Age Group	Frequency
18 to 24	7
25 to 34	32
35 to 44	34
45 to 54	60
55 to 64	106
65 and older	136
Total	**375**

Using data from the U.S. Census Bureau (census.gov) for 2014, the age distribution of adults in Texas was 14% between age 18 and 24, 20% between age 25 and 34, 19% between age 35 and 44, 18% between age 45 and 54, 14% between age 55 and 64, and 15% age 65 or older. Is it reasonable to conclude that one or more of the age groups buys a disproportionate share of Texas Lottery tickets? Use a chi-square goodness-of-fit test with $\alpha = 0.05$. (Hint: See Example 12.2.)

12.4 The "Global Automotive 2016 Color Popularity Report" (Axalta Coating Systems, axaltacs.com) included data on the colors for a sample of new cars sold in North America. They reported that 25% of the cars in the sample were white, 21% were black, 16% were gray, 11% were silver, 10% were red, and 17% were some other color. Suppose that these percentages were based on a random sample of 1200 new cars sold in North America. Is there convincing evidence that the proportions of new cars sold are not the same for all six of the color categories?

12.5 A popular urban legend is that more babies than usual are born during certain phases of the lunar cycle, especially near the full moon. The paper "The Effect of the Gravitation of the Moon on Frequency of Births" (*Environmental Health Insights* [2010]: 65–69) classified a random sample of 1007 births at a large hospital in Japan according to lunar phase. In each lunar cycle (27.32 days), the moon moves 360° relative to the earth. To determine lunar phase, the researchers divided the 360° in one lunar cycle into 12 phases of 30°. The sample data are summarized in the accompanying frequency table.

Lunar Phase (degrees)	Number of Births
0–30	90
31–60	81
61–90	76
91–120	87
121–150	90
151–180	76
181–210	94
211–240	79
241–270	76
271–300	80
301–330	93
331–360	85

(continued)

The researchers concluded that the frequency of births is not related to lunar cycle. Carry out a chi-square goodness-of-fit test to determine if the data are consistent with the researchers' claim. Use a significance level of 0.05 for your test.

12.6 The authors of the paper "External Factors and the Incidence of Severe Trauma: Time, Date, Season and Moon" (*Injury* [2014]: S93–S99) classified admissions to hospitals in Germany according to season. They wondered if severe trauma injuries were more common in some seasons than others. Assume that there were 1200 trauma cases in the sample and that the sample is representative of severe trauma injuries in Germany. The data in the accompanying table are consistent with summary quantities given in the paper. Do these data support the theory that the proportion of severe trauma cases is not the same for the four seasons? Test the relevant hypotheses using a significance level of 0.05.

Season				
Winter	Spring	Summer	Fall	Total
228	332	352	288	1,200

12.7 The authors of the paper "Is It Really About Me? Message Content in Social Awareness Streams" (*Computer Supported Cooperative Work 2010*) studied a random sample of 350 Twitter users. For each Twitter user in the sample, the tweets sent during a particular time period were analyzed and the Twitter user was classified into one of the following categories based on the type of messages they usually sent:

Category	Description
IS	Information sharing
OC	Opinions and complaints
RT	Random thoughts
ME	Me now (what I am doing now)
O	Other

The accompanying table gives the observed counts for the five categories (approximate values read from a graph in the paper).

Twitter Type	IS	OC	RT	ME	O
Observed count	51	61	64	101	73

Significance level: A significance level of $\alpha = 0.05$ will be used.

Formulate a plan ➤

Test statistic: $X^2 = \sum_{\text{all cells}} \frac{(\text{observed cell count} - \text{expected cell count})^2}{\text{expected cell count}}$

Assumptions: The samples were independently selected and thought to be representative of the three populations of interest. This means that it is appropriate to use the chi-square test if the sample size is large enough. All of the expected cell counts are at least 5, so the sample is large enough to proceed with the test.

Do the work ➤ Calculation:

$$X^2 = \frac{(217 - 208.795)^2}{208.795} + \cdots + \frac{(1650 - 1631.976)^2}{1631.976} = 9.59$$

P-value: The two-way table for this example has 3 rows and 2 columns, so the appropriate number of df is $(3 - 1)(2 - 1) = 2$. The calculated value of the test statistic is between 9.21 and 10.59 in the 2-df column of Appendix Table 8, so $0.005 < P\text{-value} < 0.010$.

Interpret the results ➤ Conclusion: The P-value is less than α (0.05), so H_0 is rejected. There is convincing evidence that the proportions in the survival categories are not the same for the three groups compared. Notice that there are more people who survived in the house or townhouse and first or second floor apartment categories than would have been expected if the survival proportions were the same for all three groups. This led the researchers who collected these data to conclude that there is a smaller chance of survival for people who suffer a heart attack in an apartment that is on the third or higher floor.

Most statistical software packages can calculate expected cell counts, the value of the X^2 statistic, and the associated P-value. This is illustrated in the following example.

Example 12.6 Keeping the Weight Off

The article "Daily Weigh-ins Can Help You Keep Off Lost Pounds, Experts Say" (*Associated Press*, October 17, 2005) describes an experiment in which 291 people who had lost at least 10% of their body weight in a medical weight loss program were assigned at random to one of three groups for follow-up. One group met monthly in person, one group "met" online monthly in a chat room, and one group received a monthly newsletter by mail. After 18 months, participants in each group were classified according to whether or not they had regained more than 5 pounds, resulting in the data summarized in Table 12.4.

TABLE 12.4 Observed and Expected Counts for Example 12.6

	Amount of Weight Gained		
	Regained 5 lb or Less	**Regained More Than 5 lb**	**Row Marginal Total**
In-Person	52 (41.0)	45 (56.0)	97
Online	44 (41.0)	53 (56.0)	97
Newsletter	27 (41.0)	70 (56.0)	97

Does there appear to be a difference in the weight regained proportions for the three follow-up methods? The relevant hypotheses are

Understand the context ➤

H_0: The proportions for the two weight-regained categories are the same for the three follow-up methods.

H_a: The weight-regained category proportions are not all the same for all three follow-up methods.

Significance level: $\alpha = 0.01$.

Formulate a plan ➤ Test statistic: $X^2 = \sum_{\text{all cells}} \frac{(\text{observed cell count} - \text{expected cell count})^2}{\text{expected cell count}}$

Assumptions: Table 12.4 contains the calculated expected counts, all of which are greater than 5. The subjects in this experiment were assigned at random to the treatment groups.

Do the work ➤ Calculation: Minitab output follows. For each cell, the Minitab output includes the observed cell count, the expected cell count, and the value of $\frac{(\text{observed cell count} - \text{expected cell count})^2}{\text{expected cell count}}$ for that cell (this is the contribution to the X^2 statistic for this cell). From the output, $X^2 = 13.773$.

```
Chi-Square Test
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts
                 <=5      >5    Total
In-person         52      45       97
               41.00   56.00
               2.951   2.161
Online            44      53       97
               41.00   56.00
               0.220   0.161
Newsletter        27      70       97
               41.00   56.00
               4.780   3.500
Total            123     168      291
Chi-Sq = 13.773, DF = 2, P-Value = 0.001
```

P-value: From the Minitab output, *P*-value = 0.001.

Interpret the results ➤ Conclusion: Since the *P*-value is less than α, H_0 is rejected. The data indicate that the proportions who have regained more than 5 pounds are not the same for the three follow-up methods. Comparing the observed and expected cell counts, we can see that the observed number in the newsletter group who had regained more than 5 pounds was greater than would have been expected and the observed number in the in-person group who had regained 5 or more pounds was less than would have been expected if there were no difference in the three follow-up methods.

Testing for Independence of Two Categorical Variables

A chi-square test can also be used to investigate the possibility of an association between two categorical variables in a single population. For example, television viewers in a particular city might be categorized with respect to both preferred network (ABC, CBS, NBC, or PBS) and favorite type of programming (comedy, drama, or information and news). The question of interest is often whether knowledge of the value of one variable provides any information about the value of the other variable—that is, are the two variables independent?

Continuing the example, suppose that those who favor ABC prefer the three types of programming in proportions 0.4, 0.5, and 0.1 and that these proportions are also correct for individuals favoring any of the other three networks. Then, learning an individual's preferred network provides no added information about that individual's favorite type of programming. The categorical variables *preferred network* and *favorite program type* would be independent.

To see how expected counts are obtained in this situation, recall from Chapter 6 that if two outcomes *A* and *B* are independent, then

$$P(A \text{ and } B) = P(A)P(B)$$

TABLE 12.6 Observed and Expected Counts for Example 12.7

	Self-Report	
Facial Expression	**No Pain**	**Pain**
No Pain	17 (12.81)	40 (44.19)
Pain	3 (7.19)	29 (24.81)

Do the work ➤ Calculation: $X^2 = \frac{(17 - 12.81)^2}{12.81} + \cdots + \frac{(29 - 24.81)^2}{24.81} = 4.92$

P-value: The table has 2 rows and 2 columns, so df $= (2 - 1)(2 - 1) = 1$. The entry closest to 4.92 in the 1-df column of Appendix Table 8 is 5.02, so the approximate P-value for this test is

P-value ≈ 0.025

Interpret the results ➤ Conclusion: Since the P-value is less than α, we reject H_0 and conclude that there is convincing evidence that a nurse's assessment of facial expression and self-reported pain are not independent.

Example 12.8 Exercise and Sleep Quality

The National Sleep Foundation asked each person in a representative sample of 1000 adult Americans about activity level and sleep quality ("2013 Sleep in America Poll," February 20, 2013, sleepfoundation.org/sites/default/files/RPT336%20Summary%20of%20Findings%2002%2020%202013.pdf, retrieved May 27, 2017). Survey participants were classified into one of four activity levels (none, light, moderate, and vigorous). Each participant was also classified into one of two sleep categories. Data consistent with summary quantities given in the paper are given in Table 12.7. Expected cell counts (calculated under the assumption of no association between activity level and sleep quality) are also shown in Table 12.7.

TABLE 12.7 Observed and Expected Counts for Example 12.8

	Poor Sleep Quality	Good Sleep Quality
No activity	40 (21.96)	50 (68.04)
Light	116 (117.12)	364 (362.88)
Moderate	57 (61.00)	193 (189.00)
Vigorous	31 (43.92)	149 (136.08)

The Sleep Foundation was interested in using these sample data to determine whether there was an association between quality of sleep and activity level.

The X^2 test with a significance level of 0.01 will be used to test the relevant hypotheses:

Understand the context ➤ H_0: Quality of sleep and activity level are independent.
H_a: Quality of sleep and activity level are not independent.

Significance level: $\alpha = 0.01$.

Formulate a plan ➤ Test statistic: $X^2 = \sum_{\text{all cells}} \frac{(\text{observed cell count} - \text{expected cell count})^2}{\text{expected cell count}}$

Assumptions: All expected cell counts are at least 5. Assuming that the sample is representative of adult Americans, the X^2 test can be used.

Do the work ➤ Calculation: Minitab output is shown. From the Minitab output, $X^2 = 24.991$.

Chi-Square Test for Association: Activity Level, Sleep Quality

Rows: Activity Level Columns: Sleep Quality

	Poor Sleep Quality	Good Sleep Quality	All
None	40	50	90
	21.96	68.04	
Light	116	364	480
	117.12	362.88	
Moderate	57	193	250
	61.00	189.00	
Vigorous	31	149	180
	43.92	136.08	
All	244	756	1000

Cell Contents: Count
Expected count

Pearson Chi-Square = 24.991, DF = 3, P-Value = 0.000

P-value: From the Minitab output, *P*-value = 0.000.

Interpret the results ➤ Conclusion: Since the *P*-value is less than α, H_0 is rejected. There is convincing evidence that there is an association between quality of sleep and activity level.

In some investigations, values of more than two categorical variables are recorded for each individual in the sample. For example, in addition to the variables *quality of sleep* and *activity level*, the researchers in the study referenced in Example 12.8 might also have collected information on occupation. A number of interesting questions could then be explored: Are all three variables independent of one another? Is it possible that occupation and quality of sleep are dependent but that the relationship between them does not depend on activity level? For a particular activity level group, is there an association between quality of sleep and occupation?

The X^2 test procedure described in this section for analysis of bivariate categorical data can be extended for use with *multivariate categorical data*. Appropriate hypothesis tests can then be used to provide insight into the relationships between variables. However, the calculations required to determine expected cell counts and to calculate the value of X^2 are quite tedious, so they are seldom done without the aid of a computer. Most statistical software packages can perform this type of analysis.

EXERCISES 12.14 - 12.28

12.14 A particular state university system has six campuses. On each campus, a random sample of students will be selected, and each student will be categorized with respect to political philosophy as liberal, moderate, or conservative. The null hypothesis of interest is that the proportion of students falling in each of these three categories is the same at all six campuses.

a. On how many degrees of freedom will the resulting X^2 test be based?

b. How does the answer in Part (a) change if there are seven campuses rather than six?

c. How does the answer in Part (a) change if there are four rather than three categories for political philosophy?

12.15 A random sample of 1000 registered voters in a certain county is selected, and each voter is categorized with respect to both educational level (four categories) and preferred candidate in an upcoming election for county supervisor (five possibilities). The hypothesis of interest is that educational level and preferred candidate are independent.