Example 1

Happiness

Picture the Scenario

What contributes to your overall happiness? Is it love? Your health? Your friendships? The amount of money you make?

To investigate which variables are associated with happiness, we can use data from the General Social Survey (GSS). In each survey, the GSS asks, "Taken all together, would you say that you are very happy, pretty happy, or not too happy?" Table 11.1 uses the 2008 survey to cross-tabulate happiness with family income, here measured as the response to the question, "Compared with American families in general, would you say that your family income is below average, average, or above average?"

Table 11.1 Happiness and Family Income, from 2008 General Social Survey

	Happiness			
Income	**Not Too Happy**	**Pretty Happy**	**Very Happy**	**Total**
Above average	26	233	164	423
Average	117	473	293	883
Below average	172	383	132	687

Questions to Explore

- How can you determine if there is an association between happiness and family income in the population of all adult Americans?
- If there is an association, what is its nature? For example, do people with above-average family income tend to be happier than people with below-average family income?
- Can you think of a variable that might have a stronger association with happiness than family income?

Thinking Ahead

Both variables in Table 11.1 are categorical. The focus of this chapter is learning how to describe associations between categorical variables. To do this, we'll continue to apply the basic tools of statistical inference. These inferential methods help us answer questions such as, "Do people with higher family incomes tend to be happier?" and "Are married people happier than unmarried people?" We'll analyze Table 11.1 in Examples 3 and 4 and other data on happiness in Example 8 and in the exercises.

Example 2

Belief in Life After Death and Race

Picture the Scenario

Table 11.4 cross-tabulates belief in life after death with race, using data from the 2008 GSS. The table also shows sample conditional distributions for life after death, given race, and the overall percentages for the life after death categories.

Table 11.4 Sample Conditional Distributions for Belief in Life After Death, Given Race

	Postlife		
Race	**Yes**	**No**	**Total**
White	1132 (82.1%)	247 (17.9%)	1379 (100%)
Black	203 (82.2%)	44 (17.8%)	247 (100%)
Other	120 (74.5%)	41 (25.5%)	161 (100%)
Overall	1455 (81.4%)	332 (18.6%)	1787 (100%)

Source: Data from CSM, UC Berkeley.

Question to Explore

Are race and belief in life after death independent or dependent?

Think It Through

The conditional distributions in Table 11.4 are similar but not exactly identical. So it is tempting to conclude that the variables are dependent. For example, if A = yes for belief in life after death, we estimate $P(A \mid \text{white}) = 1132/1379 = 0.821$, $P(A \mid \text{black}) = 0.822$, $P(A \mid \text{other}) = 0.745$, whereas we estimate that $P(A) = 1455/1787 = 0.814$. However, the definition of independence between variables refers to the *population*. Since Table 11.4 refers to a *sample* rather than a population, it provides evidence but does not definitively tell us whether these variables are independent or dependent.

Even if the variables *were* independent, we would not expect the *sample* conditional distributions to be identical. Because of sampling variability, each sample percentage typically differs somewhat from the true population percentage. We would expect to observe *some* differences between sample conditional distributions such as we see in Table 11.4 even if *no* differences exist in the population.

Insight

If the observations in Table 11.4 were the entire population, the variables would be dependent. But the association would not necessarily be practically important, because the percentages are similar from one category to another of race. The next section presents a significance test of the hypothesis that the variables are independent.

Try Exercise 11.2

Example 3

Happiness and Family Income

Picture the Scenario

Table 11.5 showed the observed and expected cell counts for the data on family income and happiness. Table 11.6 shows MINITAB output for the chi-squared test. The margin on the next page shows screen shots from the TI-83+/84. The matrix [B] referred to with the word **Expected** in the second screen shot contains the expected cell counts.

Table 11.6 Chi-Squared Test of Independence of Happiness and Family Income

This table shows how MINITAB reports, for each cell, the observed and expected cell counts and the contribution [(observed − expected)^2/expected] to X^2, as well as the overall X^2 value and its P-value for testing H_0: independence.

Rows: income   Columns: happy

	not	pretty	very		
above	26	233	164	423	Observed cell count
	66.86	231.13	125.01		Expected cell count
	24.968	0.015	12.160		Contribution to X^2
average	117	473	293	883	
	139.56	482.48	260.96		
	3.647	0.186	3.935		
below	172	383	132	687	
	108.58	375.39	203.03		
	37.039	0.154	24.851		
Total	315	1089	589	1993	

```
Cell contents:        Count
                      Expected count
                      Contribution to Chi-square
Pearson Chi-Square = 106.955, DF = 4, P-Value = 0.000
```

Questions to Explore

a. State the null and alternative hypotheses for this test.

b. Report the X^2 statistic and explain how it was calculated.

Think It Through

a. The hypotheses for the chi-squared test are

H_0: Happiness and family income are independent.

H_a: Happiness and family income are dependent (associated).

The alternative hypothesis is that there's an association between happiness and family income.

b. Finding X^2 involves a fair amount of computation, and it's best to let software do the work for us. Beneath the observed and expected cell counts in a cell, MINITAB reports the contribution of that cell to the X^2 statistic. For the first cell (above average income and not too happy), for instance,

$$(\text{observed count} - \text{expected count})^2/\text{expected count} = (26 - 66.86)^2/66.86 = 25.0.$$

For all nine cells,

$$X^2 = \frac{(26-66.86)^2}{66.86} + \frac{(233-231.13)^2}{231.13} + \frac{(164-125.01)^2}{125.01} + \frac{(117-139.56)^2}{139.56} + \frac{(473-482.48)^2}{482.48}$$
$$+ \frac{(293-260.96)^2}{260.96} + \frac{(172-108.56)^2}{108.56} + \frac{(383-375.39)^2}{375.39} + \frac{(132-203.03)^2}{203.03}$$
$$= 24.968 + 0.015 + 12.160 + 3.647 + 0.186 + 3.935 + 37.039 + 0.154 + 24.851 = 106.955.$$

Insight

Table 11.6 has some large differences between observed and expected cell counts, so X^2 is large. The larger the X^2 value, the greater the evidence against H_0: independence and in support of the alternative hypothesis that happiness and income are associated. We next learn how to interpret the magnitude of the X^2 test statistic, so we know what is small and what is large and how to find the P-value.

Try Exercise 11.11

Example 4

Happiness and Income

Picture the Scenario

For testing the null hypothesis of independence of happiness and family income, Table 11.6 in Example 3 reported a test statistic value of $X^2 = 106.955$.

Questions to Explore

a. What is the P-value for the chi-squared test of independence for these data?

b. State your decision for a significance level of 0.05 and interpret in context.

Think It Through

a. Since the table has $r = 3$ rows and $c = 3$ columns, $df = (r - 1) \times (c - 1) = 4$. In Table 11.7, for $df = 4$ the largest chi-squared value shown is 18.47. It has tail probability $= 0.001$ (see the margin figure). Since $X^2 = 106.955$ falls well above this, it has a smaller right-tail probability. Thus, the P-value is < 0.001. The actual P-value would be 0 to many decimal places. MINITAB reports P-value $= 0.000$, as Table 11.6 showed.

b. Since the P-value is below 0.05, we can reject H_0. Based on this sample, we have evidence to support that an association exists between happiness and family income in the population.

Insight

The extremely small P-value provides very strong evidence against H_0: independence. If the variables were independent, it would be highly unusual for a random sample to have this large a chi-squared statistic.

Try Exercise 11.10

Example 5

Aspirin and Cancer Death Rates Revisited

Picture the Scenario

Examples 2–4 in Chapter 10 discussed a meta-study on the effects of aspirin on cancer death rates, in which subjects were randomly assigned to take aspirin or placebo regularly. That example analyzed a 2×2 contingency table that compared the proportions of cancer deaths during a five-year period for those who took placebo and for those who took aspirin. Table 11.9 shows the data again with the expected cell counts, as part of MINITAB output for a chi-squared test of independence (or homogeneity).

Table 11.9 Annotated MINITAB Output for Chi-Squared Test of Independence of Group (Placebo, Aspirin) and Whether or Not Subject Died of Cancer

The same P-value results as with a two-sided Z test comparing the two population proportions.

```
Rows:  group  Columns:  heart

            yes        no        Total
placebo     347        11188     11535     ← Cell counts
            304        11230               ← Expected cell counts
aspirin     327        13708     14035
            370        13665
Cell Contents:         Count
                       Expected count
Pearson  Chi−Square = 11.35,  DF = 1,  P−Value = 0.001
```

Denote the population proportion of cancer deaths by p_1 for the placebo treatment and by p_2 for the aspirin treatment. For H_0: $p_1 = p_2$, the z test of Section 10.1 has a test statistic value of $z \approx 3.45$ (Exercise 10.8). Its P-value is 0.001 for H_a: $p_1 \neq p_2$.

Questions to Explore

a. What are the hypotheses for the chi-squared test for these data?

b. Report the test statistic and P-value for the chi-squared test. How do these relate to results from the z test comparing the proportions?

Think It Through

a. The null hypothesis is that whether or not a someone dies of certain cancers is not associated with whether he or she takes placebo or aspirin. This is equivalent to H_0: $p_1 = p_2$, the population proportion of cancer deaths being the same for each group. The alternative hypothesis is that there's an association. This is equivalent to H_a: $p_1 \neq p_2$.

b. From Table 11.9, more subjects taking placebo died of cancer than we would expect if the variables were not associated (347 observed versus 304 expected). Fewer subjects taking aspirin died of cancer than we would expect if the variables were not associated (327 observed versus 370 expected). Table 11.9 reports $X^2 = 11.35$. This approximates the square of the z test statistic, $X^2 = z^2 = (3.45)^2 = 11.9$, which is close to the 11.35 value. (Note that the numerical values are not exact due to rounding error.) Table 11.9 reports a chi-squared P-value of 0.001. This is very strong evidence that the population proportions of cancer deaths differed for those taking aspirin and for those taking placebo. The sample proportions suggest that the aspirin group had a lower rate of cancer deaths than the placebo group.

Insight

For 2×2 tables, $df = (r - 1) \times (c - 1) = (2 - 1) \times (2 - 1) = 1$. Whenever a statistic has a standard normal distribution, the square of that statistic has a chi-squared distribution with $df = 1$. The chi-squared test and the two-sided z test necessarily have the same P-value and the same conclusion. Here, both the z and X^2 statistics show extremely strong evidence against the null hypothesis of equal population proportions. An advantage of the z test over X^2 is that it also can be used with one-sided alternative hypotheses. The direction of the effect is lost in squaring z and using X^2.

Try Exercise 11.18

Example 6

Student Stress, Depression, and Gender

Picture the Scenario

Every year, the Higher Education Research Institute at UCLA conducts a large-scale survey of college freshmen on a variety of issues. From the survey of 283,000 freshmen in 2002, Table 11.13 compares females and males on the percent who reported feeling frequently stressed (overwhelmed by all they have to do) and the percent who reported feeling frequently depressed during the past year. P-values for chi-squared tests of independence were 0.000 (rounded) for each data set.

Table 11.13 Conditional Distributions of Stress and Depression, by Gender

	Stress				Depression		
Gender	Yes	No	Total	Gender	Yes	No	Total
Female	35%	65%	100%	Female	8%	92%	100%
Male	16%	84%	100%	Male	6%	94%	100%

Question to Explore

Which response variable, stress or depression, has the stronger sample association with gender?

Think It Through

The difference of proportions between females and males was $0.35 - 0.16 = 0.19$ for feeling stressed. It was $0.08 - 0.06 = 0.02$ for feeling depressed. Since 0.19 is much larger than 0.02, there is evidence of a greater difference between reports of females and males on their feeling stress than on depression. In the sample, stress has the stronger association with gender.

Insight

Although the difference of proportions can be as large as 1 or -1, in practice in comparisons of groups, it is rare to find differences near these limits.

Try Exercise 11.24

Example 7

Seat Belt Use and Outcome of Auto Accidents

Picture the Scenario

Based on records of automobile accidents in a recent year, the Department of Highway Safety and Motor Vehicles in Florida reported Table 11.14 on the counts of nonfatalities and fatalities. The data have $X^2 = 2338$, with P-value $= 0.00000. \ldots$

Table 11.14 Outcome of Auto Accident by Whether or Not Subject Wore Seat Belt

	Outcome		
Wore Seat Belt	**Survived**	**Died**	**Total**
Yes	412,368	510	412,878
No	162,527	1601	164,128

Source: Department of Highway Safety and Motor Vehicles, Florida.

Think It Through

In Table 11.14, the adverse outcome is death, rather than survival. The relative risk is formed for that outcome. For those who wore a seat belt, the proportion who died equaled $510/412{,}878 = 0.00124$. For those who did not wear a seat belt, the proportion who died was $1601/164{,}128 = 0.00975$. The relative risk is the ratio, $0.00124/0.00975 = 0.127$. The proportion of subjects wearing a seat belt who died was 0.127 times the proportion of subjects not wearing a seat belt who died.

Equivalently, since $0.00975/0.00124 = 1/0.127 = 7.9$, the proportion of subjects *not wearing* a seat belt who died was nearly eight times the proportion of subjects *wearing* seat belts who died. This reciprocal value is the relative risk when the rows are interchanged, with no for seat-belt use in row 1.

Insight

The ordering of rows is arbitrary. Many people find it easier to interpret the relative risk for the ordering for which its value is *above* 1.0. A relative risk of 7.9 represents a strong association. This is far from the value of 1.0 that would occur if the proportion of deaths were the same for each group. Wearing a seat belt has a practically significant effect in enhancing the chance of surviving an auto accident.

Try Exercises 11.27, part c, and 11.30

Example 8

General Happiness and Marital Happiness

Picture the Scenario

Earlier in the chapter we studied the association between happiness and family income. For married subjects, a possible predictor of happiness is their reported happiness with their marriage. Table 11.15 cross-tabulates these variables, using data from the 2008 GSS. It also reports conditional distributions with happiness as the response variable. The data have chi-squared statistic $X^2 = 263.9$, with $df = 4$ and a P-value of $0.0000. \ldots$

Table 11.15 Happiness and Marital Happiness

The percentages are the conditional distributions of happiness in each row. When percentages are compared in rows 1 and 3, the highlighted cells reveal a strong association.

	Happiness			
Marital Happiness	**Not Too Happy**	**Pretty Happy**	**Very Happy**	**Total**
Not Too Happy	13 (43.3%)	15 (50.0%)	2 (6.7%)	30 (100%)
Pretty Happy	44 (12.8%)	259 (75.5%)	40 (11.7%)	343 (100%)
Very Happy	24 (4.0%)	215 (36.1%)	356 (59.8%)	595 (100%)

Source: Data from CSM, UC Berkeley.

Question to Explore

Describe the association between marital happiness and general happiness, using the not too happy and very happy categories.

Think It Through

Let's consider the Very Happy category of happiness. The difference between the proportions in this category for the Not Too Happy and Very Happy categories of marital happiness (that is, rows 1 and 3) is $0.067 - 0.598 = -0.531$. Those who are not too happy with their marriage are much less likely to be very happy overall. Consider next the Not Too Happy category of happiness. The difference between the proportions in this category for the Not Too Happy and Very Happy categories of marital happiness is $0.433 - 0.04 = 0.393$. This is also a substantial difference.

Likewise, we could use the *ratio* of proportions. For instance, consider the Not Too Happy category of happiness as an undesirable outcome. The relative risk of this outcome, comparing the Not Too Happy and Very Happy categories of marital happiness, is $0.433/0.04 = 10.825$. The estimated proportion who are not too happy is nearly 11 times larger for those who have not too happy marriages than for those who have very happy marriages.

This ratio of 10.825 is very far from 1.0, indicating that this part of the table reveals an extremely strong association.

Insight

It is unusual to find such strong associations. For instance, family income also has a substantial association with happiness, but it is not as strong. Using Table 11.1 (partly shown in the margin), you can check that the relative risk of being not too happy equals 4.07 when you compare the below-average and above-average levels of income. This is much less than the relative risk of 10.825 and is a weaker association.

Try Exercises 11.26 and 11.27

Example 9

Religiosity and Gender

Picture the Scenario

Table 11.16 shows some artificial tables comparing females and males on their religious attendance. Let's look at some actual data on religiosity. Table 11.17 displays observed and expected cell counts and the standardized residuals for the association between gender and response to the question, "To what extent do you consider yourself a religious person?" The possible responses were (very religious, moderately religious, slightly religious, not religious at all). The data are from the 2008 GSS. The table is in the form provided by MINITAB (which actually calls the standardized residuals "adjusted residuals").

Table 11.17 Religiosity by Gender, With Expected Counts and Standardized Residuals

Large positive standardized residuals (in green) indicate strong evidence of a higher population cell proportion than expected under independence. Large negative standardized residuals (in red) indicate strong evidence of a lower population proportion than expected.

Rows: gender	Columns: religiosity					
	Very	Moderately	Slightly	Not	All	
Female	241	499	208	131	1079	← Counts
	201.8	454.8	251.4	171.0	1079.0	← Expected counts
	4.513	4.016	−4.606	−4.916		← Standardized residuals
Male	133	344	258	186	921	
	172.2	388.2	214.6	146.0	921.0	
	−4.513	−4.016	4.606	4.916		
All	374	843	466	317	2000	

Cell Contents: Count
Expected count
Adjusted residual ← This is the name MINITAB uses for standardized residuals.

Source: Data from CSM, UC Berkeley.

Questions to Explore

a. How would you interpret the standardized residual of 4.513 in the first cell?

b. Interpret the standardized residuals in the entire table.

Think It Through

a. The cell for females who are very religious has observed count = 241, expected count = 201.8, and the standardized residual = 4.513. The difference between the observed and expected counts is more than 3 standard errors. In fact it exceeds 4 in absolute value, which tells us this cell shows a much greater discrepancy than we'd expect if the variables were truly independent. There's strong evidence that the population proportion for that cell (female and very religious) is higher than independence predicts.

b. Table 11.17 exhibits large positive residuals above 4 in four out of eight cells. These are the cells in which the observed count is much larger than the expected count. Those cells had strong evidence of more frequent occurrence than if religiosity and gender were independent. The table exhibits large negative residuals in the other four cells. In these cells, the observed count is much smaller than the expected count. There's strong evidence that the population has fewer subjects in these cells than if the variables were independent.

Insight

The standardized residuals help to describe the pattern of this association: Compared to what we'd expect if religiosity and gender were independent it appears based on this sample, there are more females who are very religious or moderately religious and more males who are slightly religious or not at all religious.

Try Exercise 11.33

Example 10

A Tea-Tasting Experiment

Picture the Scenario

In introducing his small-sample test of independence in 1935, Fisher described the following experiment: Fisher enjoyed taking a break in the afternoon for tea. One day, his colleague Dr. Muriel Bristol claimed that when drinking tea she could tell whether the milk or the tea had been added to the cup first (she preferred milk first). To test her claim, Fisher asked her to taste eight cups of tea, four of which had the milk added first and four of which had the tea added first. She was told there were four cups of each type and was asked to indicate which four had the milk added first. The order of presenting the cups to her was randomized.

Table 11.18 shows the result of the experiment. Dr. Bristol predicted three of the four cups correctly that had milk poured first.

Table 11.18 Result of Tea-Tasting Experiment

The table cross-tabulates what was actually poured first (milk or tea) by what Dr. Bristol predicted was poured first. She had to indicate which four of the eight cups had the milk poured first.

	Prediction		
Actual	**Milk**	**Tea**	**Total**
Milk	3	1	4
Tea	1	3	4
Total	4	4	8

For those cases where milk was poured first (row 1), denote p_1 as the probability of guessing that milk was added first. For those cases where tea was poured first, denote p_2 as the probability of guessing that milk was added first. If she really could predict when milk was poured first better than with random guessing, then $p_1 > p_2$. So we shall test H_0: $p_1 = p_2$ (prediction independent of actual pouring order) against H_a: $p_1 > p_2$.

Questions to Explore

a. Show the sample space of the five possible contingency table outcomes that could have occurred for this experiment.

b. Table 11.19 shows the result of using SPSS software to conduct the chi-squared test of the null hypothesis that her predictions were independent of the actual order of pouring. Is this test and its P-value appropriate for these data? Why or why not?

Table 11.19 Result of Fisher's Exact Test for Tea-Tasting Experiment

The chi-squared P-value is listed under Asymp. Sig. and the Fisher's exact test P-values are listed under Exact Sig. "Sig" is short for *significance* and "asymp." is short for *asymptotic.*

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.000	1	.157		
Fisher's Exact Test				.486	.243

4 cells (100.0%) have expected count less than 5.

c. Table 11.19 also shows the result for Fisher's exact test of the same null hypothesis. The one-sided version of the test pertains to the alternative that her predictions are better than random guessing. How does the one-sided P-value relate to the probabilities for the five tables in part a? Does this P-value suggest that she had the ability to predict better than random guessing?

Think It Through

a. The possible sample tables have four observations in each row, because there were four cups with milk poured first and four cups with tea poured first. The tables also have four observations in each column, because Dr. Bristol was told there were four cups of each type and was asked to predict which four had the milk added first. Of the four cups where milk was poured first (row 1 of the table), the number she could predict correctly is 4, 3, 2, 1, or 0. These outcomes correspond to the sample tables:

	Guess Poured First									
Poured First	**Milk**	**Tea**	**Milk**	**Tea**	**Milk**	**Tea**	**Milk**	**Tea**	**Milk**	**Tea**
Milk	4	0	3	1	2	2	1	3	0	4
Tea	0	4	1	3	2	2	3	1	4	0

These are the five possible tables with totals of 4 in each row and 4 in each column.

b. The cell counts in Table 11.18 are small. Each cell has expected count $(4 \times 4)/8 = 2$. Since they are less than 5, the chi-squared test is not appropriate. In fact, Table 11.19 warns us that all four cells have expected counts less than 5. The table reports $X^2 = 2.0$ with P-value $= 0.157$, but software provides results whether or not they are valid, and it's up to us to know when we can use a method.

c. The one-sided P-value reported by SPSS is for the alternative H_a: $p_1 > p_2$. It equals 0.243. This is the probability, presuming H_0 to be true, of the observed table (Table 11.18) and the more extreme table giving even more evidence in favor of $p_1 > p_2$. That more extreme table is the one with counts (4, 0) in the first row and (0, 4) in the second row, corresponding to guessing all four cups correctly. The P-value of 0.243 does not give much evidence against the null hypothesis.

Insight

This experiment did not establish an association between the actual order of pouring and Dr. Bristol's predictions. If she had predicted all four cups correctly, the one-sided P-value would have been 0.014. We might then believe her claim. With a larger sample, we would not need such extreme results to get a small P-value.

Try Exercise 11.41