ contingency table where data (f_{ijl}) are collected as described previously, but only for two species and two locations. Note that throughout the following discussions the sum of the expected frequencies for a given row, column, or tier equals the sum of the observed frequencies for that row, column, or tier.

EXAMPLE 23.8 Test for Mutual Independence in a $2 \times 2 \times 2$ Contingency Table

H_0: Disease occurrence, species, and location are all mutually independent in the population sampled.

H_A: Disease occurrence, species, and location are not all mutually independent in the population sampled.

The observed frequencies (f_{ijl}):

	Disease present		**Disease absent**		**Species totals**
	Location 1	*Location 2*	*Location 1*	*Location 2*	$(r = 2)$
Species 1	44	12	38	10	$R_1 = 104$
Species 2	28	22	20	18	$R_2 = 88$
Disease totals $(t = 2)$:	$T_1 = 106$		$T_2 = 86$		*Grand total:*
Location totals: $(c = 2)$:	$C_1 = 130, C_2 = 62$				$n = 192$

The expected frequencies ($\hat{f}_{ijl}$):

	Disease present		Disease absent		
	Location 1	*Location 2*	*Location 1*	*Location 2*	**Species totals**
Species 1	38.8759	18.5408	31.5408	15.0425	$R_1 = 104$
Species 2	32.8950	15.6884	26.6884	12.7283	$R_2 = 88$
Disease totals:	$T_1 = 106$		$T_2 = 86$		*Grand total:*
Location totals:	$C_1 = 130, C_2 = 62$				$n = 192$

$$\chi^2 = \sum\sum\sum \frac{(f_{ijl} - \hat{f}_{ijl})^2}{\hat{f}_{ijl}}$$

$$\chi^2 = \frac{(44 - 38.8759)^2}{38.8759} + \frac{(12 - 18.5408)^2}{18.5408} + \frac{(38 - 31.5408)^2}{31.5408}$$
$$+ \frac{(10 - 15.0425)^2}{15.0425} + \frac{(28 - 32.8950)^2}{32.8950} + \frac{(22 - 15.6884)^2}{15.6884}$$
$$+ \frac{(20 - 26.6884)^2}{26.6884} + \frac{(18 - 12.7283)^2}{12.7283}$$
$$= 0.6754 + 2.3075 + 1.3228 + 1.6903 + 0.7284 + 2.5392$$
$$+ 1.6762 + 2.1834$$
$$= 13.123$$

$$\nu = rct - r - c - t + 2 = (2)(2)(2) - 2 - 2 - 2 + 2 = 4$$

$$\chi^2_{0.05,4} = 9.488$$

Reject H_0.

$0.01 < P < 0.025$ $[P = 0.011]$

(a) Mutual Independence. We can test more than one null hypothesis using multidimensional contingency-table data. An overall kind of hypothesis is that which states mutual independence among all the variables. Another way of expressing this H_0 is that there are no interactions (either three-way or two-way) among any of the variables. For this hypothesis, the expected frequency in row i, column j, and tier l is

$$\hat{f}_{ijl} = \frac{R_i C_j T_l}{n^2}, \tag{23.10}$$

where n is the total of all the frequencies in the entire contingency table.

In Example 23.8 this null hypothesis would imply that presence or absence of the disease occurred independently of species and location. For three dimensions, this null hypothesis is tested by computing

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{l=1}^{t} \frac{(f_{ijl} - \hat{f}_{ijl})^2}{\hat{f}_{ijl}}, \tag{23.11}$$

which is a simple extension of the chi-square calculation for a two-dimensional table (by Equation 23.1). The degrees of freedom for this test are the sums of the degrees of freedom for all interactions:

$$\nu = (r - 1)(c - 1)(t - 1) + (r - 1)(c - 1) + (r - 1)(t - 1) + (c - 1)(t - 1), \tag{23.12}$$

which is equivalent to

$$\nu = rct - r - c - t + 2. \tag{23.13}$$

(b) Partial Independence. If the preceding null hypothesis is not rejected, then we conclude that all three variables are mutually independent and the analysis proceeds no further. If, however, H_0 is rejected, then we may test further to conclude between which variables dependencies and independencies exist. For example, we may test whether one of the three variables is independent of the other two, a situation known as *partial independence.**

For the hypothesis of rows being independent of columns and tiers, we need total frequencies for rows and total frequencies for combinations of columns and tiers. Designating the total frequency in column j and tier l as $(CT)_{jl}$, expected frequencies are calculated as

$$\hat{f}_{ijl} = \frac{R_i(CT)_{jl}}{n}, \tag{23.14}$$

and Equation 23.11 is used with degrees of freedom

$$\nu = (r - 1)(c - 1)(t - 1) + (r - 1)(c - 1) + (r - 1)(t - 1), \tag{23.15}$$

which is equivalent to

$$\nu = rct - ct - r + 1. \tag{23.16}$$

For the null hypothesis of columns being independent of rows and tiers, we compute expected frequencies using column totals, C_j, and the totals for row and tier combinations, $(RT)_{il}$:

$$\hat{f}_{ijl} = \frac{C_j(RT)_{il}}{n}, \tag{23.17}$$

and

$$\nu = rct - rt - c + 1. \tag{23.18}$$

And, for the null hypothesis of tiers being independent of rows and columns, we use tier totals, T_l, and the totals for row and column combinations, $(RC)_{ij}$:

$$\hat{f}_{ijl} = \frac{T_l(RC)_{ij}}{n}; \tag{23.19}$$

$$\nu = rct - rc - t + 1. \tag{23.20}$$

*A different hypothesis is that of *conditional independence*, where two of the variables are said to be independent in each level of the third (but each may have dependence on the third). This is discussed in the references cited at the beginning of this section.

In Example 23.9, all three pairs of hypotheses for partial independence are tested. In one of the three (the last), H_0 is not rejected; thus we conclude that presence of disease is independent of species and location. However, the hypothesis test of Example 23.8 concluded that all three variables are not independent of each other. Therefore, we suspect that species and location are not independent. The independence of these two variables may be tested using a two-dimensional contingency table, as described earlier, in Section 23.3, and demonstrated in Example 23.10. In the present case, the species-location interaction is tested by way of a 2×2 contingency table, and we conclude that these two factors are not independent (i.e., species occurrence depends on geographic location).

In general, hypotheses to be tested should be stated before the data are collected. But the hypotheses proposed in Example 23.10 were suggested *after* the data were examined. Therefore, instead of accepting the present conclusion of the analysis in Example 23.10, such a conclusion should be reached by testing this pair of hypotheses upon obtaining a new set of data from the population of interest and stating the hypotheses in advance of the testing.

EXAMPLE 23.9 Test for Partial Independence in a 2 × 2 × 2 Contingency Table. As the H_0 of Overall Independence Was Rejected in Example 23.8, We May Test the Following Three Pairs of Hypotheses

H_0: Species is independent of location and disease.

H_A: Species is not independent of location and disease.

The expected frequencies ($\hat{f}_{ijl}$):

	Disease present		Disease absent		
	Location 1	*Location 2*	*Location 1*	*Location 2*	**Species totals**
Species 1	39.0000	18.4167	31.4167	15.1667	$R_1 = 104$
Species 2	33.0000	15.5833	26.5833	12.8333	$R_2 = 88$
Location and disease totals:	$(CT)_{11} = 72$	$(CT)_{12} = 34$	$(CT)_{21} = 58$	$(CT)_{22} = 28$	*Grand total:* $n = 192$

$$\chi^2 = \frac{(44 - 39.0000)^2}{39.0000} + \frac{(12 - 18.4167)^2}{18.4167} + \frac{(38 - 31.4167)^2}{31.4167} + \cdots + \frac{(18 - 12.8333)^2}{12.8333}$$

$$= 0.6410 + 2.2357 + 1.3795 + 1.7601 + 0.7576 + 2.6422 + 1.6303 + 2.0801$$

$$= 13.126$$

$$\nu = rct - ct - r + 1 = (2)(2)(2) - (2)(2) - 2 + 1 = 3$$

$$\chi^2_{0.05,3} = 7.815$$

Reject H_0. Species is not independent of location and presence of disease.

$$0.005 < P < 0.001 \quad [P = 0.0044]$$

H_0: Location is independent of species and disease.

H_A: Location is not independent of species and disease.

The expected frequencies ($\hat{f}_{ijl}$):

	Disease present		**Disease absent**		
	Species 1	*Species 2*	*Species 1*	*Species 2*	**Location totals**
Location 1	37.91677	33.8542	32.5000	25.7292	$C_1 = 130$
Location 2	18.0833	16.1458	15.5000	12.2708	$C_2 = 62$
Species and disease totals:	$(RT)_{11} = 56$	$(RT)_{12} = 50$	$(RT)_{21} = 48$	$(RT)_{22} = 38$	*Grand total:* $n = 192$

$$\chi^2 = \frac{(44 - 37.9167)^2}{37.9167} + \frac{(28 - 33.8542)^2}{33.8542} + \cdots + \frac{(18 - 12.2708)^2}{12.2708}$$

$$= 0.9760 + 1.0123 + 0.9308 + 1.2757 + 2.0464 + 2.1226$$

$$+ 1.9516 + 2.6749$$

$$= 12.990$$

$$\nu = rct - rt - c + 1 = (2)(2)(2) - (2)(2) - 2 + 1 = 3$$

$$\chi^2_{0.05,3} = 7.815$$

Reject H_0. Location is not independent of species and presence of disease.

$$0.001 < P < 0.005 \quad [P = 0.0047]$$

H_0: Presence of disease is independent of species and location.

H_A: Presence of disease is not independent of species and location.

The expected frequencies ($\hat{f}_{ijl}$):

	Species 1		**Species 2**		
	Location 1	*Location 2*	*Location 1*	*Location 2*	**Disease totals**
Disease present	45.2708	12.1458	26.5000	22.0833	$T_1 = 106$
Disease absent	36.7292	9.8542	21.5000	17.9167	$T_2 = 86$
Species and location totals:	$(RC)_{11} = 82$	$(RC)_{12} = 22$	$(RC)_{21} = 48$	$(RC)_{22} = 40$	*Grand total:* $n = 192$

$$\chi^2 = \frac{(44 - 45.2708)^2}{45.2708} + \frac{(12 - 12.1458)^2}{12.1458} + \cdots + \frac{(18 - 17.9167)^2}{17.9167}$$
$$= 0.0357 + 0.0018 + 0.0849 + 0.0003 + 0.0440 + 0.0022$$
$$+ 0.1047 + 0.0004$$
$$= 0.274$$

$\nu = rct - rc - t + 1 = (2)(2)(2) - (2)(2) - 2 + 1 = 3$

$\chi^2_{0.05,3} = 7.815$

Do not reject H_0.

$0.95 < P < 0.975 \quad [P = 0.96]$

EXAMPLE 23.10 Test for Independence of Two Variables, Following Tests for Partial Dependence

The hypothesis test of Example 23.8 concluded that all three variables are not mutually independent, while the last test in Example 23.9 concluded that presence of disease is independent of species and location. Therefore, it is desirable (and permissible) to test the following two-dimensional contingency table:

H_0: Species occurrence is independent of location.

H_A: Species occurrence is not independent of location.

	Location 1	*Location 2*	*Total*
Species 1	82	22	104
Species 2	48	40	88
Total	130	62	192

$\chi^2 = 12.874$

$\nu = (r - 1)(c - 1) = 1$

$\chi^2_{0.05,1} = 3.841$

Reject H_0.

$P < 0.001 \quad [P = 0.00033]$

(c) The Log-Likelihood Ratio. The log-likelihood ratio of Section 23.7 can be expanded to contingency tables with more than two dimensions. While some authors have chosen this procedure over chi-square testing and it is found in some statistical computer packages, others (e.g., Haber, 1984; Hosmane, 1987; Koehler, 1986; Larntz, 1978; Rudas, 1986; and Stelzl, 2000) have concluded that χ^2 is preferable. With χ^2 in contrast to G, the probability of a Type I error is generally closer to α.

EXERCISES

3.1. Consider the following data for the abundance of a certain species of bird.

(a) Using chi-square, test the null hypothesis that the ratio of numbers of males to females was the same in all four seasons.

(b) Apply the G test to that hypothesis.

Sex	*Spring*	*Summer*	*Fall*	*Winter*
Males	163	135	71	43
Females	86	77	40	38

3.2. The following data are frequencies of skunks found with and without rabies in two different geographic areas.

(a) Using chi-square, test the null hypothesis that the incidence of rabies in skunks is the same in both areas.

(b) Apply the G test to that hypothesis.

Area	*With rabies*	*Without rabies*
E	14	29
W	12	38

3.3. Data were collected as in Exercise 23.2, but with the additional tabulation of the sex of each skunk recorded, as follows. Test for mutual independence; and, if H_0 is rejected, test for partial independence.

Area	**With rabies**		**Without rabies**	
	Male	*Female*	*Male*	*Female*
E	42	33	55	63
W	84	51	34	48

23.4. A sample of 150 was obtained of men with each of three types of cancer, and the following data are the frequencies of blood types for the men.

(a) Using chi-square, test the null hypothesis that, in the sampled population, the frequency distribution of the three kinds of cancer is the same for men with each of the four blood types (which is the same as testing the H_0 that the frequency distribution of the four blood types is the same in men with each of the three kinds of cancer).

(b) Apply the G test to the same hypothesis.

	Blood type				
Cancer type	O	A	B	AB	**Total**
Colon	61	65	18	6	$R_1 = 150$
Lung	69	57	15	9	$R_2 = 150$
Prostate	73	60	12	5	$R_3 = 150$