12 Chi-Square Tests and Nonparametric Tests

Learning Objectives

In this chapter, you learn:

- How and when to use the chi-square test for contingency tables
- How to use the Marascuilo procedure for determining pairwise differences when evaluating more than two proportions
- How and when to use the McNemar test
- How to use the chi-square test for a variance or standard deviation
- How and when to use nonparametric tests

USING STATISTICS

@ T.C. Resort Properties

You are the manager of T.C. Resort Properties, a collection of five upscale hotels located on two tropical islands. Guests who are satisfied with the quality of services during their stay are more likely to return on a future vacation and to recommend the hotel to friends and relatives. You have defined the business objective as improving the return rate at the hotels. To assess the quality of services being provided by your hotels, guests are encouraged to complete a satisfaction survey when they check out. You need to analyze the data from these surveys to determine the overall satisfaction with the services provided, the likelihood that the guests will return to the hotel, and the reasons some guests indicate that they will not return. For example, on one island, T.C. Resort Properties operates the Beachcomber and Windsurfer hotels. Is the perceived quality at the Beachcomber Hotel the same as at the Windsurfer Hotel? If there is a difference, how can you use this information to improve the overall quality of service at T.C. Resort Properties? Furthermore, if guests indicate that they are not planning to return, what are the most common reasons given for this decision? Are the reasons given unique to a certain hotel or common to all hotels operated by T.C. Resort Properties?

In the preceding three chapters, you used hypothesis-testing procedures to analyze both numerical and categorical data. Chapter 9 presented some one-sample tests, and Chapter 10 developed several two-sample tests. Chapter 11 discussed the analysis of variance (ANOVA), which you use to study one or two factors of interest. This chapter extends hypothesis testing to analyze differences between population proportions based on two or more samples, to test the hypothesis of *independence* in the joint responses to two categorical variables, and to test for a population variance or standard deviation. The chapter concludes with nonparametric tests as alternatives to several hypothesis tests considered in Chapters 10 and 11.

12.1 Chi-Square Test for the Difference Between Two Proportions

In Section 10.3, you studied the Z test for the difference between two proportions. In this section, the data are examined from a different perspective. The hypothesis-testing procedure uses a test statistic that is approximated by a chi-square (χ^2) distribution. The results of this χ^2 test are equivalent to those of the Z test described in Section 10.3.

If you are interested in comparing the counts of categorical responses between two independent groups, you can develop a two-way **contingency table** (see Section 2.2) to display the frequency of occurrence of items of interest and items not of interest for each group. In Chapter 4, contingency tables were used to define and study probability.

To illustrate the contingency table, return to the Using Statistics scenario concerning T.C. Resort Properties. On one of the islands, T.C. Resort Properties has two hotels (the Beachcomber and the Windsurfer). You define the business objective as improving the quality of service at T.C. Resort Properties. You collect data from customer satisfaction surveys and focus on the responses to the single question "Are you likely to choose this hotel again?" You organize the results of the survey and determine that 163 of 227 guests at the Beachcomber responded yes to "Are you likely to choose this hotel again?" and 154 of 262 guests at the Windsurfer responded yes to "Are you likely to choose this hotel again?" You want to analyze the results to determine whether, at the 0.05 level of significance, there is evidence of a significant difference in guest satisfaction (as measured by likelihood to return to the hotel) between the two hotels.

The contingency table displayed in Table 12.1, which has two rows and two columns, is called a **2 × 2 contingency table**. The cells in the table indicate the frequency for each row and column combination.

TABLE 12.1

Layout of a 2 × 2 Contingency Table

	COLUMN VARIABLE (GROUP)		
ROW VARIABLE	**1**	**2**	**Totals**
Items of interest	X_1	X_2	X
Items not of interest	$n_1 - X_1$	$n_2 - X_2$	$n - X$
Totals	n_1	n_2	n

where

$$
\begin{aligned}
X_1 &= \text{number of items of interest in group 1} \\
X_2 &= \text{number of items of interest in group 2} \\
n_1 - X_1 &= \text{number of items that are not of interest in group 1} \\
n_2 - X_2 &= \text{number of items that are not of interest in group 2} \\
X &= X_1 + X_2\text{, the total number of items of interest} \\
n - X &= (n_1 - X_1) + (n_2 - X_2)\text{, the total number of items that are not of interest} \\
n_1 &= \text{sample size in group 1} \\
n_2 &= \text{sample size in group 2} \\
n &= n_1 + n_2 = \text{total sample size}
\end{aligned}
$$

Table 12.2 contains the contingency table for the hotel guest satisfaction study. The contingency table has two rows, indicating whether the guests would return to the hotel or would not return to the hotel, and two columns, one for each hotel. The cells in the table indicate the frequency of each row and column combination. The row totals indicate the number of guests who would return to the hotel and those who would not return to the hotel. The column totals are the sample sizes for each hotel location.

TABLE 12.2

2 × 2 Contingency Table for the Hotel Guest Satisfaction Survey

	HOTEL		
CHOOSE HOTEL AGAIN?	**Beachcomber**	**Windsurfer**	**Total**
Yes	163	154	317
No	64	108	172
Total	227	262	489

To test whether the population proportion of guests who would return to the Beachcomber, π_1, is equal to the population proportion of guests who would return to the Windsurfer, π_2, you can use the χ^2 **test for the difference between two proportions**. To test the null hypothesis that there is no difference between the two population proportions:

$$H_0: \pi_1 = \pi_2$$

against the alternative that the two population proportions are not the same:

$$H_1: \pi_1 \neq \pi_2$$

you use the χ^2_{STAT} test statistic, shown in Equation (12.1).

χ^2 TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

The χ^2_{STAT} test statistic is equal to the squared difference between the observed and expected frequencies, divided by the expected frequency in each cell of the table, summed over all cells of the table.

$$\chi^2_{STAT} = \sum_{all\text{ cells}} \frac{(f_o - f_e)^2}{f_e} \tag{12.1}$$

where

f_o = **observed frequency** in a particular cell of a contingency table

f_e = **expected frequency** in a particular cell if the null hypothesis is true

The χ^2_{STAT} test statistic approximately follows a chi-square distribution with 1 degree of freedom.[1]

[1]In general, the degrees of freedom in a contingency table are equal to (number of rows − 1) multiplied by (number of columns − 1).

To compute the expected frequency, f_e, in any cell, you need to understand that if the null hypothesis is true, the proportion of items of interest in the two populations will be equal. Then the sample proportions you compute from each of the two groups would differ from each other only by chance. Each would provide an estimate of the common population parameter, π. A statistic that combines these two separate estimates together into one overall estimate of the population parameter provides more information than either of the two separate estimates could provide by itself. This statistic, given by the symbol $\bar{p}$, represents the estimated overall proportion of items of interest for the two groups combined (i.e., the total number of items of interest divided by the total sample size). The complement of $\bar{p}$, $1 - \bar{p}$, represents the estimated overall proportion of items that are not of interest in the two groups. Using the notation presented in Table 12.1 on page 468, Equation (12.2) defines $\bar{p}$.

COMPUTING THE ESTIMATED OVERALL PROPORTION FOR TWO GROUPS

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n} \quad \textbf{(12.2)}$$

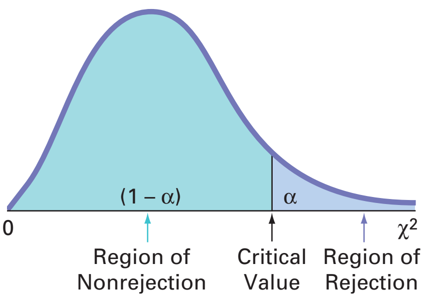
To compute the expected frequency, f_e, for cells that involve items of interest (i.e., the cells in the first row in the contingency table), you multiply the sample size (or column total) for a group by $\bar{p}$. To compute the expected frequency, f_e, for cells that involve items that are not of interest (i.e., the cells in the second row in the contingency table), you multiply the sample size (or column total) for a group by $1 - \bar{p}$.

The χ^2_{STAT} test statistic shown in Equation (12.1) on page 469 approximately follows a **chi-square (χ^2) distribution** (see Table E.4) with 1 degree of freedom. Using a level of significance α, you reject the null hypothesis if the computed χ^2_{STAT} test statistic is greater than χ^2_α, the upper-tail critical value from the χ^2 distribution with 1 degree of freedom. Thus, the decision rule is

$$\text{Reject } H_0 \text{ if } \chi^2_{STAT} > \chi^2_\alpha;$$
$$\text{otherwise, do not reject } H_0.$$

Figure 12.1 illustrates the decision rule.

FIGURE 12.1
Regions of rejection and nonrejection when using the chi-square test for the difference between two proportions, with level of significance α



If the null hypothesis is true, the computed χ^2_{STAT} test statistic should be close to zero because the squared difference between what is actually observed in each cell, f_o, and what is theoretically expected, f_e, should be very small. If H_0 is false, then there are differences in the population proportions, and the computed χ^2_{STAT} test statistic is expected to be large. However, what is a large difference in a cell is relative. The same actual difference between f_o and f_e from a cell with a small number of expected frequencies contributes more to the χ^2_{STAT} test statistic than a cell with a large number of expected frequencies.

To illustrate the use of the chi-square test for the difference between two proportions, return to the Using Statistics scenario concerning T.C. Resort Properties on page 467 and the corresponding contingency table displayed in Table 12.2 on page 469. The null hypothesis (H_0: $\pi_1 = \pi_2$) states that there is no difference between the proportion of guests who are likely to choose either of these hotels again. To begin,

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{163 + 154}{227 + 262} = \frac{317}{489} = 0.6483$$

$\bar{p}$ is the estimate of the common parameter π, the population proportion of guests who are likely to choose either of these hotels again if the null hypothesis is true. The estimated proportion of guests who are *not* likely to choose these hotels again is the complement of $\bar{p}$, $1 - 0.6483 = 0.3517$. Multiplying these two proportions by the sample size for the Beachcomber Hotel gives the number of guests expected to choose the Beachcomber again and the number not expected to choose this hotel again. In a similar manner, multiplying the two proportions by the Windsurfer Hotel's sample size yields the corresponding expected frequencies for that group.

EXAMPLE 12.1

Computing the Expected Frequencies

Compute the expected frequencies for each of the four cells of Table 12.2 on page 469.

SOLUTION

Yes—Beachcomber: $\bar{p} = 0.6483$ and $n_1 = 227$, so $f_e = 147.16$

Yes—Windsurfer: $\bar{p} = 0.6483$ and $n_2 = 262$, so $f_e = 169.84$

No—Beachcomber: $1 - \bar{p} = 0.3517$ and $n_1 = 227$, so $f_e = 79.84$

No—Windsurfer: $1 - \bar{p} = 0.3517$ and $n_2 = 262$, so $f_e = 92.16$

Table 12.3 presents these expected frequencies next to the corresponding observed frequencies.

TABLE 12.3

Comparing the Observed (f_o) and Expected (f_e) Frequencies

	HOTEL				
	BEACHCOMBER		WINDSURFER		
CHOOSE HOTEL AGAIN?	**Observed**	**Expected**	**Observed**	**Expected**	**Total**
Yes	163	147.16	154	169.84	317
No	64	79.84	108	92.16	172
Total	227	227.00	262	262.00	489

To test the null hypothesis that the population proportions are equal:

$$H_0: \pi_1 = \pi_2$$

against the alternative that the population proportions are not equal:

$$H_1: \pi_1 \neq \pi_2$$

you use the observed and expected frequencies from Table 12.3 to compute the χ^2_{STAT} test statistic given by Equation (12.1) on page 469. Table 12.4 presents the calculations.

TABLE 12.4

Computing the χ^2_{STAT} Test Statistic for the Hotel Guest Satisfaction Survey

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
163	147.16	15.84	250.91	1.71
154	169.84	−15.84	250.91	1.48
64	79.84	−15.84	250.91	3.14
108	92.16	15.84	250.91	2.72
				9.05

The chi-square (χ^2) distribution is a right-skewed distribution whose shape depends solely on the number of degrees of freedom. You find the critical value for the χ^2 test from Table E.4, a portion of which is presented as Table 12.5.

TABLE 12.5

Finding the Critical Value from the Chi-Square Distribution with 1 Degree of Freedom, Using the 0.05 Level of Significance

	Cumulative Probabilities						
	.005	**.01**	**...**	**.95**	**.975**	**.99**	**.995**
	Upper-Tail Area						
Degrees of Freedom	**.995**	**.99**	**...**	**.05**	**.025**	**.01**	**.005**
1			...	3.841	5.024	6.635	7.879
2	0.010	0.020	...	5.991	7.378	9.210	10.597
3	0.072	0.115	...	7.815	9.348	11.345	12.838
4	0.207	0.297	...	9.488	11.143	13.277	14.860
5	0.412	0.554	...	11.071	12.833	15.086	16.750

The values in Table 12.5 refer to selected upper-tail areas of the χ^2 distribution. A 2×2 contingency table has $(2 - 1)(2 - 1) = 1$ degree of freedom. Using $\alpha = 0.05$, with 1 degree of freedom, the critical value of χ^2 from Table 12.5 is 3.841. You reject H_0 if the computed χ^2_{STAT} test statistic is greater than 3.841 (see Figure 12.2). Because $\chi^2_{STAT} = 9.05 > 3.841$, you reject H_0. You conclude that the proportion of guests who would return to the Beachcomber is different from the proportion of guests who would return to the Windsurfer.

FIGURE 12.2

Regions of rejection and nonrejection when finding the χ^2 critical value with 1 degree of freedom, at the 0.05 level of significance

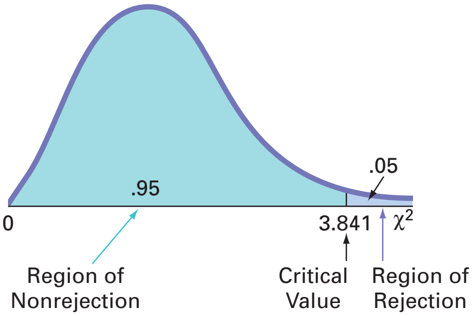


Figure 12.3 shows the results for the Table 12.2 guest satisfaction contingency table on page 469.

FIGURE 12.3

Excel and Minitab chi-square test results for the two-hotel guest satisfaction data

	A	B	C	D	E	F	G
1	Chi-Square Test						
2							
3	Observed Frequencies						
4		Hotel				Calculations	
5	Choose Again?	Beachcomber	Windsurfer	Total		fo-fe	
6	Yes	163	154	317		15.8446	-15.8446
7	No	64	108	172		-15.8446	15.8446
8	Total	227	262	489			
9							
10	Expected Frequencies						
11		Hotel					
12	Choose Again?	Beachcomber	Windsurfer	Total		(fo-fe)^2/fe	
13	Yes	147.1554	169.8446	317		1.7060	1.4781
14	No	79.8446	92.1554	172		3.1442	2.7242
15	Total	227	262	489			
16							
17	Data						
18	Level of Significance	0.05					
19	Number of Rows	2					
20	Number of Columns	2					
21	Degrees of Freedom	1	=(B19 - 1) * (B20 - 1)				
22							
23	Results						
24	Critical Value	3.8415	=CHIINV(B18, B21)				
25	Chi-Square Test Statistic	9.0526	=SUM(F13:G14)				
26	*p*-Value	0.0026	=CHIDIST(B25, B21)				
27	Reject the null hypothesis		=IF(B26 < B18, "Reject the null hypothesis",				
28			"Do not reject the null hypothesis")				
29	*Expected frequency assumption*						
30	*is met.*		=IF(OR(B13 < 5, C13 < 5, B14 < 5, C14 < 5), " is violated.", " is met.")				

```
Chi-Square Test: Beachcomber, Windsurfer
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

       Beachcomber  Windsurfer  Total
   1           163         154    317
            147.16      169.84
             1.706       1.478

   2            64         108    172
             79.84       92.16
             3.144       2.724

Total          227         262    489

Chi-Sq = 9.053, DF = 1, P-Value = 0.003
```

These results include the expected frequencies, χ^2_{STAT}, degrees of freedom, and p-value. The computed χ^2_{STAT} test statistic is 9.0526, which is greater than the critical value of 3.8415 (or the p-value $= 0.0026 < 0.05$), so you reject the null hypothesis that there is no difference in guest satisfaction between the two hotels. The p-value, equal to 0.0026, is the probability of observing sample proportions as different as or more different from the actual difference between the Beachcomber and Windsurfer $(0.718 - 0.588 = 0.13)$ observed in the sample data, if the population proportions for the Beachcomber and Windsurfer hotels are equal. Thus, there is strong evidence to conclude that the two hotels are significantly different with respect to guest satisfaction, as measured by whether a guest is likely to return to the hotel again. From Table 12.3 on page 471 you can see that a greater proportion of guests are likely to return to the Beachcomber than to the Windsurfer.

For the χ^2 test to give accurate results for a 2 $\times$ 2 table, you must assume that each expected frequency is at least 5. If this assumption is not satisfied, you can use alternative procedures, such as Fisher's exact test (see references 1, 2, and 4).

In the hotel guest satisfaction survey, both the Z test based on the standardized normal distribution (see Section 10.3) and the χ^2 test based on the chi-square distribution provide the same conclusion. You can explain this result by the interrelationship between the standardized normal distribution and a chi-square distribution with 1 degree of freedom. For such situations, the χ^2_{STAT} test statistic is the square of the Z_{STAT} test statistic. For instance, in the guest satisfaction study, the computed Z_{STAT} test statistic is $+3.0088$ and the computed χ^2_{STAT} test statistic is 9.0526. Except for rounding error, this 9.0526 value is the square of $+3.0088$ [i.e., $(+3.0088)^2 \cong 9.0526$]. Also, if you compare the critical values of the test statistics from the two distributions, at the 0.05 level of significance, the χ^2 value of 3.841 with 1 degree of freedom is the square of the Z value of ± 1.96. Furthermore, the p-values for both tests are equal. Therefore, when testing the null hypothesis of equality of proportions:

$$H_0\colon \pi_1 = \pi_2$$

against the alternative that the population proportions are not equal:

$$H_1\colon \pi_1 \neq \pi_2$$

the Z test and the χ^2 test are equivalent.

If you are interested in determining whether there is evidence of a *directional* difference, such as $\pi_1 > \pi_2$, you must use the Z test, with the entire rejection region located in one tail of the standardized normal distribution.

In Section 12.2, the χ^2 test is extended to make comparisons and evaluate differences between the proportions among more than two groups. However, you cannot use the Z test if there are more than two groups.

Problems for Section 12.1

LEARNING THE BASICS

12.1 Determine the critical value of χ^2 with 1 degree of freedom in each of the following circumstances:

a. $\alpha = 0.01$
b. $\alpha = 0.005$
c. $\alpha = 0.10$

12.2 Determine the critical value of χ^2 with 1 degree of freedom in each of the following circumstances:

a. $\alpha = 0.05$
b. $\alpha = 0.025$
c. $\alpha = 0.01$

12.3 Use the following contingency table:

	A	B	Total
1	20	30	50
2	30	45	75
Total	50	75	125

a. Compute the expected frequency for each cell.
b. Compare the observed and expected frequencies for each cell.
c. Compute χ^2_{STAT}. Is it significant at $\alpha = 0.05$?

12.4 Use the following contingency table:

	A	B	Total
1	20	30	50
2	30	20	50
Total	50	50	100

a. Compute the expected frequency for each cell.
b. Compute χ^2_{STAT}. Is it significant at $\alpha = 0.05$?

APPLYING THE CONCEPTS

12.5 A sample of 500 shoppers was selected in a large metropolitan area to determine various information concerning consumer behavior. Among the questions asked was, "Do you enjoy shopping for clothing?" The results are summarized in the following contingency table:

ENJOY SHOPPING FOR CLOTHING	GENDER		
	Male	Female	Total
Yes	136	224	360
No	104	36	140
Total	240	260	500

a. Is there evidence of a significant difference between the proportion of males and females who enjoy shopping for clothing at the 0.01 level of significance?
b. Determine the *p*-value in (a) and interpret its meaning.
c. What are your answers to (a) and (b) if 206 males enjoyed shopping for clothing and 34 did not?
d. Compare the results of (a) through (c) to those of Problem 10.29 (a), (b), and (d) on page 391.

12.6 Has the ease of removing your name from an e-mail list changed? A study of 100 large online retailers revealed the following:

YEAR	NEED THREE OR MORE CLICKS TO BE REMOVED	
	Yes	No
2009	39	61
2008	7	93

Source: Data extracted from "More Clicks to Escape an Email List," *The New York Times,* March 29, 2010, p. B2.

a. Set up the null and alternative hypotheses to try to determine whether the effort it takes to be removed from an e-mail list has changed.
b. Conduct the hypothesis test defined in (a), using the 0.05 level of significance.
c. Why shouldn't you compare the results in (a) to those of Problem 10.30 (b) on page 391?

12.7 A survey was conducted of 665 consumer magazines on the practices of their websites. Of these, 273 magazines reported that online-only content is copy-edited as rigorously as print content; 379 reported that online-only content is fact-checked as rigorously as print content. (Data extracted from S. Clifford, "Columbia Survey Finds a Slack Editing Process of Magazine Web Sites," *The New York Times*, March 1, 2010, p. B6.) Suppose that a sample of 500 newspapers revealed that 252 reported that online-only content is copy-edited as rigorously as print content and 296 reported that online-only content is fact-checked as rigorously as print content.
a. At the 0.05 level of significance, is there evidence of a difference between consumer magazines and newspapers in the proportion of online-only content that is copy-edited as rigorously as print content ?
b. Find the *p*-value in (a) and interpret its meaning.
c. At the 0.05 level of significance, is there evidence of a difference between consumer magazines and newspapers in the proportion of online-only content that is fact-checked as rigorously as print content?
d. Find the *p*-value in (c) and interpret its meaning.

SELF Test **12.8** Do people of different age groups differ in their response to e-mail messages? A survey by the Center for the Digital Future of the University of Southern California reported that 70.7% of users over age 70 believe that e-mail messages should be answered quickly, as compared to 53.6% of users 12 to 50 years old. (Data extracted from A. Mindlin, "Older E-mail Users Favor Fast Replies," *The New York Times*, July 14, 2008, p. B3.) Suppose that the survey was based on 1,000 users over age 70 and 1,000 users 12 to 50 years old.
a. At the 0.01 level of significance, is there evidence of a significant difference between the two age groups in their belief that e-mail messages should be answered quickly?
b. Find the *p*-value in (a) and interpret its meaning.
c. Compare the results of (a) and (b) to those of Problem 10.32 on page 391.

12.9 Different age groups use different media sources for news. A study on this issue explored the use of cell phones for accessing news. The study reported that 47% of users under age 50 and 15% of users age 50 and over accessed news on their cell phones. (Data extracted from "Cellphone Users Who Access News on Their Phones," *USA Today*, March 1, 2010, p. 1A.) Suppose that the survey consisted of 1,000 users under age 50, of whom 470 accessed news on their cell phones, and 891 users age 50 and over, of whom 134 accessed news on their cell phones.
a. Construct a 2×2 contingency table.
b. Is there evidence of a significant difference in the proportion that accessed the news on their cell phones between users under age 50 and users 50 years and older? (Use $\alpha = 0.05$.)
c. Determine the *p*-value in (b) and interpret its meaning.
d. Compare the results of (b) and (c) to those of Problem 10.35 (a) and (b) on page 392.

12.10 How do Americans feel about ads on websites? A survey of 1,000 adult Internet users found that 670 opposed ads on websites. (Data extracted from S. Clifford, "Tracked for Ads? Many Americans Say No Thanks," *The New York Times*, September 30, 2009, p. B3.) Suppose that a survey of 1,000 Internet users age 12–17 found that 510 opposed ads on websites.
a. At the 0.05 level of significance, is there evidence of a difference between adult Internet users and Internet users age 12–17 in the proportion who oppose ads?
b. Find the *p*-value in (a) and interpret its meaning.

12.2 Chi-Square Test for Differences Among More Than Two Proportions

In this section, the χ^2 test is extended to compare more than two independent populations. The letter c is used to represent the number of independent populations under consideration. Thus, the contingency table now has two rows and c columns. To test the null hypothesis that there are no differences among the c population proportions:

$$H_0: \pi_1 = \pi_2 = \cdots = \pi_c$$

against the alternative that not all the c population proportions are equal:

$$H_1: \text{Not all } \pi_j \text{ are equal (where } j = 1, 2, \ldots, c)$$

you use Equation (12.1) on page 469:

$$\chi^2_{STAT} = \sum_{all\ \text{cells}} \frac{(f_o - f_e)^2}{f_e}$$

where

f_o = observed frequency in a particular cell of a $2 \times c$ contingency table
f_e = expected frequency in a particular cell if the null hypothesis is true

If the null hypothesis is true and the proportions are equal across all c populations, the c sample proportions should differ only by chance. In such a situation, a statistic that combines these c separate estimates into one overall estimate of the population proportion, π, provides more information than any one of the c separate estimates alone. To expand on Equation (12.2) on page 470, the statistic $\bar{p}$ in Equation (12.3) represents the estimated overall proportion for all c groups combined.

COMPUTING THE ESTIMATED OVERALL PROPORTION FOR c GROUPS

$$\bar{p} = \frac{X_1 + X_2 + \cdots + X_c}{n_1 + n_2 + \cdots + n_c} = \frac{X}{n} \quad \textbf{(12.3)}$$

To compute the expected frequency, f_e, for each cell in the first row in the contingency table, multiply each sample size (or column total) by $\bar{p}$. To compute the expected frequency, f_e, for each cell in the second row in the contingency table, multiply each sample size (or column total) by $(1 - \bar{p})$. The test statistic shown in Equation (12.1) on page 469 approximately follows a chi-square distribution, with degrees of freedom equal to the number of rows in the contingency table minus 1, multiplied by the number of columns in the table minus 1. For a **$2 \times c$ contingency table**, there are $c - 1$ degrees of freedom:

$$\text{Degrees of freedom} = (2 - 1)(c - 1) = c - 1$$

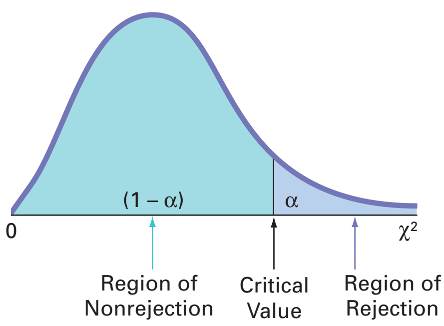
Using the level of significance α, you reject the null hypothesis if the computed χ^2_{STAT} test statistic is greater than χ^2_α, the upper-tail critical value from a chi-square distribution with $c - 1$ degrees of freedom. Therefore, the decision rule is

$$\text{Reject } H_0 \text{ if } \chi^2_{STAT} > \chi^2_\alpha;$$
$$\text{otherwise, do not reject } H_0.$$

Figure 12.4 illustrates this decision rule.

FIGURE 12.4

Regions of rejection and nonrejection when testing for differences among c proportions using the χ^2 test



To illustrate the χ^2 test for equality of proportions when there are more than two groups, return to the Using Statistics scenario on page 467 concerning T.C. Resort Properties. Once again, you define the business objective as improving the quality of service, but this time, three hotels located on a different island are to be surveyed. Data are collected from customer satisfaction surveys at these three hotels. You organize the responses into the contingency table shown in Table 12.6.

TABLE 12.6

2×3 Contingency Table for Guest Satisfaction Survey

	HOTEL			
CHOOSE HOTEL AGAIN?	**Golden Palm**	**Palm Royale**	**Palm Princess**	**Total**
Yes	128	199	186	513
No	88	33	66	187
Total	216	232	252	700

Because the null hypothesis states that there are no differences among the three hotels in the proportion of guests who would likely return again, you use Equation (12.3) to calculate an estimate of π, the population proportion of guests who would likely return again:

$$\bar{p} = \frac{X_1 + X_2 + \cdots + X_c}{n_1 + n_2 + \cdots + n_c} = \frac{X}{n}$$

$$= \frac{(128 + 199 + 186)}{(216 + 232 + 252)} = \frac{513}{700}$$

$$= 0.733$$

The estimated overall proportion of guests who would *not* be likely to return again is the complement, $(1 - \bar{p})$, or 0.267. Multiplying these two proportions by the sample size for each hotel yields the expected number of guests who would and would not likely return.

EXAMPLE 12.2

Computing the Expected Frequencies

Compute the expected frequencies for each of the six cells in Table 12.6.

SOLUTION

Yes—Golden Palm: $\bar{p} = 0.733$ and $n_1 = 216$, so $f_e = 158.30$

Yes—Palm Royale: $\bar{p} = 0.733$ and $n_2 = 232$, so $f_e = 170.02$

Yes—Palm Princess: $\bar{p} = 0.733$ and $n_3 = 252$, so $f_e = 184.68$

No—Golden Palm: $1 - \bar{p} = 0.267$ and $n_1 = 216$, so $f_e = 57.70$

No—Palm Royale: $1 - \bar{p} = 0.267$ and $n_2 = 232$, so $f_e = 61.98$

No—Palm Princess: $1 - \bar{p} = 0.267$ and $n_3 = 252$, so $f_e = 67.32$

Table 12.7 presents these expected frequencies.

TABLE 12.7

Contingency Table of Expected Frequencies from a Guest Satisfaction Survey of Three Hotels

	HOTEL			
CHOOSE HOTEL AGAIN?	**Golden Palm**	**Palm Royale**	**Palm Princess**	**Total**
Yes	158.30	170.02	184.68	513
No	57.70	61.98	67.32	187
Total	216.00	232.00	252.00	700

To test the null hypothesis that the proportions are equal:

$$H_0: \pi_1 = \pi_2 = \pi_3$$

against the alternative that not all three proportions are equal:

$$H_1: \text{Not all } \pi_j \text{ are equal (where } j = 1, 2, 3)$$

you use the observed frequencies from Table 12.6 and the expected frequencies from Table 12.7 to compute the χ^2_{STAT} test statistic [given by Equation (12.1) on page 469]. Table 12.8 presents the calculations.

TABLE 12.8

Computing the χ^2_{STAT} Test Statistic for the Guest Satisfaction Survey of Three Hotels

f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
128	158.30	−30.30	918.09	5.80
199	170.02	28.98	839.84	4.94
186	184.68	1.32	1.74	0.01
88	57.70	30.30	918.09	15.91
33	61.98	−28.98	839.84	13.55
66	67.32	−1.32	1.74	0.02
				40.23

You use Table E.4 to find the critical value of the χ^2 test statistic. In the guest satisfaction survey, because there are three hotels, there are $(2 - 1)(3 - 1) = 2$ degrees of freedom. Using $\alpha = 0.05$, the χ^2 critical value with 2 degrees of freedom is 5.991 (see Figure 12.5). Because the computed χ^2_{STAT} test statistic is 40.23, which is greater than this critical value, you reject the null hypothesis.

FIGURE 12.5

Regions of rejection and nonrejection when testing for differences in three proportions at the 0.05 level of significance, with 2 degrees of freedom

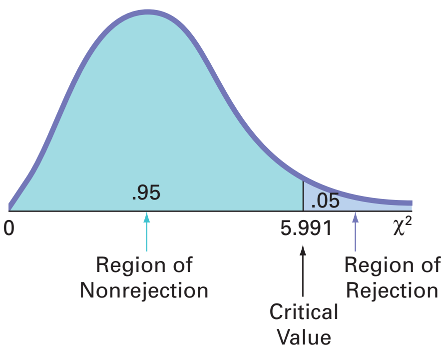


Figure 12.6 shows the results for this problem. The results also report the p-value. Because the p-value is (approximately) 0.0000, less than $\alpha = 0.05$, you reject the null hypothesis. Further, this p-value indicates that there is virtually no chance that there will be differences this large or larger among the three sample proportions, if the population proportions for the three hotels are equal. Thus, there is sufficient evidence to conclude that the hotel properties are different with respect to the proportion of guests who are likely to return.

FIGURE 12.6

Excel and Minitab chi-square test results for the Table 12.6 guest satisfaction data

	A	B	C	D	E	F	G	H	I
1	Chi-Square Test								
2									
3	Observed Frequencies								
4		Hotel					Calculations		
5	Choose Again?	Golden Palm	Palm Royale	Palm Princess	Total		fo - fe		
6	Yes	128	199	186	513		-30.2971	28.9771	1.3200
7	No	88	33	66	187		30.2971	-28.9771	-1.3200
8	Total	216	232	252	700				
9									
10	Expected Frequencies								
11		Hotel							
12	Choose Again?	Golden Palm	Palm Royale	Palm Princess	Total		(fo - fe)^2/fe		
13	Yes	158.2971	170.0229	184.68	513		5.7987	4.9386	0.0094
14	No	57.7029	61.9771	67.32	187		15.9077	13.5481	0.0259
15	Total	216	232	252	700				
16									
17	Data								
18	Level of Significance	0.05							
19	Number of Rows	2							
20	Number of Columns	3							
21	Degrees of Freedom	2	=(B19 - 1) * (B20 - 1)						
22									
23	Results								
24	Critical Value	5.9915	=CHIINV(B18, B21)						
25	Chi-Square Test Statistic	40.2284	=SUM(G13:I14)						
26	*p*-Value	0.0000	=CHIDIST(B25, B21)						
27	Reject the null hypothesis		=IF(B26 < B18, "Reject the null hypothesis",						
28			"Do not reject the null hypothesis")						
29	***Expected frequency assumption***								
30	***is met.***		=IF(OR(B13 < 1, C13 < 1, D13 < 1, B14 < 1, C14 < 1, D14 < 1),						
			" is violated.", " is met.")						

Chi-Square Test: Golden Palm, Palm Royale, Palm Princess

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        Golden    Palm      Palm
          Palm  Royale  Princess  Total
    1      128     199       186    513
        158.30  170.02    184.68
         5.799   4.939     0.009

    2       88      33        66    187
         57.70   61.98     67.32
        15.908  13.548     0.026

Total      216     232       252    700

Chi-Sq = 40.228, DF = 2, P-Value = 0.000
```

For the χ^2 test to give accurate results when dealing with $2 \times c$ contingency tables, all expected frequencies must be large. The definition of "large" has led to research among statisticians. Some statisticians (see reference 5) have found that the test gives accurate results as long as all expected frequencies are at least 0.5. Other statisticians, more conservative in their approach, believe that no more than 20% of the cells should contain expected frequencies less than 5, and no cells should have expected frequencies less than 1 (see reference 3). As a reasonable compromise between these points of view, to assure the validity of the test, you should make sure that each expected frequency is at least 1. To do this, you may need to collapse two or more low-expected-frequency categories into one category in the contingency table before performing the test. If combining categories is undesirable, you can use one of the available alternative procedures (see references 1, 2, and 7).

The Marascuilo Procedure

Rejecting the null hypothesis in a χ^2 test of equality of proportions in a $2 \times c$ table only allows you to reach the conclusion that not all c population proportions are equal. To determine which proportions differ, you use a multiple comparisons procedure such as the Marascuilo procedure.

The **Marascuilo procedure** enables you to make comparisons between all pairs of groups. First, you compute the sample proportions. Then, you use Equation (12.4) to compute the critical ranges for the Marascuilo procedure. You compute a different critical range for each pairwise comparison of sample proportions.

CRITICAL RANGE FOR THE MARASCUILO PROCEDURE

$$\text{Critical range} = \sqrt{\chi^2_\alpha}\sqrt{\frac{p_j(1-p_j)}{n_j} + \frac{p_{j'}(1-p_{j'})}{n_{j'}}} \tag{12.4}$$

Then, you compare each of the $c(c-1)/2$ pairs of sample proportions against its corresponding critical range. You declare a specific pair significantly different if the absolute difference in the sample proportions, $|p_j - p_{j'}|$, is greater than its critical range.

To apply the Marascuilo procedure, return to the guest satisfaction survey. Using the χ^2 test, you concluded that there was evidence of a significant difference among the population proportions. From Table 12.6 on page 476, the three sample proportions are

$$p_1 = \frac{X_1}{n_1} = \frac{128}{216} = 0.5926$$

$$p_2 = \frac{X_2}{n_2} = \frac{199}{232} = 0.8578$$

$$p_3 = \frac{X_3}{n_3} = \frac{186}{252} = 0.7381$$

Next, you compute the absolute differences in sample proportions and their corresponding critical ranges. Because there are three hotels, there are $(3)(3 - 1)/2 = 3$ pairwise comparisons. Using Table E.4 and an overall level of significance of 0.05, the upper-tail critical value for a chi-square distribution having $(c - 1) = 2$ degrees of freedom is 5.991. Thus,

$$\sqrt{\chi^2_\alpha} = \sqrt{5.991} = 2.4477$$

Absolute Difference in Proportions	**Critical Range**
$\lvert p_j - p_{j'} \rvert$	$2.4477\sqrt{\dfrac{p_j(1 - p_j)}{n_j} + \dfrac{p_{j'}(1 - p_{j'})}{n_{j'}}}$
$\lvert p_1 - p_2 \rvert = \lvert 0.5926 - 0.8578 \rvert = 0.2652$	$2.4477\sqrt{\dfrac{(0.5926)(0.4074)}{216} + \dfrac{(0.8578)(0.1422)}{232}} = 0.0992$
$\lvert p_1 - p_3 \rvert = \lvert 0.5926 - 0.7381 \rvert = 0.1455$	$2.4477\sqrt{\dfrac{(0.5926)(0.4074)}{216} + \dfrac{(0.7381)(0.2619)}{252}} = 0.1063$
$\lvert p_2 - p_3 \rvert = \lvert 0.8578 - 0.7381 \rvert = 0.1197$	$2.4477\sqrt{\dfrac{(0.8578)(0.1422)}{232} + \dfrac{(0.7381)(0.2619)}{252}} = 0.0880$

Figure 12.7 shows the Excel results for this example.

FIGURE 12.7

Excel Marascuilo procedure results worksheet for the guest satisfaction survey

Minitab does not contain a command to perform the Marascuilo procedure.

	A	B	C	D
1	Marascuilo Procedure for Guest Satisfaction Analysis			
2				
3	Level of Significance	0.05	=ChiSquare2x3!B18	
4	Square Root of Critical Value	2.4477	=SQRT(ChiSquare2x3!B24)	
5				
6	Group Sample Proportions			
7	1: Golden Palm	0.5926	=ChiSquare2x3!B6/ChiSquare2x3!B8	
8	2: Palm Royale	0.8578	=ChiSquare2x3!C6/ChiSquare2x3!C8	
9	3: Palm Princess	0.7381	=ChiSquare2x3!D6/ChiSquare2x3!D8	
10				
11	MARASCUILO TABLE			
12	Proportions	Absolute Differences	Critical Range	
13	| Group 1 - Group 2 |	0.2652	0.0992	Significant
14	| Group 1 - Group 3 |	0.1455	0.1063	Significant
15				
16	| Group 2 - Group 3 |	0.1197	0.0880	Significant

As the final step, you compare the absolute differences to the critical ranges. If the absolute difference is greater than its critical range, the proportions are significantly different. At the 0.05 level of significance, you can conclude that guest satisfaction is higher at the Palm Royale $(p_2 = 0.858)$ than at either the Golden Palm $(p_1 = 0.593)$ or the Palm Princess $(p_3 = 0.738)$ and that guest satisfaction is also higher at the Palm Princess than at the Golden Palm. These results clearly suggest that you should investigate possible reasons for these differences. In particular, you should try to determine why satisfaction is significantly lower at the Golden Palm than at the other two hotels.

Online Topic: The Analysis of Proportions (ANOP)

The ANOP procedure provides a confidence interval approach that allows you to determine which, if any, of the *c* groups has a proportion significantly different from the overall mean of all the group proportions combined. To study this topic, read the **ANOP** online topic file that is available on this book's companion website. (See Appendix C to learn how to access the online topic files.)

Problems for Section 12.2

LEARNING THE BASICS

12.11 Consider a contingency table with two rows and five columns.

a. How many degrees of freedom are there in the contingency table?
b. Determine the critical value for $\alpha = 0.05$.
c. Determine the critical value for $\alpha = 0.01$.

12.12 Use the following contingency table:

	A	B	C	Total
1	10	30	50	90
2	40	45	50	135
Total	50	75	100	225

a. Compute the expected frequencies for each cell.
b. Compute χ^2_{STAT}. Is it significant at $\alpha = 0.05$?

12.13 Use the following contingency table:

	A	B	C	Total
1	20	30	25	75
2	30	20	25	75
Total	50	50	50	150

a. Compute the expected frequencies for each cell.
b. Compute χ^2_{STAT}. Is it significant at $\alpha = 0.05$?
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which groups are different.

APPLYING THE CONCEPTS

12.14 How do Americans feel about online ads tailored to their individual interests? A survey of 1,000 adult Internet users found that 55% of the 18 to 24 year olds, 59% of 25 to 34 year olds, 66% of 35 to 49 year olds, 77% of 50 to 64 year olds, and 82% of 65 to 89 year olds opposed such ads. (Data extracted from S. Clifford, "Tracked for Ads? Many Americans Say No Thanks," *The New York Times*, September 30, 2009, p. B3.) Suppose that the survey was based on 200 respondents in each of five age groups: 18 to 24, 25 to 34, 35 to 49, 50 to 64, and 65 to 89.

a. At the 0.05 level of significance, is there evidence of a difference among the age groups in the opposition to ads on web pages tailored to their interests?
b. Determine the *p*-value in (a) and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which age groups are different.

12.15 How do Americans feel about online discounts tailored to their individual interests? A survey of 1,000 adult Internet users found that 37% of the 18 to 24 year olds, 44% of 25 to 34 year olds, 50% of 35 to 49 year olds, 58% of 50 to 64 year olds, and 70% of 65 to 89 year olds opposed such discounts. (Data extracted from S. Clifford, "Tracked for Ads? Many Americans Say No Thanks," *The New York Times*, September 30, 2009, p. B3.) Suppose that the survey was based on 200 respondents in each of five age groups: 18 to 24, 25 to 34, 35 to 49, 50 to 64, and 65 to 89.

a. At the 0.05 level of significance, is there evidence of a difference among the age groups in the opposition to discounts on web pages tailored to their interests?
b. Compute the *p*-value and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which age groups are different.

SELF Test **12.16** More shoppers do the majority of their grocery shopping on Saturday than any other day of the week. However, is there a difference in the various age groups in the proportion of people who do the majority of their grocery shopping on Saturday? A study showed the results for the different age groups. (Data extracted from "Major Shopping by Day," *Progressive Grocer Annual*

Report, April 30, 2002.) The data were reported as percentages, and no sample sizes were given:

	AGE		
MAJOR SHOPPING DAY	**Under 35**	**35–54**	**Over 54**
Saturday	24%	28%	12%
A day other than Saturday	76%	72%	88%

Assume that 200 shoppers for each age group were surveyed.

a. Is there evidence of a significant difference among the age groups with respect to major grocery shopping day? (Use $\alpha = 0.05$.)
b. Determine the p-value in (a) and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which age groups are different.
d. Discuss the managerial implications of (a) and (c). How can grocery stores use this information to improve marketing and sales? Be specific.

12.17 Repeat (a) and (b) of Problem 12.16, assuming that only 50 shoppers for each age group were surveyed. Discuss the implications of sample size on the χ^2 test for differences among more than two populations.

12.18 Is there a generation gap in music? A study reported that 45% of 16 to 29 year olds, 42% of 30 to 49 year olds, and 33% of 50 to 64 year olds often listened to rock music. (Data extracted from A. Tugend, "Bridging the Workplace Generation Gap: It Starts with a Text," *The New York Times*, November 7, 2009, p. B5.) Suppose that the study was based on a sample of 200 respondents in each group.

a. Is there evidence of a significant difference among the age groups with respect to the proportion who often listened to rock music? (Use $\alpha = 0.05$.)
b. Determine the p-value in (a) and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$. to determine which age groups are different.

12.19 Is there a generation gap in music? A study reported that 25% of 16 to 29 year olds, 21% of 30 to 49 year olds, and 31% of 50 to 64 year olds often listened to country music. (Data extracted from A. Tugend, "Bridging the Workplace Generation Gap: It Starts with a Text," *The New York Times*, November 7, 2009, p. B5.) Suppose that the study was based on a sample of 200 respondents in each group.

a. Is there evidence of a significant difference among the age groups with respect to the proportion who often listened to country music? (Use $\alpha = 0.05$.)
b. Determine the p-value in (a) and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$. to determine which age groups are different.

12.3 Chi-Square Test of Independence

In Sections 12.1 and 12.2, you used the χ^2 test to evaluate potential differences among population proportions. For a contingency table that has r rows and c columns, you can generalize the χ^2 test as a *test of independence* for two categorical variables.

For a test of independence, the null and alternative hypotheses follow:

H_0: The two categorical variables are independent (i.e., there is no relationship between them).

H_1: The two categorical variables are dependent (i.e., there is a relationship between them).

Once again, you use Equation (12.1) on page 469 to compute the test statistic:

$$\chi^2_{STAT} = \sum_{\textit{all}\text{ cells}} \frac{(f_o - f_e)^2}{f_e}$$

You reject the null hypothesis at the α level of significance if the computed value of the χ^2_{STAT} test statistic is greater than χ^2_α, the upper-tail critical value from a chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom (see Figure 12.8 on page 482). Thus, the decision rule is

$$\text{Reject } H_0 \text{ if } \chi^2_{STAT} > \chi^2_\alpha;$$
$$\text{otherwise, do not reject } H_0.$$

The χ^2 **test of independence** is similar to the χ^2 test for equality of proportions. The test statistics and the decision rules are the same, but the null and alternative hypotheses and conclusions are different. For example, in the guest satisfaction survey of Sections 12.1 and 12.2, there is evidence of a significant difference between the hotels with respect to the proportion

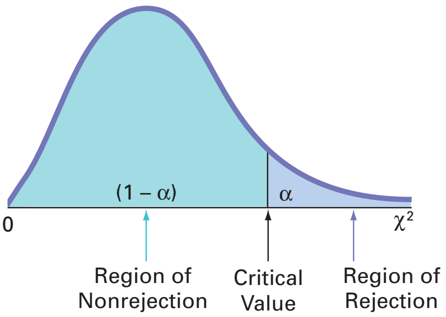


FIGURE 12.8

Regions of rejection and nonrejection when testing for independence in an $r \times c$ contingency table, using the χ^2 test

of guests who would return. From a different viewpoint, you could conclude that there is a significant relationship between the hotels and the likelihood that a guest would return. However, the two types of tests differ in how the samples are selected.

In a test for equality of proportions, there is one factor of interest, with two or more levels. These levels represent samples drawn from independent populations. The categorical responses in each group or level are classified into two categories, such as *item of interest* and *not an item of interest*. The objective is to make comparisons and evaluate differences between the proportions of the *items of interest* among the various levels. However, in a test for independence, there are two factors of interest, each of which has two or more levels. You select one sample and tally the joint responses to the two categorical variables into the cells of a contingency table.

To illustrate the χ^2 test for independence, suppose that, in the survey on hotel guest satisfaction, respondents who stated that they were not likely to return were asked what was the primary reason for their unwillingness to return to the hotel. Table 12.9 presents the resulting 4×3 contingency table.

In Table 12.9, observe that of the primary reasons for not planning to return to the hotel, 67 were due to price, 60 were due to location, 31 were due to room accommodation, and 29 were due to other reasons. As in Table 12.6 on page 476, there were 88 guests at the Golden Palm, 33 guests at the Palm Royale, and 66 guests at the Palm Princess who were not planning

TABLE 12.9

Contingency Table of Primary Reason for Not Returning and Hotel

PRIMARY REASON FOR NOT RETURNING	HOTEL			
	Golden Palm	Palm Royale	Palm Princess	Total
Price	23	7	37	67
Location	39	13	8	60
Room accommodation	13	5	13	31
Other	13	8	8	29
Total	88	33	66	187

to return. The observed frequencies in the cells of the 4×3 contingency table represent the joint tallies of the sampled guests with respect to primary reason for not returning and the hotel where they stayed. The null and alternative hypotheses are

H_0: There is no relationship between the primary reason for not returning and the hotel.

H_1: There is a relationship between the primary reason for not returning and the hotel.

To test this null hypothesis of independence against the alternative that there is a relationship between the two categorical variables, you use Equation (12.1) on page 469 to compute the test statistic:

$$\chi^2_{STAT} = \sum_{all\text{ cells}} \frac{(f_o - f_e)^2}{f_e}$$

where

f_o = observed frequency in a particular cell of the $r \times c$ contingency table

f_e = expected frequency in a particular cell if the null hypothesis of independence is true

To compute the expected frequency, f_e, in any cell, you use the multiplication rule for independent events discussed on page 160 [see Equation (4.7)]. For example, under the null hypothesis of independence, the probability of responses expected in the upper-left-corner cell representing primary reason of price for the Golden Palm is the product of the two separate probabilities $P(\text{Price})$ and $P(\text{Golden Palm})$. Here, the proportion of reasons that are due to price, $P(\text{Price})$, is $67/187 = 0.3583$, and the proportion of all responses from the Golden Palm, $P(\text{Golden Palm})$, is $88/187 = 0.4706$. If the null hypothesis is true, then the primary reason for not returning and the hotel are independent:

$$\begin{aligned} P(\text{Price } \textit{and} \text{ Golden Palm}) &= P(\text{Price}) \times P(\text{Golden Palm}) \\ &= (0.3583) \times (0.4706) \\ &= 0.1686 \end{aligned}$$

The expected frequency is the product of the overall sample size, n, and this probability, $187 \times 0.1686 = 31.53$. The f_e values for the remaining cells are calculated in a similar manner (see Table 12.10).

Equation (12.5) presents a simpler way to compute the expected frequency.

COMPUTING THE EXPECTED FREQUENCY

The expected frequency in a cell is the product of its row total and column total, divided by the overall sample size.

$$f_e = \frac{\text{Row total} \times \text{Column total}}{n} \tag{12.5}$$

where

$$\begin{aligned} \text{Row total} &= \text{sum of the frequencies in the row} \\ \text{Column total} &= \text{sum of the frequencies in the column} \\ n &= \text{overall sample size} \end{aligned}$$

For example, using Equation (12.5) for the upper-left-corner cell (price for the Golden Palm),

$$f_e = \frac{\text{Row total} \times \text{Column total}}{n} = \frac{(67)(88)}{187} = 31.53$$

and for the lower-right-corner cell (other reason for the Palm Princess),

$$f_e = \frac{\text{Row total} \times \text{Column total}}{n} = \frac{(29)(66)}{187} = 10.24$$

Table 12.10 lists the entire set of f_e values.

TABLE 12.10

Contingency Table of Expected Frequencies of Primary Reason for Not Returning with Hotel

PRIMARY REASON FOR NOT RETURNING	HOTEL: Golden Palm	Palm Royale	Palm Princess	Total
Price	31.53	11.82	23.65	67
Location	28.24	10.59	21.18	60
Room accommodation	14.59	5.47	10.94	31
Other	13.65	5.12	10.24	29
Total	88.00	33.00	66.00	187

To perform the test of independence, you use the χ^2_{STAT} test statistic shown in Equation (12.1) on page 469. The χ^2_{STAT} test statistic approximately follows a chi-square distribution, with degrees of freedom equal to the number of rows in the contingency table minus 1, multiplied by the number of columns in the table minus 1:

$$\text{Degrees of freedom} = (r - 1)(c - 1)$$
$$= (4 - 1)(3 - 1) = 6$$

Table 12.11 illustrates the computations for the χ^2_{STAT} test statistic.

TABLE 12.11
Computing the χ^2_{STAT} Test Statistic for the Test of Independence

Cell	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Price/Golden Palm	23	31.53	−8.53	72.76	2.31
Price/Palm Royale	7	11.82	−4.82	23.23	1.97
Price/Palm Princess	37	23.65	13.35	178.22	7.54
Location/Golden Palm	39	28.24	10.76	115.78	4.10
Location/Palm Royale	13	10.59	2.41	5.81	0.55
Location/Palm Princess	8	21.18	−13.18	173.71	8.20
Room/Golden Palm	13	14.59	−1.59	2.53	0.17
Room/Palm Royale	5	5.47	−0.47	0.22	0.04
Room/Palm Princess	13	10.94	2.06	4.24	0.39
Other/Golden Palm	13	13.65	−0.65	0.42	0.03
Other/Palm Royale	8	5.12	2.88	8.29	1.62
Other/Palm Princess	8	10.24	−2.24	5.02	0.49
					27.41

Using the $\alpha = 0.05$ level of significance, the upper-tail critical value from the chi-square distribution with 6 degrees of freedom is 12.592 (see Table E.4). Because $\chi^2_{STAT} = 27.41 > 12.592$, you reject the null hypothesis of independence (see Figure 12.9).

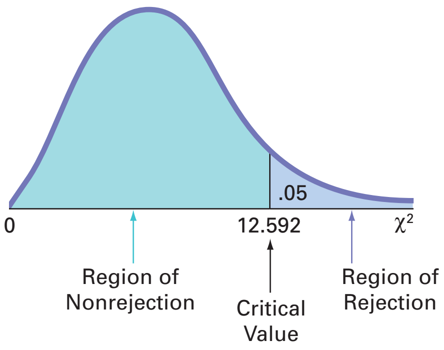


FIGURE 12.9
Regions of rejection and nonrejection when testing for independence in the hotel guest satisfaction survey example at the 0.05 level of significance, with 6 degrees of freedom

The results for this test, shown in Figure 12.10, include the p-value, 0.0001. Since $\chi^2_{STAT} = 27.4104 > 12.592$, you reject the null hypothesis of independence. Using the p-value approach, you reject the null hypothesis of independence because the p-value $= 0.0001 < 0.05$. The p-value indicates that there is virtually no chance of having a relationship this strong or stronger between the hotel and the primary reasons for not returning in a sample, if the primary reasons for not returning are independent of the specific hotels in the entire population. Thus, there is strong evidence of a relationship between primary reason for not returning and the hotel.

Examination of the observed and expected frequencies (see Table 12.11 above) reveals that price is underrepresented as a reason for not returning to the Golden Palm (i.e., $f_o = 23$ and $f_e = 31.53$) but is overrepresented at the Palm Princess. Guests are more satisfied with the

FIGURE 12.10

Excel and Minitab chi-square test results for the Table 12.9 primary reason for not returning and hotel data

	A	B	C	D	E	
1	Chi-Square Test of Independence					
2						
3		Observed Frequencies				
4			Hotel			
5	Reason for Not Returning	Golden Palm	Palm Royale	Palm Princess	Total	
6	Price	23	7	37	67	
7	Location	39	13	8	60	
8	Room accommodation	13	5	13	31	
9	Other	13	8	8	29	
10	Total	88	33	66	187	
11						
12		Expected Frequencies				
13			Hotel			
14	Reason for Not Returning	Golden Palm	Palm Royale	Palm Princess	Total	
15	Price	31.5294	11.8235	23.6471	67	
16	Location	28.2353	10.5882	21.1765	60	
17	Room accommodation	14.5882	5.4706	10.9412	31	
18	Other	13.6471	5.1176	10.2353	29	
19	Total	88	33	66	187	
20						
21	Data					
22	Level of Significance	0.05				
23	Number of Rows	4				
24	Number of Columns	3				
25	Degrees of Freedom	6	=(B23 - 1) * (B24 - 1)			
26						
27	Results					
28	Critical Value	12.5916	=CHIINV(B22, B25)			
29	Chi-Square Test Statistic	27.4104	=SUM(G15:I18)			
30	*p*-Value	0.0001	=CHIDIST(B29, B25)			
31	Reject the null hypothesis		=IF(B30 < B22, "Reject the null hypothesis", "Do not reject the null hypothesis")			
32						
33	*Expected frequency assumption*					
34	*is met.*		=IF(OR(B15 < 1, C15 < 1, D15 < 1, B16 < 1, C16 < 1 ,D16 < 1, B17 < 1, C17 < 1, D17 < 1, B18 < 1, C18 < 1, D18 < 1), " is violated.", " is met.")			

Chi-Square Test: Golden Palm, Palm Royale, Palm Princess

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        Golden    Palm      Palm
          Palm  Royale  Princess  Total
   1        23       7        37     67
         31.53   11.82     23.65
         2.307   1.968     7.540

   2        39      13         8     60
         28.24   10.59     21.18
         4.104   0.549     8.199

   3        13       5        13     31
         14.59    5.47     10.94
         0.173   0.040     0.387

   4        13       8         8     29
         13.65    5.12     10.24
         0.031   1.623     0.488

Total       88      33        66    187

Chi-Sq = 27.410, DF = 6, P-Value = 0.000
```

price at the Golden Palm than at the Palm Princess. Location is overrepresented as a reason for not returning to the Golden Palm but greatly underrepresented at the Palm Princess. Thus, guests are much more satisfied with the location of the Palm Princess than with that of the Golden Palm.

To ensure accurate results, all expected frequencies need to be large in order to use the χ^2 test when dealing with $r \times c$ contingency tables. As in the case of $2 \times c$ contingency tables in Section 12.2, all expected frequencies should be at least 1. For contingency tables in which one or more expected frequencies are less than 1, you can use the chi-square test after collapsing two or more low-frequency rows into one row (or collapsing two or more low-frequency columns into one column). Merging rows or columns usually results in expected frequencies sufficiently large to assure the accuracy of the χ^2 test.

Problems for Section 12.3

LEARNING THE BASICS

12.20 If a contingency table has three rows and four columns, how many degrees of freedom are there for the χ^2 test of independence?

12.21 When performing a χ^2 test of independence in a contingency table with r rows and c columns, determine the upper-tail critical value of the test statistic in each of the following circumstances:

a. $\alpha = 0.05, r = 4$ rows, $c = 5$ columns
b. $\alpha = 0.01, r = 4$ rows, $c = 5$ columns
c. $\alpha = 0.01, r = 4$ rows, $c = 6$ columns
d. $\alpha = 0.01, r = 3$ rows, $c = 6$ columns
e. $\alpha = 0.01, r = 6$ rows, $c = 3$ columns

APPLYING THE CONCEPTS

12.22 The owner of a restaurant serving Continental-style entrées has the business objective of learning more about the patterns of patron demand during the Friday-to-Sunday weekend time period. Data were collected from 630 customers on the type of entrée ordered and the type of dessert ordered and organized into the following table:

TYPE OF DESSERT	TYPE OF ENTRÉE				
	Beef	Poultry	Fish	Pasta	Total
Ice cream	13	8	12	14	47
Cake	98	12	29	6	145
Fruit	8	10	6	2	26
None	124	98	149	41	412
Total	243	128	196	63	630

At the 0.05 level of significance, is there evidence of a relationship between type of dessert and type of entrée?

12.23 Is there a generation gap in the type of music that people listen to? The following table represents the type of favorite music for a sample of 1,000 respondents classified according to their age group:

FAVORITE TYPE	AGE				
	16–29	30–49	50–64	65 and over	Total
Rock	71	62	51	27	211
Rap or hip-hop	40	21	7	3	71
Rhythm and blues	48	46	46	40	180
Country	43	53	59	79	234
Classical	22	28	33	46	129
Jazz	18	26	36	43	123
Salsa	8	14	18	12	52
Total	250	250	250	250	1000

At the 0.05 level of significance, is there evidence of a relationship between favorite type of music and age group?

SELF Test **12.24** A large corporation is interested in determining whether a relationship exists between the commuting time of its employees and the level of stress-related problems observed on the job. A study of 116 workers reveals the following:

COMMUTING TIME	STRESS LEVEL			
	High	Moderate	Low	Total
Under 15 min.	9	5	18	32
15–45 min.	17	8	28	53
Over 45 min.	18	6	7	31
Total	44	19	53	116

a. At the 0.01 level of significance, is there evidence of a significant relationship between commuting time and stress level?

b. What is your answer to (a) if you use the 0.05 level of significance?

12.25 Where people turn for news is different for various age groups. A study indicated where different age groups primarily get their news:

MEDIA	AGE GROUP		
	Under 36	36–50	50 +
Local TV	107	119	133
National TV	73	102	127
Radio	75	97	109
Local newspaper	52	79	107
Internet	95	83	76

At the 0.05 level of significance, is there evidence of a significant relationship between the age group and where people primarily get their news? If so, explain the relationship.

12.26 *USA Today* reported on when the decision of what to have for dinner is made. Suppose the results were based on a survey of 1,000 respondents and considered whether the household included any children under 18 years old. The results are cross-classified in the following table:

WHEN DECISION MADE	TYPE OF HOUSEHOLD		
	One Adult/No Children	Two or More Adults/ Children	Two or More Adults/No Children
Just before eating	162	54	154
In the afternoon	73	38	69
In the morning	59	58	53
A few days before	21	64	45
The night before	15	50	45
Always eat the same thing on this night	2	16	2
Not sure	7	6	7

Source: Data extracted from "What's for Dinner," **www.usatoday.com**, January 10, 2000.

At the 0.05 level of significance, is there evidence of a significant relationship between when the decision is made of what to have for dinner and the type of household?

12.4 McNemar Test for the Difference Between Two Proportions (Related Samples)

In Section 10.3, you used the Z test, and in Section 12.1, you used the chi-square test to examine whether there was a difference in the proportion of items of interest between two populations. These tests require independent samples from each population. However, sometimes when you are testing differences between the proportion of items of interest, the data are collected from repeated measurements or matched samples. For example, in marketing, these situations can occur when you want to determine whether there has been a change in attitude, perception, or behavior from one time period to another. To test whether there is evidence of a difference between the proportions when the data have been collected from two related samples, you can use the **McNemar test**.

Table 12.12 presents the 2×2 table needed for the McNemar test.

TABLE 12.12

2×2 Contingency Table for the McNemar Test

	CONDITION (GROUP) 2		
CONDITION (GROUP) 1	**Yes**	**No**	**Totals**
Yes	A	B	$A + B$
No	C	D	$C + D$
Totals	$A + C$	$B + D$	n

where

A = number of respondents who answer yes to condition 1 and yes to condition 2

B = number of respondents who answer yes to condition 1 and no to condition 2

C = number of respondents who answer no to condition 1 and yes to condition 2

D = number of respondents who answer no to condition 1 and no to condition 2

n = number of respondents in the sample

The sample proportions are

$$p_1 = \frac{A + B}{n} = \text{proportion of respondents in the sample who answer yes to condition 1}$$

$$p_2 = \frac{A + C}{n} = \text{proportion of respondents in the sample who answer yes to condition 2}$$

The population proportions are

π_1 = proportion in the population who would answer yes to condition 1

π_2 = proportion in the population who would answer yes to condition 2

When testing differences between the proportions, you can use a two-tail test or a one-tail test. In both cases, you use a test statistic that approximately follows the normal distribution. Equation (12.6) presents the McNemar test statistic used to test H_0: $\pi_1 = \pi_2$.

McNEMAR TEST STATISTIC

$$Z_{STAT} = \frac{B - C}{\sqrt{B + C}} \qquad \textbf{(12.6)}$$

where the Z_{STAT} test statistic is approximately normally distributed.

To illustrate the McNemar test, suppose that the business problem facing a cell phone provider was to determine the effect of a marketing campaign on the brand loyalty of cell phone customers.

Data were collected from $n = 600$ participants. In the study, the participants were initially asked to state their preferences for two competing cell phone providers, Sprint and Verizon. Initially, 282 panelists said they preferred Sprint and 318 said they preferred Verizon. After exposing the set of participants to an intensive marketing campaign strategy for Verizon, the same 600 participants are again asked to state their preferences. Of the 282 panelists who previously preferred Sprint, 246 maintained their brand loyalty, but 36 switched their preference to Verizon. Of the 318 participants who initially preferred Verizon, 306 remained brand loyal, but 12 switched their preference to Sprint. These results are organized into the contingency table presented in Table 12.13.

You use the McNemar test for these data because you have repeated measurements from the same set of panelists. Each participant gave a response about whether he or she preferred Sprint or Verizon before exposure to the intensive marketing campaign and then again after exposure to the campaign.

TABLE 12.13

Brand Loyalty of Cell Phone Providers

	AFTER MARKETING CAMPAIGN		
BEFORE MARKETING CAMPAIGN	**Sprint**	**Verizon**	**Total**
Sprint	246	36	282
Verizon	12	306	318
Total	258	342	600

To determine whether the intensive marketing campaign was effective, you want to investigate whether there is a difference between the population proportion who favor Sprint before the campaign, π_1, versus the proportion who favor Sprint after the campaign, π_2. The null and alternative hypotheses are

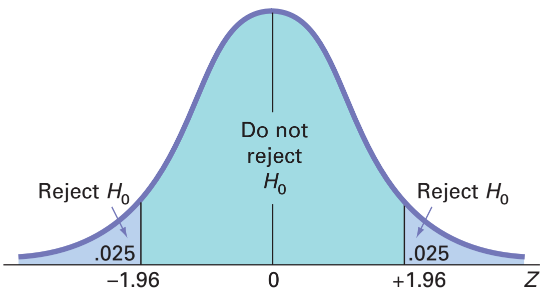
$$H_0: \pi_1 = \pi_2$$
$$H_1: \pi_1 \neq \pi_2$$

Using a 0.05 level of significance, the critical values are -1.96 and $+1.96$ (see Figure 12.11), and the decision rule is

$$\text{Reject } H_0 \text{ if } Z_{STAT} < -1.96 \text{ or if } Z_{STAT} > +1.96;$$
$$\text{otherwise, do not reject } H_0.$$

FIGURE 12.11

Two-tail McNemar test at the 0.05 level of significance



For the data in Table 12.13,

$$A = 246 \quad B = 36 \quad C = 12 \quad D = 306$$

so that

$$p_1 = \frac{A + B}{n} = \frac{246 + 36}{600} = \frac{282}{600} = 0.47 \text{ and } p_2 = \frac{A + C}{n} = \frac{246 + 12}{600} = \frac{258}{600} = 0.43$$

Using Equation (12.6),

$$Z = \frac{B - C}{\sqrt{B + C}} = \frac{36 - 12}{\sqrt{36 + 12}} = \frac{24}{\sqrt{48}} = 3.4641$$

Because $Z_{STAT} = 3.4641 > 1.96$, you reject H_0. Using the p-value approach (see Figure 12.12), the p-value is 0.0005. Because $0.0005 < 0.05$, you reject H_0. You can conclude that the population proportion who prefer Sprint before the intensive marketing campaign is different from the population proportion who prefer Sprint after exposure to the intensive Verizon marketing campaign. In fact, from Figure 12.12, observe that preference for Verizon increased after exposure to the intensive marketing campaign.

FIGURE 12.12

Excel results for the McNemar test for brand loyalty of cell phone providers

Minitab does not contain a command to perform the McNemar test.

	A	B	C	D	
1	**McNemar Test**				
2					
3	**Observed Frequencies**				
4		After Campaign			
5	Before Campaign	Sprint	Verizon	Total	
6	Sprint	246	36	282	
7	Verizon	12	306	318	
8	Total	258	342	600	
9					
10	**Data**				
11	Level of Significance	0.05			
12					
13	Intermediate Calculations				
14	Numerator	24	=C6 - B7		
15	Denominator	6.9282	=SQRT(C6 + B7)		
16	*Z* Test Statistic	3.4641	=B14/B15		
17					
18	**Two-Tail Test**				
19	Lower Critical Value	-1.9600	=NORMSINV(B11/2)		
20	Upper Critical Value	1.9600	=NORMSINV(1 - B11/2)		
21	*p*-Value	0.0005	=2 * (1 - NORMSDIST(ABS(B16)))		
22	Reject the null hypothesis		=IF(B21 < B11, "Reject the null hypothesis", "Do not reject the null hypothesis")		

Problems for Section 12.4

LEARNING THE BASICS

12.27 Given the following table for two related samples:

GROUP 1	GROUP 2: Yes	No	Total
Yes	46	25	71
No	16	59	75
Total	62	84	146

a. Compute the McNemar test statistic.

b. At the 0.05 level of significance, is there evidence of a difference between group 1 and group 2?

APPLYING THE CONCEPTS

SELF Test **12.28** A market researcher wanted to determine whether the proportion of coffee drinkers who preferred Brand *A* increased as a result of an advertising campaign. A random sample of 200 coffee drinkers was selected. The results indicating preference for Brand *A* or Brand *B* prior to the beginning of the advertising campaign and after its completion are shown in the following table:

PREFERENCE PRIOR TO ADVERTISING CAMPAIGN	PREFERENCE AFTER COMPLETION OF ADVERTISING CAMPAIGN: Brand *A*	Brand *B*	Total
Brand *A*	101	9	110
Brand *B*	22	68	90
Total	123	77	200

a. At the 0.05 level of significance, is there evidence that the proportion of coffee drinkers who prefer Brand *A* is lower at the beginning of the advertising campaign than at the end of the advertising campaign?

b. Compute the p-value in (a) and interpret its meaning.

12.29 Two candidates for governor participated in a televised debate. A political pollster recorded the preferences of 500 registered voters in a random sample prior to and after the debate:

PREFERENCE PRIOR TO DEBATE	PREFERENCE AFTER DEBATE		
	Candidate *A*	**Candidate *B***	**Total**
Candidate *A*	269	21	290
Candidate *B*	36	174	210
Total	305	195	500

a. At the 0.01 level of significance, is there evidence of a difference in the proportion of voters who favored Candidate *A* prior to and after the debate?

b. Compute the p-value in (a) and interpret its meaning.

12.30 A taste-testing experiment compared two brands of Chilean merlot wines. After the initial comparison, 60 preferred Brand *A*, and 40 preferred Brand *B*. The 100 respondents were then exposed to a very professional and powerful advertisement promoting Brand *A*. The 100 respondents were then asked to taste the two wines again and declare which brand they preferred. The results are shown in the following table:

PREFERENCE PRIOR TO ADVERTISING	PREFERENCE AFTER ADVERTISING		
	Brand *A*	**Brand *B***	**Total**
Brand *A*	55	5	60
Brand *B*	15	25	40
Total	70	30	100

a. At the 0.05 level of significance, is there evidence that the proportion who preferred Brand *A* was lower before the advertising than after the advertising?

b. Compute the p-value in (a) and interpret its meaning.

12.31 The CEO of a large metropolitan health-care facility would like to assess the effect of the recent implementation of the Six Sigma management approach on customer satisfaction. A random sample of 100 patients is selected from a list of patients who were at the facility the past week and also a year ago:

SATISFIED LAST YEAR	SATISFIED NOW		
	Yes	**No**	**Total**
Yes	67	5	72
No	20	8	28
Total	87	13	100

a. At the 0.05 level of significance, is there evidence that satisfaction was lower last year, prior to introduction of Six Sigma management?

b. Compute the p-value in (a) and interpret its meaning.

12.32 The personnel director of a large department store wants to reduce absenteeism among sales associates. She decides to institute an incentive plan that provides financial rewards for sales associates who are absent fewer than five days in a given calendar year. A sample of 100 sales associates selected at the end of the second year reveals the following:

YEAR 1	YEAR 2		
	$<$ 5 Days Absent	**$\geq$ 5 Days Absent**	**Total**
$<$ 5 Days Absent	32	4	36
$\geq$ 5 Days Absent	25	39	64
Total	57	43	100

a. At the 0.05 level of significance, is there evidence that the proportion of employees absent fewer than five days was lower in year 1 than in year 2?

b. Compute the p-value in (a) and interpret its meaning.

12.5 Chi-Square Test for the Variance or Standard Deviation

When analyzing numerical data, sometimes you need to test a hypothesis about the population variance or standard deviation. For example, in the cereal-filling process described in Section 9.1, you assumed that the population standard deviation, σ, was equal to 15 grams. To determine whether the variability of the process has changed, you need to test whether the standard deviation has changed from the previously specified level of 15 grams.

Assuming that the data are normally distributed, you use the χ^2 **test for the variance or standard deviation** defined in Equation (12.7) to test whether the population variance or standard deviation is equal to a specified value.

χ^2 TEST FOR THE VARIANCE OR STANDARD DEVIATION

$$\chi^2_{STAT} = \frac{(n - 1)S^2}{\sigma^2} \tag{12.7}$$

where

$$n = \text{sample size}$$
$$S^2 = \text{sample variance}$$
$$\sigma^2 = \text{hypothesized population variance}$$

The test statistic χ^2_{STAT} follows a chi-square distribution with $n - 1$ degrees of freedom.

To apply the test of hypothesis, return to the cereal-filling example. You are interested in determining whether the standard deviation has changed from the previously specified level of 15 grams. Thus, you use a two-tail test with the following null and alternative hypotheses:

$$H_0: \sigma^2 = 225 \text{ (that is, } \sigma = 15 \text{ grams)}$$
$$H_1: \sigma^2 \neq 225 \text{ (that is, } \sigma \neq 15 \text{ grams)}$$

If you select a sample of 25 cereal boxes, you reject the null hypothesis if the computed χ^2_{STAT} test statistic falls into either the lower or upper tail of a chi-square distribution with $25 - 1 = 24$ degrees of freedom, as shown in Figure 12.13. From Equation (12.7), observe that the χ^2_{STAT} test statistic falls into the lower tail of the chi-square distribution if the sample standard deviation (S) is sufficiently smaller than the hypothesized σ of 15 grams, and it falls into the upper tail if S is sufficiently larger than 15 grams. From Table 12.14 (or Table E.4), if you select a level of significance of 0.05, the lower and upper critical values are 12.401 and 39.364, respectively. Therefore, the decision rule is

$$\text{Reject } H_0 \text{ if } \chi^2_{STAT} < 12.401 \text{ or if } \chi^2_{STAT} > 39.364;$$
$$\text{otherwise, do not reject } H_0.$$

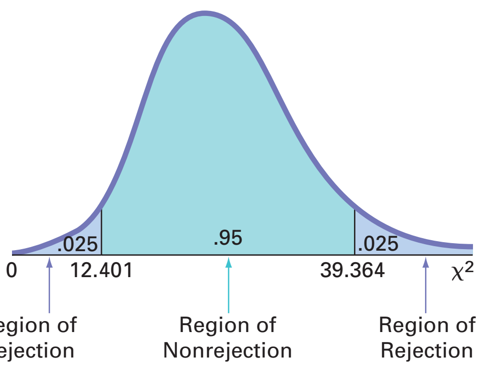


FIGURE 12.13
Determining the lower and upper critical values of a chi-square distribution with 24 degrees of freedom corresponding to a 0.05 level of significance for a two-tail test of hypothesis about a population variance or standard deviation

TABLE 12.14

Finding the Critical Values Corresponding to a 0.05 Level of Significance for a Two-Tail Test from the Chi-Square Distribution with 24 Degrees of Freedom

	Cumulative Area							
	.005	.01	$\boxed{.025}$	.05	.10	.90	.95	$\boxed{.975}$
	Upper-Tail Areas							
Degrees of Freedom	**.995**	**.99**	$\boxed{.975}$	**.95**	**.90**	**.10**	**.05**	$\boxed{.025}$
1	...	...	0.001	0.004	0.016	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076
$\boxed{24}$	9.886	10.856	$\boxed{12.401}$	13.848	15.659	33.196	36.415	$\boxed{39.364}$
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646

Source: Extracted from Table E.4.

Suppose that in the sample of 25 cereal boxes, the standard deviation, S, is 17.7 grams. Using Equation (12.7),

$$\chi^2_{STAT} = \frac{(n-1)S^2}{\sigma^2} = \frac{(25-1)(17.7)^2}{(15)^2} = 33.42$$

Because $12.401 < \chi^2_{STAT} = 33.42 < 39.364$, or because the p-value $= 0.0956 > 0.05$ (see Figure 12.14), you do not reject H_0. You conclude that there is insufficient evidence that the population standard deviation is different from 15 grams.

FIGURE 12.14

Excel results for the chi-square test for the standard deviation of the cereal-filling process

Minitab does not contain a command that directly performs this test.

	A	B	
1	**Cereal-Filling Analysis**		
2			
3	**Data**		
4	Null Hypothesis σ^2=	225	
5	Level of Significance	0.05	
6	Sample Size	25	
7	Sample Standard Deviation	17.7	
8			
9	Intermediate Calculations		
10	Degrees of Freedom	24	=B6 - 1
11	Half Area	0.025	=B5/2
12	Chi-Square Statistic	33.4176	=B10 * B7^2/B4
13			
14	**Two-Tail Test**		
15	Lower Critical Value	12.4012	=CHIINV(1 - B11, B10)
16	Upper Critical Value	39.3641	=CHIINV(B11, B10)
17	*p*-Value	0.0956	=IF(B12 - B15 < 0, 1 - CHIDIST(B12, B10), CHIDIST(B12, B10))
18	Do not reject the null hypothesis		=IF(B17 < B5/2, "Reject the null hypothesis", "Do not reject the null hypothesis")

In testing a hypothesis about a population variance or standard deviation, you assume that the values in the population are normally distributed. However, the chi-square test statistic for the variance or standard deviation is very sensitive to departures from this assumption (i.e., it is not a robust test). Thus, if the population is not normally distributed, particularly for small sample sizes, the accuracy of the test can be seriously affected.

Problems for Section 12.5

LEARNING THE BASICS

12.33 Determine the lower- and upper-tail critical values of χ^2 for each of the following two-tail tests:
a. $\alpha = 0.01, n = 26$
b. $\alpha = 0.05, n = 17$
c. $\alpha = 0.10, n = 14$

12.34 Determine the lower- and upper-tail critical values of χ^2 for each of the following two-tail tests:
a. $\alpha = 0.01, n = 24$
b. $\alpha = 0.05, n = 20$
c. $\alpha = 0.10, n = 16$

12.35 You are working with a sample of $n = 25$ selected from an underlying normal population, $S = 150$. What is the value of χ^2_{STAT} if you are testing the null hypothesis H_0: $\sigma = 100$?

12.36 You are working with a sample of $n = 16$ selected from an underlying normal population, $S = 10$. What is the value of χ^2_{STAT} if you are testing the null hypothesis H_0: $\sigma = 12$?

12.37 In Problem 12.36, how many degrees of freedom are there in the hypothesis test?

12.38 In Problems 12.36 and 12.37, what are the critical values from Table E.4 if the level of significance is $\alpha = 0.05$ and H_1 is as follows:
a. $\sigma \neq 12$?
b. $\sigma < 12$?

12.39 In Problems 12.36, 12.37, and 12.38, what is your statistical decision if H_1 is
a. $\sigma \neq 12$?
b. $\sigma < 12$?

12.40 If, in a sample of size $n = 16$ selected from a very left-skewed population, the sample standard deviation is $S = 24$, would you use the hypothesis test given in Equation (12.7) to test H_0: $\sigma = 20$? Discuss.

APPLYING THE CONCEPTS

12.41 A manufacturer of candy must monitor the temperature at which the candies are baked. Too much variation will cause inconsistency in the taste of the candy. Past records show that the standard deviation of the temperature has been 1.2°F. A random sample of 30 batches of candy is selected, and the sample standard deviation of the temperature is 2.1°F.
a. At the 0.05 level of significance, is there evidence that the population standard deviation has increased above 1.2°F?
b. What assumption do you need to make in order to perform this test?
c. Compute the p-value in (a) and interpret its meaning.

SELF Test **12.42** A market researcher for an automobile manufacturer intends to conduct a nationwide survey concerning car repairs. Among the questions included in the survey is the following: "What was the cost of all repairs performed on your car last year?" In order to determine the sample size necessary, the researcher needs to provide an estimate of the standard deviation. Using his past experience and judgment, he estimates that the standard deviation of the amount of repairs is \$200. Suppose that a small-scale study of 25 auto owners selected at random indicates a sample standard deviation of \$237.52.
a. At the 0.05 level of significance, is there evidence that the population standard deviation is different from \$200?
b. What assumption do you need to make in order to perform this test?
c. Compute the p-value in part (a) and interpret its meaning.

12.43 The marketing manager of a branch office of a local telephone operating company wants to study characteristics of residential customers served by her office. In particular, she wants to estimate the mean monthly cost of calls within the local calling region. In order to determine the sample size necessary, she needs an estimate of the standard deviation. On the basis of her past experience and judgment, she estimates that the standard deviation is equal to \$12. Suppose that a small-scale study of 15 residential customers indicates a sample standard deviation of \$9.25.
a. At the 0.10 level of significance, is there evidence that the population standard deviation is different from \$12?
b. What assumption do you need to make in order to perform this test?
c. Compute the p-value in (a) and interpret its meaning.

12.44 A manufacturer of doorknobs has a production process that is designed to provide a doorknob with a target diameter of 2.5 inches. In the past, the standard deviation of the diameter has been 0.035 inch. In an effort to reduce the variation in the process, various studies have resulted in a redesigned process. A sample of 25 doorknobs produced under the new process indicates a sample standard deviation of 0.025 inch.
a. At the 0.05 level of significance, is there evidence that the population standard deviation is less than 0.035 inch in the new process?
b. What assumption do you need to make in order to perform this test?
c. Compute the p-value in (a) and interpret its meaning.

12.45 A manufacturing company produces steel housings for electrical equipment. The main component part of the housing is a steel trough that is made out of a 14-gauge steel coil. It is produced using a 250-ton progressive punch press with a wipe-down operation that puts two 90-degree forms in the flat steel to make the trough. The distance from one side of the form to the other is critical because of weatherproofing in outdoor applications. The company requires that the width of the trough must be between 8.31 inches and 8.61 inches. In the past, the standard deviation of the width of the trough has been 0.05 inch. The file **Trough2** contains the widths of the troughs, in inches, for a sample of $n = 25$:

8.312 8.343 8.317 8.383 8.348 8.410 8.351 8.373 8.481
8.422 8.476 8.382 8.484 8.403 8.414 8.419 8.385 8.465
8.498 8.447 8.436 8.413 8.489 8.414 8.481

a. At the 0.05 level of significance, is there evidence that the standard deviation of the width of the troughs is different from 0.05 inch?

b. What assumption about the population distribution is needed in order to conduct the test in (a)?

12.46 An important quality characteristic of interest in a teabag filling process is the weight of the tea in the individual bags. The weight of the teabags should be as consistent as possible. In the past, the standard deviation of the weight of the teabags has been 0.05 grams. The file **Teabags2** contains an ordered array of the weight, in grams, of a sample of 20 tea bags produced during an eight-hour shift. Is there evidence that the standard deviation of the amount of tea per bag is greater than 0.05 grams? (Use $\alpha = 0.01$.)

12.6 Wilcoxon Rank Sum Test: Nonparametric Analysis for Two Independent Populations

"A nonparametric procedure is a statistical procedure that has (certain) desirable properties that hold under relatively mild assumptions regarding the underlying population(s) from which the data are obtained."

—Myles Hollander and Douglas A. Wolfe (reference 4, p. 1)

In Section 10.1, you used the t test for the difference between the means of two independent populations. If sample sizes are small and you cannot assume that the data in each sample are from normally distributed populations, you have two choices:

- Use a nonparametric procedure such as the Wilcoxon rank sum test, which does not depend on the assumption of normality for the two populations.
- Use the pooled-variance t test, following a *normalizing transformation* on the data (see reference 10).

This section introduces the **Wilcoxon rank sum test** for testing whether there is a difference between two *medians*. The Wilcoxon rank sum test is almost as powerful as the pooled-variance and separate-variance t tests discussed in Section 10.1 under conditions appropriate to these tests and is likely to be more powerful when the assumptions of those t tests are not met. In addition, you can use the Wilcoxon rank sum test when you have only ordinal data, as often happens in consumer behavior and marketing research.

To perform the Wilcoxon rank sum test, you replace the values in the two samples of size n_1 and n_2 with their combined ranks (unless the data contained the ranks initially). You begin by defining $n = n_1 + n_2$ as the total sample size. Next, you assign the ranks so that rank 1 is given to the smallest of the n combined values, rank 2 is given to the second smallest, and so on, until rank n is given to the largest. If several values are tied, you assign each value the average of the ranks that otherwise would have been assigned had there been no ties.

Whenever the two sample sizes are unequal, n_1 represents the smaller sample and n_2 the larger sample. The Wilcoxon rank sum test statistic, T_1, is defined as the sum of the ranks assigned to the n_1 values in the smaller sample. (For equal-sized samples, either sample may be used for determining T_1.) For any integer value n, the sum of the first n consecutive integers is $n(n + 1)/2$. Therefore, T_1 plus T_2, the sum of the ranks assigned to the n_2 items in the second sample, must equal $n(n + 1)/2$. You can use Equation (12.8) to check the accuracy of your rankings.

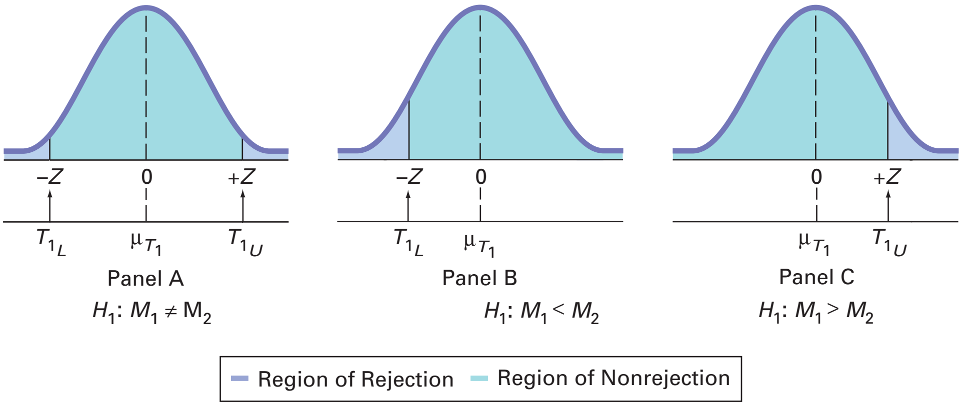
CHECKING THE RANKINGS

$$T_1 + T_2 = \frac{n(n + 1)}{2} \quad \textbf{(12.8)}$$

When n_1 and n_2 are both ≤ 10, you use Table E.6 to find the critical values of the test statistic T_1. For a two-tail test, you reject the null hypothesis (see Panel A of Figure 12.15) if the computed value of T_1 is greater than or equal to the upper critical value, or if T_1 is less than or equal to the lower critical value. For one-tail tests having the alternative hypothesis $H_1: M_1 < M_2$ [i.e., the median of population 1 (M_1) is less than the median of population 2 (M_2)], you reject the null hypothesis if the observed value of T_1 is less than or equal to the lower critical value (see Panel B of Figure 12.15). For one-tail tests having the alternative hypothesis $H_1: M_1 > M_2$, you reject the null hypothesis if the observed value of T_1 equals or is greater than the upper critical value (see Panel C of Figure 12.15).

FIGURE 12.15

Regions of rejection and nonrejection using the Wilcoxon rank sum test



For large sample sizes, the test statistic T_1 is approximately normally distributed, with the mean, μ_{T_1}, equal to

$$\mu_{T_1} = \frac{n_1(n + 1)}{2}$$

and the standard deviation, σ_{T_1}, equal to

$$\sigma_{T_1} = \sqrt{\frac{n_1 n_2(n + 1)}{12}}$$

Therefore, Equation (12.9) defines the standardized Z test statistic.

LARGE SAMPLE WILCOXON RANK SUM TEST

$$Z_{STAT} = \frac{T_1 - \dfrac{n_1(n + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n + 1)}{12}}} \quad \textbf{(12.9)}$$

where the test statistic Z_{STAT} approximately follows a standardized normal distribution.

You use Equation (12.9) for testing the null hypothesis when the sample sizes are outside the range of Table E.6. Based on α, the level of significance selected, you reject the null hypothesis if the Z_{STAT} test statistic falls in the rejection region.

To study an application of the Wilcoxon rank sum test, return to the Using Statistics scenario of Chapter 10 concerning sales of BLK Cola for the normal shelf display and end-aisle locations (see page 365). If you cannot assume that the populations are normally distributed, you can use the Wilcoxon rank sum test to evaluate possible differences in the median sales for the two display locations.[2] The data (stored in Cola) and the combined ranks are shown in Table 12.15.

[2]To test for differences in the median sales between the two locations, you must assume that the distributions of sales in both populations are identical except for differences in central tendency (i.e., the medians).

TABLE 12.15

Forming the Combined Rankings

Sales			
Normal Display ($n_1 = 10$)	**Combined Ranking**	**End-Aisle Display ($n_2 = 10$)**	**Combined Ranking**
22	1.0	52	5.5
34	3.0	71	14.0
52	5.5	76	15.0
62	10.0	54	7.0
30	2.0	67	13.0
40	4.0	83	17.0
64	11.0	66	12.0
84	18.5	90	20.0
56	8.0	77	16.0
59	9.0	84	18.5

Source: Data are taken from Table 10.1 on page 367.

Because you have not stated in advance which display location is likely to have a higher median, you use a two-tail test with the following null and alternative hypotheses:

$$H_0: M_1 = M_2 \text{ (the median sales are equal)}$$
$$H_1: M_1 \neq M_2 \text{ (the median sales are not equal)}$$

Now you need to compute T_1, the sum of the ranks assigned to the *smaller* sample. When the sample sizes are equal, as in this example, you can define either sample as the group from which to compute T_1. Choosing the normal display as the first sample,

$$T_1 = 1 + 3 + 5.5 + 10 + 2 + 4 + 11 + 18.5 + 8 + 9 = 72$$

As a check on the ranking procedure, you compute T_2 from

$$T_2 = 5.5 + 14 + 15 + 7 + 13 + 17 + 12 + 20 + 16 + 18.5 = 138$$

and then use Equation (12.8) on page 495 to show that the sum of the first $n = 20$ integers in the combined ranking is equal to $T_1 + T_2$:

$$T_1 + T_2 = \frac{n(n+1)}{2}$$
$$72 + 138 = \frac{20(21)}{2} = 210$$
$$210 = 210$$

Next, you use Table E.6 to determine the lower- and upper-tail critical values for the test statistic T_1. From Table 12.16, a portion of Table E.6, observe that for a level of significance of 0.05, the critical values are 78 and 132. The decision rule is

$$\text{Reject } H_0 \text{ if } T_1 \leq 78 \text{ or if } T_1 \geq 132;$$
$$\text{otherwise, do not reject } H_0.$$

TABLE 12.16

Finding the Lower- and Upper-Tail Critical Values for the Wilcoxon Rank Sum Test Statistic, T_1, Where $n_1 = 10$, $n_2 = 10$, and $\alpha = 0.05$

	α		n_1						
n_2	One-Tail	Two-Tail	4	5	6	7 (Lower, Upper)	8	9	$\boxed{10}$
9	.05	.10	16,40	24,51	33,63	43,76	54,90	66,105	
	.025	.05	14,42	22,53	31,65	40,79	51,93	62,109	
	.01	.02	13,43	20,55	28,68	37,82	47,97	59,112	
	.005	.01	11,45	18,57	26,70	35,84	45,99	56,115	
10	.05	.10	17,43	26,54	35,67	45,81	56,96	69,111	82,128
$\boxed{10}$	.025	.05	15,45	23,57	32,70	42,84	53,99	65,115	$\boxed{78,132}$
	.01	.02	13,47	21,59	29,73	39,87	49,103	61,119	74,136
	.005	.01	12,48	19,61	27,75	37,89	47,105	58,122	71,139

Source: Extracted from Table E.6.

Because the test statistic $T_1 = 72 < 78$, you reject H_0. There is evidence of a significant difference in the median sales for the two display locations. Because the sum of the ranks is lower for the normal display, you conclude that median sales are lower for the normal display.

Figure 12.16 shows the Wilcoxon rank sum test results (Excel) and the Mann-Whitney test results (Minitab) for the BLK Cola sales data. Although the Mann-Whitney test computes a different test statistic, the test is numerically equivalent to the Wilcoxon rank sum test (see references 1, 2, and 9). From the Figure 12.16 Excel results, you reject the null hypothesis because the p-value is 0.0126, which is less than $\alpha = 0.05$. This p-value indicates that if the medians of the two populations are equal, the chance of finding a difference at least this large in the samples is only 0.0126. The Figure 12.16 Minitab results report "The test is significant at 0.0139." The slight difference in the results is due to the fact that Minitab is computing an exact probability and Excel is using the normal approximation. Minitab also computes the p-value adjusted for ties.

FIGURE 12.16

Wilcoxon rank sum test (Excel) and Mann-Whitney test (Minitab) results for the BLK Cola sales data

	A	B	
1	**Wilcoxon Rank Sum Test**		
2			
3	**Data**		
4	Level of Significance	0.05	
5			
6	**Population 1 Sample**		
7	Sample Size	10	=COUNTIF(SortedRanks!A2:A21,"Normal")
8	Sum of Ranks	72	=SUMIF(SortedRanks!A2:A21,"Normal",SortedRanks!C2:C21)
9	**Population 2 Sample**		
10	Sample Size	10	=COUNTIF(SortedRanks!A2:A21,"End-Aisle")
11	Sum of Ranks	138	=SUMIF(SortedRanks!A2:A21,"End-Aisle",SortedRanks!C2:C21)
12			
13	Intermediate Calculations		
14	Total Sample Size n	20	=B7 + B10
15	*T1* Test Statistic	72	=IF(B7 <= B10, B8, B11)
16	*T1* Mean	105	=IF(B7 <= B10, B7 * (B14 + 1)/2, B10 * (B14 + 1)/2)
17	Standard Error of *T1*	13.2288	=SQRT(B7 * B10 * (B14 + 1)/12)
18	*Z* Test Statistic	-2.4946	=(B15 - B16)/B17
19			
20	**Two-Tail Test**		
21	Lower Critical Value	-1.9600	=NORMSINV(B4/2)
22	Upper Critical Value	1.9600	=NORMSINV(1 - B4/2)
23	*p*-Value	0.0126	=2 * (1 - NORMSDIST(ABS(B18)))
24	Reject the null hypothesis		=IF(B23 < B4, "Reject the null hypothesis", "Do not reject the null hypothesis")

Mann-Whitney Test and CI: Normal, End-Aisle

```
            N  Median
Normal     10   54.00
End-Aisle  10   73.50

Point estimate for ETA1-ETA2 is -21.50
95.5 Percent CI for ETA1-ETA2 is (-37.01,-6.00)
W = 72.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0140
The test is significant at 0.0139 (adjusted for ties)
```

Table E.6 shows the lower and upper critical values of the Wilcoxon rank sum test statistic, T_1, but only for situations in which both n_1 and n_2 are less than or equal to 10. If either one or both of the sample sizes are greater than 10, you *must* use the large-sample Z approximation formula [Equation (12.9) on page 495]. However, you can also use this approximation formula for small sample sizes. To demonstrate the large-sample Z approximation formula, consider the BLK Cola sales data. Using Equation (12.9),

$$Z_{STAT} = \frac{T_1 - \dfrac{n_1(n+1)}{2}}{\sqrt{\dfrac{n_1 n_2(n+1)}{12}}}$$

$$= \frac{72 - \dfrac{(10)(21)}{2}}{\sqrt{\dfrac{(10)(10)(21)}{12}}}$$

$$= \frac{72 - 105}{13.2288} = -2.4946$$

Because $Z_{STAT} = -2.4946 < -1.96$, the critical value of Z at the 0.05 level of significance (or p-*value* $= 0.0126 < 0.05$), you reject H_0.

Problems for Section 12.6

LEARNING THE BASICS

12.47 Using Table E.6, determine the lower- and upper-tail critical values for the Wilcoxon rank sum test statistic, T_1, in each of the following two-tail tests:

a. $\alpha = 0.10, n_1 = 6, n_2 = 8$
b. $\alpha = 0.05, n_1 = 6, n_2 = 8$
c. $\alpha = 0.01, n_1 = 6, n_2 = 8$
d. Given your results in (a) through (c), what do you conclude regarding the width of the region of nonrejection as the selected level of significance, α, gets smaller?

12.48 Using Table E.6, determine the lower-tail critical value for the Wilcoxon rank sum test statistic, T_1, in each of the following one-tail tests:

a. $\alpha = 0.05, n_1 = 6, n_2 = 8$
b. $\alpha = 0.025, n_1 = 6, n_2 = 8$
c. $\alpha = 0.01, n_1 = 6, n_2 = 8$
d. $\alpha = 0.005, n_1 = 6, n_2 = 8$

12.49 The following information is available for two samples selected from independent populations:

Sample 1: $n_1 = 7$ Assigned ranks: 4 1 8 2 5 10 11

Sample 2: $n_2 = 9$ Assigned ranks: 7 16 12 9 3 14 13 6 15

What is the value of T_1 if you are testing the null hypothesis H_0: $M_1 = M_2$?

12.50 In Problem 12.49, what are the lower- and upper-tail critical values for the test statistic T_1 from Table E.6 if you use a 0.05 level of significance and the alternative hypothesis is H_1: $M_1 \neq M_2$?

12.51 In Problems 12.49 and 12.50, what is your statistical decision?

12.52 The following information is available for two samples selected from independent and similarly shaped right-skewed populations:

Sample 1: $n_1 = 5$ 1.1 2.3 2.9 3.6 14.7

Sample 2: $n_2 = 6$ 2.8 4.4 4.4 5.2 6.0 18.5

a. Replace the observed values with the corresponding ranks (where 1 = smallest value; $n = n_1 + n_2 = 11$ = largest value) in the combined samples.
b. What is the value of the test statistic T_1?
c. Compute the value of T_2, the sum of the ranks in the larger sample.
d. To check the accuracy of your rankings, use Equation (12.8) on page 495 to demonstrate that

$$T_1 + T_2 = \frac{n(n+1)}{2}$$

12.53 From Problem 12.52, at the 0.05 level of significance, determine the lower-tail critical value for the Wilcoxon rank sum test statistic, T_1, if you want to test the null hypothesis, $H_0: M_1 \geq M_2$, against the one-tail alternative, $H_1: M_1 < M_2$.

12.54 In Problems 12.52 and 12.53, what is your statistical decision?

APPLYING THE CONCEPTS

12.55 A vice president for marketing recruits 20 college graduates for management training. Each of the 20 individuals is randomly assigned, 10 each, to one of two groups. A "traditional" method of training (*T*) is used in one group, and an "experimental" method (*E*) is used in the other. After the graduates spend six months on the job, the vice president ranks them on the basis of their performance, from 1 (worst) to 20 (best), with the following results (stored in the file **TestRank**):

T:	1	2	3	5	9	10	12	13	14	15
E:	4	6	7	8	11	16	17	18	19	20

Is there evidence of a difference in the median performance between the two methods? (Use $\alpha = 0.05$.)

12.56 Wine experts Gaiter and Brecher use a six-category scale when rating wines: Yech, OK, Good, Very Good, Delicious, and Delicious! (Data extracted from D. Gaiter and J. Brecher, "A Good U.S. Cabernet Is Hard to Find," *The Wall Street Journal*, May 19, 2006, p. W7.) Suppose Gaiter and Brecher tested a random sample of eight inexpensive California Cabernets and a random sample of eight inexpensive Washington Cabernets. *Inexpensive* is defined as a suggested retail value in the United States of under \$20. The data, stored in **Cabernet**, are as follows:

California—Good, Delicious, Yech, OK, OK, Very Good, Yech, OK

Washington—Very Good, OK, Delicious!, Very Good, Delicious, Good, Delicious, Delicious!

a. Are the data collected by rating wines using this scale nominal, ordinal, interval, or ratio?

b. Why is the two-sample *t* test defined in Section 10.1 inappropriate to test the mean rating of California Cabernets versus Washington Cabernets?

c. Is there evidence of a significance difference in the median rating of California Cabernets and Washington Cabernets? (Use $\alpha = 0.05$.)

12.57 A problem with a telephone line that prevents a customer from receiving or making calls is upsetting to both the customer and the telephone company. The file **Phone** contains samples of 20 problems reported to two different offices of a telephone company and the time to clear these problems (in minutes) from the customers' lines:

Central Office I Time to Clear Problems (Minutes)

1.48	1.75	0.78	2.85	0.52	1.60	4.15	3.97	1.48	3.10
1.02	0.53	0.93	1.60	0.80	1.05	6.32	3.93	5.45	0.97

Central Office II Time to Clear Problems (Minutes)

7.55	3.75	0.10	1.10	0.60	0.52	3.30	2.10	0.58	4.02
3.75	0.65	1.92	0.60	1.53	4.23	0.08	1.48	1.65	0.72

a. Is there evidence of a difference in the median time to clear problems between offices? (Use $\alpha = 0.05$.)

b. What assumptions must you make in (a)?

c. Compare the results of (a) with those of Problem 10.9(a) on page 375.

SELF Test **12.58** The management of a hotel has the business objective of increasing the return rate for hotel guests. One aspect of first impressions by guests relates to the time it takes to deliver a guest's luggage to the room after check-in to the hotel. A random sample of 20 deliveries on a particular day were selected in Wing A of the hotel, and a random sample of 20 deliveries were selected in Wing B. The results are stored in **Luggage**.

a. Is there evidence of a difference in the median delivery times in the two wings of the hotel? (Use $\alpha = 0.05$.)

b. Compare the results of (a) with those of Problem 10.67 on page 402.

12.59 The lengths of life (in hours) of a sample of 40 100-watt light bulbs produced by Manufacturer A and a sample of 40 100-watt light bulbs produced by Manufacturer B are stored in **Bulbs**.

a. Using a 0.05 level of significance, is there evidence of a difference in the median life of bulbs produced by the two manufacturers?

b. What assumptions must you make in (a)?

c. Compare the results of (a) with those of Problem 10.66 (a) on page 402. Discuss.

12.60 Nondestructive evaluation is used to measure the properties of components or materials without causing any permanent physical change to the components or materials. It includes the determination of properties of materials and the classification of flaws by size, shape, type, and location. Nondestructive evaluation is very effective for detecting surface flaws and characterizing surface properties of electrically conductive materials. Data were collected that classified each component as having a flaw or not, based on manual inspection and operator judgment, and also reported the size of the crack in the material. The results in terms of crack size (in inches) are stored in **Crack** and shown below. (Data extracted from B. D. Olin and W. Q. Meeker, "Applications of Statistical Methods to Nondestructive Evaluation," *Technometrics*, 38, 1996, p. 101.)

Unflawed

0.003	0.004	0.012	0.014	0.021	0.023	0.024	0.030	0.034
0.041	0.041	0.042	0.043	0.045	0.057	0.063	0.074	0.076

Flawed

0.022 0.026 0.026 0.030 0.031 0.034 0.042 0.043 0.044
0.046 0.046 0.052 0.055 0.058 0.060 0.060 0.070 0.071
0.073 0.073 0.078 0.079 0.079 0.083 0.090 0.095 0.095
0.096 0.100 0.102 0.103 0.105 0.114 0.119 0.120 0.130
0.160 0.306 0.328 0.440

a. Using a 0.05 level of significance, is there evidence that the median crack size is smaller for unflawed components than for flawed components?
b. What assumptions must you make in (a)?
c. Compare the results of (a) with those of Problem 10.17 (a) on page 376. Discuss.

12.61 A bank with a branch located in a commercial district of a city has developed an improved process for serving customers during the noon-to-1 P.M. lunch period. The bank has the business objective of reducing the waiting time (defined as the time elapsed from when the customer enters the line until he or she reaches the teller window) to increase customer satisfaction. A random sample of 15 customers is selected (and stored in Bank1); the results (in minutes) are as follows:

4.21 5.55 3.02 5.13 4.77 2.34 3.54 3.20
4.50 6.10 0.38 5.12 6.46 6.19 3.79

Another branch, located in a residential area, is also concerned with the noon-to-1 P.M. lunch period. A random sample of 15 customers is selected (and stored in the file Bank2); the results (in minutes) are as follows:

9.66 5.90 8.02 5.79 8.73 3.82 8.01 8.35
10.49 6.68 5.64 4.08 6.17 9.91 5.47

a. Is there evidence of a difference in the median waiting time between the two branches? (Use $\alpha = 0.05$.)
b. What assumptions must you make in (a)?
c. Compare the results (a) with those of Problem 10.12 (a) on page 376. Discuss.

12.62 Digital cameras have taken over the majority of the point-and-shoot camera market. One of the important features of a camera is the battery life, as measured by the number of shots taken until the battery needs to be recharged. The file DigitalCameras contains the battery life of 29 subcompact cameras and 16 compact cameras. (Data extracted from "Digital Cameras," *Consumer Reports*, July 2009, pp. 28–29.)
a. Is there evidence of a difference in the median battery life between subcompact cameras and compact cameras? (Use $\alpha = 0.05$.)
b. What assumptions must you make in (a)?
c. Compare the results of (a) with those of Problem 10.11 (a) on page 376. Discuss.

12.7 Kruskal-Wallis Rank Test: Nonparametric Analysis for the One-Way ANOVA

If the normality assumption of the one-way ANOVA F test is violated, you can use the Kruskal-Wallis rank test. The Kruskal-Wallis rank test for differences among c medians (where $c > 2$) is an extension of the Wilcoxon rank sum test for two independent populations, discussed in Section 12.6. Thus, the Kruskal-Wallis test has the same power relative to the one-way ANOVA F test that the Wilcoxon rank sum test has relative to the t test.

You use the **Kruskal-Wallis rank test** to test whether c independent groups have equal medians. The null hypothesis is

$$H_0: M_1 = M_2 = \cdots = M_c$$

and the alternative hypothesis is

$$H_1: \text{Not all } M_j \text{ are equal (where } j = 1, 2, \ldots, c\text{)}.$$

To use the Kruskal-Wallis rank test, you first replace the values in the c samples with their combined ranks (if necessary). Rank 1 is given to the smallest of the combined values and rank n to the largest of the combined values (where $n = n_1 + n_2 + \cdots + n_c$). If any values are tied, you assign each of them the mean of the ranks they would have otherwise been assigned if ties had not been present in the data.

The Kruskal-Wallis test is an alternative to the one-way ANOVA F test. Instead of comparing each of the c group means against the grand mean, the Kruskal-Wallis test compares the mean rank in each of the c groups against the overall mean rank, based on all n combined values. Equation (12.10) defines the Kruskal-Wallis test statistic, H.

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCES AMONG C MEDIANS

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right] - 3(n+1) \quad \textbf{(12.10)}$$

where

n = total number of values over the combined samples

n_j = number of values in the jth sample ($j = 1, 2, \ldots, c$)

T_j = sum of the ranks assigned to the jth sample

T_j^2 = square of the sum of the ranks assigned to the jth sample

c = number of groups

If there is a significant difference among the c groups, the mean rank differs considerably from group to group. In the process of squaring these differences, the test statistic H becomes large. If there are no differences present, the test statistic H is small because the mean of the ranks assigned in each group should be very similar from group to group.

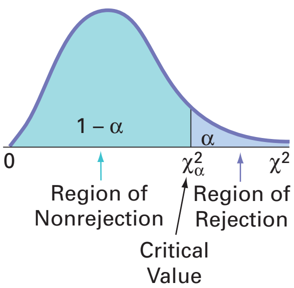
As the sample sizes in each group get large (i.e., at least 5), you can approximate the test statistic, H, by the chi-square distribution with $c - 1$ degrees of freedom. Thus, you reject the null hypothesis if the computed value of H is greater than the upper-tail critical value (see Figure 12.17). Therefore, the decision rule is

$$\text{Reject } H_0 \text{ if } H > \chi^2_\alpha;$$

$$\text{otherwise, do not reject } H_0.$$

FIGURE 12.17

Determining the rejection region for the Kruskal-Wallis test



To illustrate the Kruskal-Wallis rank test for differences among c medians, return to the Using Statistics scenario from Chapter 11 on page 415, concerning the strength of parachutes. If you cannot assume that the tensile strength of the parachutes is normally distributed in all c groups, you can use the Kruskal-Wallis rank test.

The null hypothesis is that the median tensile strengths of parachutes for the four suppliers are equal. The alternative hypothesis is that at least one of the suppliers differs from the others:

$$H_0: M_1 = M_2 = M_3 = M_4$$

$$H_1: \text{Not all } M_j \text{ are equal (where } j = 1, 2, 3, 4).$$

Table 12.17 presents the data (stored in the file **Parachute**), along with the corresponding ranks.

TABLE 12.17

Tensile Strength and Ranks of Parachutes Woven from Synthetic Fibers from Four Suppliers

Supplier							
1		2		3		4	
Amount	Rank	Amount	Rank	Amount	Rank	Amount	Rank
18.5	4	26.3	20	20.6	8	25.4	19
24.0	13.5	25.3	18	25.2	17	19.9	5.5
17.2	1	24.0	13.5	20.8	9	22.6	11
19.9	5.5	21.2	10	24.7	16	17.5	2
18.0	3	24.5	15	22.9	12	20.4	7

In converting the 20 tensile strengths to ranks, observe in Table 12.17 that the third parachute for Supplier 1 has the lowest tensile strength, 17.2. It is assigned a rank of 1. The fourth value for Supplier 1 and the second value for Supplier 4 each have a value of 19.9. Because they are tied for ranks 5 and 6, each is assigned the rank 5.5. Finally, the first value for Supplier 2 is the largest value, 26.3, and is assigned a rank of 20.

After all the ranks are assigned, you compute the sum of the ranks for each group:

$$\text{Rank sums: } T_1 = 27 \;\; T_2 = 76.5 \;\; T_3 = 62 \;\; T_4 = 44.5$$

As a check on the rankings, recall from Equation (12.8) on page 495 that for any integer n, the sum of the first n consecutive integers is $n(n + 1)/2$. Therefore,

$$\begin{aligned} T_1 + T_2 + T_3 + T_4 &= \frac{n(n+1)}{2} \\ 27 + 76.5 + 62 + 44.5 &= \frac{(20)(21)}{2} \\ 210 &= 210 \end{aligned}$$

To test the null hypothesis of equal population medians, you calculate the test statistic H using Equation (12.10) on page 501:

$$\begin{aligned} H &= \left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right] - 3(n+1) \\ &= \left\{\frac{12}{(20)(21)}\left[\frac{(27)^2}{5} + \frac{(76.5)^2}{5} + \frac{(62)^2}{5} + \frac{(44.5)^2}{5}\right]\right\} - 3(21) \\ &= \left(\frac{12}{420}\right)(2{,}481.1) - 63 = 7.8886 \end{aligned}$$

The test statistic H approximately follows a chi-square distribution with $c - 1$ degrees of freedom. Using a 0.05 level of significance, χ^2_α, the upper-tail critical value of the chi-square distribution with $c - 1 = 3$ degrees of freedom, is 7.815 (see Table 12.18). Because the computed value of the test statistic $H = 7.8886$ is greater than the critical value, you reject the null hypothesis and conclude that the median tensile strength is not the same for all the suppliers. You reach the same conclusion by using the p-value approach, because, as shown in Figure 12.18, $p\text{-value} = 0.0484 < 0.05$. At this point, you could simultaneously compare all pairs of suppliers to determine which ones differ (see reference 2).

TABLE 12.18

Finding χ^2_{α}, the Upper-Tail Critical Value for the Kruskal-Wallis Rank Test, at the 0.05 Level of Significance with 3 Degrees of Freedom

	Cumulative Area									
	.005	.01	.025	.05	.10	.25	.75	.90	$\boxed{.95}$	.975
	Upper-Tail Area									
Degrees of Freedom	.995	.99	.975	.95	.90	.75	.25	.10	$\boxed{.05}$	.025
1	—	—	0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378
$\boxed{3}$	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	$\boxed{7.815}$	9.348
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.071	12.833

Source: Extracted from Table E.4.

FIGURE 12.18

Excel and Minitab Kruskal-Wallis rank test results for differences among the four median tensile strengths of parachutes

	A	B	C	D	E	F	G
1	Kruskal-Wallis Rank Test						
2							
3	Data			Calculations			
4	Level of Significance	0.05		Group	Sample Size	Sum of Ranks	Mean Ranks
5				1	5	27	5.4
6	Intermediate Calculations			2	5	76.5	15.3
7	Sum of Squared Ranks/Sample Size	2481.1		3	5	62	12.4
8	Sum of Sample Sizes	20		4	5	44.5	8.9
9	Number of Groups	4					
10							
11	Test Result						
12	*H* Test Statistic	7.8886	=(12/(B8 * (B8 + 1))) * B7 - (3 * (B8 + 1))				
13	Critical Value	7.8147	=CHIINV(B4, B9 - 1)				
14	*p*-Value	0.0484	=CHIDIST(B12, B9 - 1)				
15	Reject the null hypothesis		=IF(B14 < B4, "Reject the null hypothesis", "Do not reject the null hypothesis")				

Also
Cell B7: =(G5 * F5) + (G6 * F6) + (G7 * F7) + (G8 * F8)
Cell B8: =SUM(E5:E8)

Kruskal-Wallis Test: Strength versus Supplier

```
Kruskal-Wallis Test on Strength

Supplier     N  Median  Ave Rank      Z
Supplier 1   5   18.50       5.4  -2.23
Supplier 2   5   24.50      15.3   2.09
Supplier 3   5   22.90      12.4   0.83
Supplier 4   5   20.40       8.9  -0.70
Overall     20              10.5

H = 7.89  DF = 3  P = 0.048
H = 7.90  DF = 3  P = 0.048  (adjusted for ties)
```

The following assumptions are needed to use the Kruskal-Wallis rank test:

- The c samples are randomly and independently selected from their respective populations.
- The underlying variable is continuous.
- The data provide at least a set of ranks, both within and among the c samples.
- The c populations have the same variability.
- The c populations have the same shape.

The Kruskal-Wallis procedure makes less stringent assumptions than does the F test. If you ignore the last two assumptions (variability and shape), you can still use the Kruskal-Wallis rank test to determine whether at least one of the populations differs from the other populations in some characteristic—such as central tendency, variation, or shape.

To use the F test, you must assume that the c samples are from normal populations that have equal variances. When the more stringent assumptions of the F test hold, you should use the F test instead of the Kruskal-Wallis test because it has slightly more power to detect significant differences among groups. However, if the assumptions of the F test do not hold, you should use the Kruskal-Wallis test.

Problems for Section 12.7

LEARNING THE BASICS

12.63 What is the upper-tail critical value from the chi-square distribution if you use the Kruskal-Wallis rank test for comparing the medians in six populations at the 0.01 level of significance?

12.64 For this problem, use the results of Problem 12.63.

a. State the decision rule for testing the null hypothesis that all six groups have equal population medians.

b. What is your statistical decision if the computed value of the test statistic H is 13.77?

APPLYING THE CONCEPTS

12.65 A pet food company has the business objective of expanding its product line beyond its current kidney- and shrimp-based cat foods. The company developed two new products—one based on chicken livers and the other based on salmon. The company conducted an experiment to compare the two new products with its two existing ones as well as a generic beef-based product sold in a supermarket chain.

For the experiment, a sample of 50 cats from the population at a local animal shelter was selected. Ten cats were randomly assigned to each of the five products being tested. Each of the cats was then presented with 3 ounces of the selected food in a dish at feeding time. The researchers defined the variable to be measured as the number of ounces of food that the cat consumed within a 10-minute time interval that began when the filled dish was presented. The results for this experiment are summarized in the following table and stored in CatFood:

Kidney	Shrimp	Chicken Liver	Salmon	Beef
2.37	2.26	2.29	1.79	2.09
2.62	2.69	2.23	2.33	1.87
2.31	2.25	2.41	1.96	1.67
2.47	2.45	2.68	2.05	1.64
2.59	2.34	2.25	2.26	2.16
2.62	2.37	2.17	2.24	1.75
2.34	2.22	2.37	1.96	1.18
2.47	2.56	2.26	1.58	1.92
2.45	2.36	2.45	2.18	1.32
2.32	2.59	2.57	1.93	1.94

a. At the 0.05 level of significance, is there evidence of a significant difference in the median amount of food eaten among the various products?

b. Compare the results of (a) with those of Problem 11.13 (a) on page 429.

c. Which test is more appropriate for these data, the Kruskal-Wallis rank test or the one-way ANOVA F test? Explain.

SELF Test **12.66** A hospital conducted a study of the waiting time in its emergency room. The hospital has a main campus, along with three satellite locations. Management had a business objective of reducing waiting time for emergency room cases that did not require immediate attention. To study this, a random sample of 15 emergency room cases at each location were selected on a particular day, and the waiting time (recorded from check-in to when the patient was called into the clinic area) was measured. The results are stored in ERWaiting.

a. At the 0.05 level of significance, is there evidence of a difference in the median waiting times in the four locations?

b. Compare the results of (a) with those of Problem 11.9 (a) on page 428.

12.67 The per-store daily customer count (i.e., the mean number of customers in a store in one day) for a nationwide convenience store chain that operates nearly 10,000 stores has been steady, at 900, for some time. To increase the customer count, the chain is considering cutting prices for coffee beverages. Management needs to determine how much prices can be cut in order to increase the daily customer count without reducing the gross margin on coffee sales too much. You decide to carry out an experiment in a sample of 24 stores where customer counts have been running almost exactly at the national average of 900. In 6 of the stores, a small coffee will be \$0.59, in another 6 stores the price will be \$0.69, in a third group of 6 stores, the price will be \$0.79, and in a fourth group of 6 stores, the price will now be \$0.89. After four weeks, the daily customer count in the stores is stored in CoffeeSales.

a. At the 0.05 level of significance, is there evidence of a difference in the daily customer count based on the price of a small coffee?

b. Compare the results of (a) with those of Problem 11.11 (a) on page 428.

12.68 An advertising agency has been hired by a manufacturer of pens to develop an advertising campaign for the upcoming holiday season. To prepare for this project, the research director decides to initiate a study of the effect of advertising on product perception. An experiment is designed to compare five different advertisements. Advertisement A greatly undersells the pen's characteristics. Advertisement B slightly undersells the pen's characteristics. Advertisement C slightly oversells the pen's characteristics. Advertisement D greatly oversells the pen's characteristics. Advertisement E attempts to correctly state the pen's characteristics. A sample of 30 adult respondents, taken from a larger focus group, is randomly assigned to the five advertisements (so that there are six respondents to each). After reading the advertisement and developing a sense of product expectation, all respondents unknowingly receive the same pen to evaluate. The respondents are permitted to test the pen and the plausibility of the advertising copy. The respondents are then asked to rate the pen from 1 to 7 on the product characteristic scales of appearance, durability, and writing performance. The *combined* scores of three ratings (appearance, durability, and writing performance) for the 30 respondents (stored in **Pen**) are as follows:

A	B	C	D	E
15	16	8	5	12
18	17	7	6	19
17	21	10	13	18
19	16	15	11	12
19	19	14	9	17
20	17	14	10	14

a. At the 0.05 level of significance, is there evidence of a difference in the median ratings of the five advertisements?

b. Compare the results of (a) with those of Problem 11.10 (a) on page 428.

c. Which test is more appropriate for these data, the Kruskal-Wallis rank test or the one-way ANOVA F test? Explain.

12.69 A sporting goods manufacturing company wanted to compare the distance traveled by golf balls produced using each of four different designs. Ten balls of each design were manufactured and brought to the local golf course for the club professional to test. The order in which the balls were hit with the same club from the first tee was randomized so that the pro did not know which type of ball was being hit. All 40 balls were hit in a short period of time, during which the environmental conditions were essentially the same. The results (distance traveled in yards) for the four designs are stored in **Golfball**:

a. At the 0.05 level of significance, is there evidence of a difference in the median distances traveled by the golf balls with different designs?

b. Compare the results of (a) with those of Problem 11.14 (a) on page 429.

12.70 Students in a business statistics course performed an experiment to test the strength of four brands of trash bags. One-pound weights were placed into a bag, one at a time, until the bag broke. A total of 40 bags were used (10 for each brand). The file **Trashbags** gives the weight (in pounds) required to break the trash bags.

a. At the 0.05 level of significance, is there evidence of a difference in the median strength of the four brands of trash bags?

b. Compare the results in (a) to those in Problem 11.8 (a) on page 428.

12.8 *Online Topic:* Wilcoxon Signed Ranks Test: Nonparametric Analysis for Two Related Populations

In Section 10.2, you used the paired t test to compare the means of two populations when you had repeated measures or matched samples. The paired t test assumes that the data are measured on an interval or a ratio scale and are normally distributed. If you cannot make these assumptions, you can use the nonparametric **Wilcoxon signed ranks test** to test for the median difference. To study this topic, read the **Section 12.8** online topic file that is available on this book's companion website. (See Appendix C to learn how to access the online topic files.)

12.9 *Online Topic:* Friedman Rank Test: Nonparametric Analysis for the Randomized Block Design

When analyzing a randomized block design, sometimes the data consist of only the ranks within each block. Other times, you cannot assume that the data from each of the c groups are from normally distributed populations. In these situations, you can use the **Friedman rank test**. To study this topic, read the **Section 12.9** online topic file that is available on this book's companion website. (See Appendix C to learn how to access the online topic files.)

USING STATISTICS @ T.C. Resort Properties Revisited

In the Using Statistics scenario, you were the manager of T.C. Resort Properties, a collection of five upscale hotels located on two tropical islands. To assess the quality of services being provided by your hotels, guests are encouraged to complete a satisfaction survey when they check out. You analyzed the data from these surveys to determine the overall satisfaction with the services provided, the likelihood that the guests will return to the hotel, and the reasons given by some guests for not wanting to return.

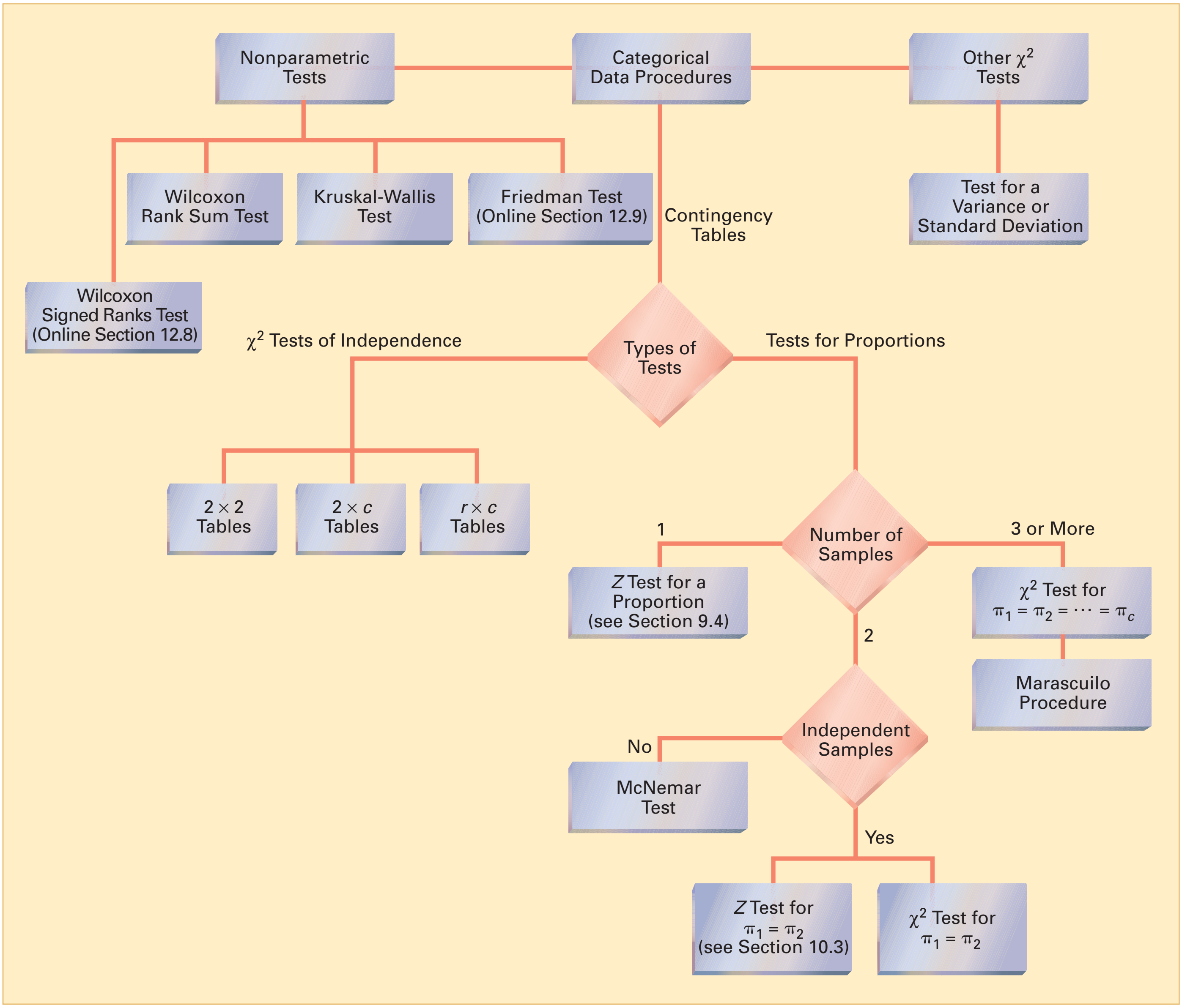
On one island, T.C. Resort Properties operates the Beachcomber and Windsurfer hotels. You performed a chi-square test for the difference in two proportions and concluded that a greater proportion of guests are willing to return to the Beachcomber Hotel than to the Windsurfer. On the other island, T.C. Resort Properties operates the Golden Palm, Palm Royale, and Palm Princess hotels. To see if guest satisfaction was the same among the three hotels, you performed a chi-square test for the differences among more than two proportions. The test confirmed that the three proportions are not equal, and guests are most likely to return to the Palm Royale and least likely to return to the Golden Palm.

In addition, you investigated whether the reasons given for not returning to the Golden Palm, Palm Royale, and Palm Princess were unique to a certain hotel or common to all three hotels. By performing a chi-square test of independence, you determined that the reasons given for wanting to return or not depended on the hotel where the guests had been staying. By examining the observed and expected frequencies, you concluded that guests were more satisfied with the price at the Golden Palm and were much more satisfied with the location of the Palm Princess. Guest satisfaction with room accommodations was not significantly different among the three hotels.

SUMMARY

Figure 12.19 presents a roadmap for this chapter. First, you used hypothesis testing for analyzing categorical response data from two independent samples and from more than two independent samples. In addition, the rules of probability from Section 4.2 were extended to the hypothesis of independence in the joint responses to two categorical variables. You also used the McNemar test to study situations where the samples were not independent. In addition, you used the chi-square distribution to test a variance or standard deviation. You also studied two nonparametric tests. You used the Wilcoxon rank sum test when the assumptions of the t test for two independent samples were violated and the Kruskal-Wallis test when the assumptions of the one-way ANOVA F test were violated.

FIGURE 12.19
Roadmap of Chapter 12



KEY EQUATIONS

χ^2 Test for the Difference Between Two Proportions

$$\chi^2_{STAT} = \sum_{\text{all cells}} \frac{(f_0 - f_e)^2}{f_e} \tag{12.1}$$

Computing the Estimated Overall Proportion for Two Groups

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n} \tag{12.2}$$

Computing the Estimated Overall Proportion for *c* Groups

$$\bar{p} = \frac{X_1 + X_2 + \cdots + X_c}{n_1 + n_2 + \cdots + n_c} = \frac{X}{n} \tag{12.3}$$

Critical Range for the Marascuilo Procedure

$$\text{Critical range} = \sqrt{\chi^2_\alpha}\sqrt{\frac{p_j(1 - p_j)}{n_j} + \frac{p_{j'}(1 - p_{j'})}{n_{j'}}} \tag{12.4}$$

Computing the Expected Frequency

$$f_e = \frac{\text{Row total} \times \text{Column total}}{n} \tag{12.5}$$

McNemar Test Statistic

$$Z_{STAT} = \frac{B - C}{\sqrt{B + C}} \tag{12.6}$$

χ^2 Test for the Variance or Standard Deviation

$$\chi^2_{STAT} = \frac{(n - 1)S^2}{\sigma^2} \tag{12.7}$$

Checking the Rankings

$$T_1 + T_2 = \frac{n(n + 1)}{2} \tag{12.8}$$

Large Sample Wilcoxon Rank Sum Test

$$Z_{STAT} = \frac{T_1 - \dfrac{n_1(n + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n + 1)}{12}}} \tag{12.9}$$

Kruskal-Wallis Rank Test for Differences Among *c* Medians

$$H = \left[\frac{12}{n(n + 1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right] - 3(n + 1) \tag{12.10}$$

KEY TERMS

CHAPTER REVIEW PROBLEMS

CHECKING YOUR UNDERSTANDING

12.71 Under what conditions should you use the χ^2 test to determine whether there is a difference between the proportions of two independent populations?

12.72 Under what conditions should you use the χ^2 test to determine whether there is a difference among the proportions of more than two independent populations?

12.73 Under what conditions should you use the χ^2 test of independence?

12.74 Under what conditions should you use the McNemar test?

12.75 What is a nonparametric procedure?

12.76 Under what conditions should you use the Wilcoxon rank sum test instead of the *t* test for the difference between the means?

12.77 Under what conditions should you use the Kruskal-Wallis rank test instead of the one-way ANOVA?

APPLYING THE CONCEPTS

12.78 Undergraduate students at Miami University in Oxford, Ohio, were surveyed in order to evaluate the effect of gender and price on purchasing a pizza from Pizza Hut. Students were told to suppose that they were planning to have a large two-topping pizza delivered to their residence that evening. The students had to decide between ordering from Pizza Hut at a reduced price of \$8.49 (the regular price for a large two-topping pizza from the Oxford Pizza Hut at this time was \$11.49) and ordering a pizza from a different pizzeria. The results from this question are summarized in the following contingency table:

	PIZZERIA		
GENDER	Pizza Hut	Other	Total
Female	4	13	17
Male	6	12	18
Total	10	25	35

A subsequent survey evaluated purchase decisions at other prices. These results are summarized in the following contingency table:

	PRICE			
PIZZERIA	**\$8.49**	**\$11.49**	**\$14.49**	**Total**
Pizza Hut	10	5	2	17
Other	25	23	27	75
Total	35	28	29	92

a. Using a 0.05 level of significance and using the data in the first contingency table, is there evidence of a significant difference between males and females in their pizzeria selection?
b. What is your answer to (a) if nine of the male students selected Pizza Hut and nine selected another pizzeria?
c. Using a 0.05 level of significance and using the data in the second contingency table, is there evidence of a difference in pizzeria selection based on price?
d. Determine the *p*-value in (c) and interpret its meaning.

12.79 According to U.S. Census estimates, there were about 20 million children between 8 and 12 years old (referred to as *tweens*) in the United States in 2009. A recent survey of 1,223 8- to 12-year-old children (S. Jayson, "It's Cooler Than Ever to Be a Tween," *USA Today*, February 4, 2009, pp. 1A, 2A) reported the following results. Suppose that the survey was based on 600 boys and 623 girls.

What Tweens Did in the Past Week	Boys	Girls
Played a game on a video game system	498	243
Read a book for fun	276	324
Gave product advice to parents	186	181
Shopped at a mall	144	262

For *each type of activity*, determine whether there is a difference between boys and girls at the 0.05 level of significance.

12.80 A company is considering an organizational change involving the use of self-managed work teams. To assess the attitudes of employees of the company toward this change, a sample of 400 employees is selected and asked whether they favor the institution of self-managed work teams in the organization. Three responses are permitted: favor, neutral, or oppose. The results of the survey, cross-classified by type of job and attitude toward self-managed work teams, are summarized as follows:

	SELF-MANAGED WORK TEAMS			
TYPE OF JOB	**Favor**	**Neutral**	**Oppose**	**Total**
Hourly worker	108	46	71	225
Supervisor	18	12	30	60
Middle management	35	14	26	75
Upper management	24	7	9	40
Total	185	79	136	400

a. At the 0.05 level of significance, is there evidence of a relationship between attitude toward self-managed work teams and type of job?

The survey also asked respondents about their attitudes toward instituting a policy whereby an employee could take one additional vacation day per month without pay. The results, cross-classified by type of job, are as follows:

	VACATION TIME WITHOUT PAY			
TYPE OF JOB	**Favor**	**Neutral**	**Oppose**	**Total**
Hourly worker	135	23	67	225
Supervisor	39	7	14	60
Middle management	47	6	22	75
Upper management	26	6	8	40
Total	247	42	111	400

b. At the 0.05 level of significance, is there evidence of a relationship between attitude toward vacation time without pay and type of job?

12.81 A company that produces and markets continuing education programs on DVDs for the educational testing industry has traditionally mailed advertising to prospective customers. A market research study was undertaken to compare two approaches: mailing a sample DVD upon request that contained highlights of the full DVD and sending an e-mail containing a link to a website from which sample material could be downloaded. Of those who responded to either the mailing or the e-mail, the results were as follows in terms of purchase of the complete DVD:

	TYPE OF MEDIA USED		
PURCHASED	**Mailing**	**E-mail**	**Total**
Yes	26	11	37
No	227	247	474
Total	253	258	511

a. At the 0.05 level of significance, is there evidence of a difference in the proportion of DVDs purchased on the basis of the type of media used?

b. On the basis of the results of (a), which type of media should the company use in the future? Explain the rationale for your decision.

The company also wanted to determine which of three sales approaches should be used to generate sales among those who either requested the sample DVD by mail or downloaded the sample DVD but did not purchase the full DVD: (1) targeted e-mail, (2) a DVD that contained additional features, or (3) a telephone call to prospective customers. The 474 respondents who did not initially purchase the full DVD were randomly assigned to each of the three sales approaches. The results, in terms of purchases of the full-program DVD, are as follows:

	SALES APPROACH			
ACTION	Targeted E-mail	More Complete DVD	Telephone Call	Total
Purchase	5	17	18	40
Don't purchase	153	141	140	434
Total	158	158	158	474

c. At the 0.05 level of significance, is there evidence of a difference in the proportion of DVDs purchased on the basis of the sales strategy used?

d. On the basis of the results of (c), which sales approach do you think the company should use in the future? Explain the rationale for your decision.

12.82 A market researcher investigated consumer preferences for Coca-Cola and Pepsi before a taste test and after a taste test. The following table summarizes the results from a sample of 200 consumers:

PREFERENCE BEFORE TASTE TEST	PREFERENCE AFTER TASTE TEST		
	Coca-Cola	Pepsi	Total
Coca-Cola	104	6	110
Pepsi	14	76	90
Total	118	82	200

a. Is there evidence of a difference in the proportion of respondents who prefer Coca-Cola before and after the taste tests? (Use $\alpha = 0.10$.)

b. Compute the p-value and interpret its meaning.

c. Show how the following table was derived from the table above:

	SOFT DRINK		
PREFERENCE	Coca-Cola	Pepsi	Total
Before taste test	110	90	200
After taste test	118	82	200
Total	228	172	400

d. Using the second table, is there evidence of a difference in preference for Coca-Cola before and after the taste test? (Use $\alpha = 0.05$.)

e. Determine the p-value and interpret its meaning.

f. Explain the difference in the results of (a) and (d). Which method of analyzing the data should you use? Why?

12.83 A market researcher was interested in studying the effect of advertisements on brand preference of people purchasing a new personal computer. Prospective purchasers of new computers were first asked whether they preferred Apple or Dell and then watched video advertisements of comparable models of the two brands. After viewing the ads, the prospective customers again indicated their preferences. The results are summarized in the following table:

	PREFERENCE AFTER ADS		
PREFERENCE BEFORE ADS	Apple	Dell	Total
Apple	97	3	100
Dell	11	89	100
Total	108	92	200

a. Is there evidence of a difference in the proportion of respondents who prefer Apple before and after viewing the ads? (Use $\alpha = 0.05$.)

b. Compute the p-value and interpret its meaning.

c. Show how the following table was derived from the table above:

	MANUFACTURER		
PREFERENCE	Apple	Dell	Total
Before ad	100	100	200
After ad	108	92	200
Total	208	192	400

d. Using the second table, is there evidence of a difference in preference for Apple before and after viewing the ads? (Use $\alpha = 0.05$.)

e. Determine the p-value and interpret its meaning.

f. Explain the difference in the results of (a) and (d). Which method of analyzing the data should you use? Why?

TEAM PROJECT

The file **Bond Funds** contains information regarding eight variables from a sample of 184 bond mutual funds:

Type—Type of bonds comprising the bond mutual fund (intermediate government or short-term corporate)
Assets—In millions of dollars
Fees—Sales charges (no or yes)
Expense ratio—Ratio of expenses to net assets in percentage
Return 2009—Twelve-month return in 2009
Three-year return—Annualized return, 2007–2009
Five-year return—Annualized return, 2005–2009
Risk—Risk-of-loss factor of the bond mutual fund (below average, average, or above average)

12.84 **a.** Construct a 2 × 2 contingency table, using fees as the row variable and type as the column variable.
b. At the 0.05 level of significance, is there evidence of a significant relationship between the type of bond mutual fund and whether there is a fee?

12.85 **a.** Construct a 2 × 3 contingency table, using fees as the row variable and risk as the column variable.
b. At the 0.05 level of significance, is there evidence of a significant relationship between the perceived risk of a bond mutual fund and whether there is a fee?

12.86 **a.** Construct a 3 × 2 contingency table, using risk as the row variable and category as the column variable.
b. At the 0.05 level of significance, is there evidence of a significant relationship between the category of a bond mutual fund and its perceived risk?

STUDENT SURVEY DATABASE

12.87 Problem 1.27 on page 14 describes a survey of 62 undergraduate students (stored in **UndergradSurvey**). For these data, construct contingency tables, using gender, major, plans to go to graduate school, and employment status. (You need to construct six tables, taking two variables at a time.) Analyze the data at the 0.05 level of significance to determine whether any significant relationships exist among these variables.

12.88 Problem 1.27 on page 14 describes a survey of 62 undergraduate students (stored in **UndergradSurvey**).
a. Select a sample of undergraduate students at your school and conduct a similar survey for those students.
b. For the data collected in (a), repeat Problem 12.87.
c. Compare the results of (b) to those of Problem 12.87.

12.89 Problem 1.28 on page 15 describes a survey of 44 MBA students (see the file **GradSurvey**). For these data, construct contingency tables, using gender, undergraduate major, graduate major, and employment status. (You need to construct six tables, taking two variables at a time.) Analyze the data at the 0.05 level of significance to determine whether any significant relationships exist among these variables.

12.90 Problem 1.28 on page 15 describes a survey of 44 MBA students (stored in **GradSurvey**).
a. Select a sample of graduate students in your MBA program and conduct a similar survey for those students.
b. For the data collected in (a), repeat Problem 12.89.
c. Compare the results of (b) to those of Problem 12.89.

MANAGING ASHLAND MULTICOMM SERVICES

Phase 1

Reviewing the results of its research, the marketing department team concluded that a segment of Ashland households might be interested in a discounted trial subscription to the AMS *3-For-All* cable/phone/Internet service. The team decided to test various discounts before determining the type of discount to offer during the trial period. It decided to conduct an experiment using three types of discounts plus a plan that offered no discount during the trial period:

1. No discount for the *3-For-All* cable/phone/Internet service. Subscribers would pay $24.99 per week for the *3-For-All* cable/phone/Internet service during the 90-day trial period.
2. Moderate discount for the *3-For-All* cable/phone/Internet service. Subscribers would pay $19.99 per week for the *3-For-All* cable/phone/Internet service during the 90-day trial period.
3. Substantial discount for the *3-For-All* cable/phone/Internet service. Subscribers would pay $14.99 per week for the *3-For-All* cable/phone/Internet service during the 90-day trial period.
4. Discount restaurant card. Subscribers would be given a card providing a discount of 15% at selected restaurants in Ashland during the trial period.

Each participant in the experiment was randomly assigned to a discount plan. A random sample of 100 subscribers to each plan during the trial period was tracked to determine how many would continue to subscribe to the *3-For-All* service after the trial period. Table AMS12.1 summarizes the results.

TABLE AMS12.1

Number of Subscribers Who Continue Subscriptions After Trial Period with Four Discount Plans

CONTINUE SUBSCRIPTIONS AFTER TRIAL PERIOD	DISCOUNT PLANS				
	No Discount	Moderate Discount	Substantial Discount	Restaurant Card	Total
Yes	24	30	38	51	143
No	76	70	62	49	257
Total	100	100	100	100	400

Exercise

1. Analyze the results of the experiment. Write a report to the team that includes your recommendation for which discount plan to use. Be prepared to discuss the limitations and assumptions of the experiment.

Phase 2

The marketing department team discussed the results of the survey presented in Chapter 8, on pages 317–318. The team realized that the evaluation of individual questions was providing only limited information. In order to further understand the market for the *3-For-All* cable/phone/Internet service, the data were organized in the following contingency tables:

HAS AMS TELEPHONE SERVICE	HAS AMS INTERNET SERVICE		
	Yes	No	Total
Yes	55	28	83
No	207	128	335
Total	262	156	418

TYPE OF SERVICE	DISCOUNT TRIAL		
	Yes	No	Total
Basic	8	156	164
Enhanced	32	222	254
Total	40	378	418

TYPE OF SERVICE	WATCHES PREMIUM OR ON-DEMAND SERVICES				
	Almost Every Day	Several Times a Week	Almost Never	Never	Total
Basic	2	5	127	30	164
Enhanced	12	30	186	26	254
Total	14	35	313	56	418

DISCOUNT	WATCHES PREMIUM OR ON-DEMAND SERVICES				
	Almost Every Day	Several Times a Week	Almost Never	Never	Total
Yes	4	5	27	4	40
No	10	30	286	52	378
Total	14	35	313	56	418

DISCOUNT	METHOD FOR CURRENT SUBSCRIPTION					
	Toll-Free Phone	AMS Website	Direct Mail Reply Card	Good Tunes & More	Other	Total
Yes	11	21	5	1	2	40
No	219	85	41	9	24	378
Total	230	106	46	10	26	418

GOLD CARD	METHOD FOR CURRENT SUBSCRIPTION					
	Toll-Free Phone	AMS Website	Direct Mail Reply Card	Good Tunes & More	Other	Total
Yes	10	20	5	1	2	38
No	220	86	41	9	24	380
Total	230	106	46	10	26	418

Exercise

2. Analyze the results of the contingency tables. Write a report for the marketing department team and discuss the marketing implications of the results for Ashland Multi-Comm Services.

DIGITAL CASE

Apply your knowledge of testing for the difference between two proportions in this Digital Case, which extends the T.C. Resort Properties Using Statistics scenario of this chapter.

As T.C. Resort Properties seeks to improve its customer service, the company faces new competition from SunLow Resorts. SunLow has recently opened resort hotels on the islands where T.C. Resort Properties has its five hotels. SunLow is currently advertising that a random survey of 300 customers revealed that about 60% of the customers preferred its "Concierge Class" travel reward program over the T.C. Resorts "TCRewards Plus" program.

Open and review **ConciergeClass.pdf**, an electronic brochure that describes the Concierge Class program and compares it to the T.C. Resorts program. Then answer the following questions:

1. Are the claims made by SunLow valid?

2. What analyses of the survey data would lead to a more favorable impression about T.C. Resort Properties?

3. Perform one of the analyses identified in your answer to step 2.

4. Review the data about the T.C. Resorts properties customers presented in this chapter. Are there any other questions that you might include in a future survey of travel reward programs? Explain.

REFERENCES

1. Conover, W. J., *Practical Nonparametric Statistics*, 3rd ed. (New York: Wiley, 2000).
2. Daniel, W. W., *Applied Nonparametric Statistics*, 2nd ed. (Boston: PWS Kent, 1990).
3. Dixon, W. J., and F. J. Massey, Jr., *Introduction to Statistical Analysis*, 4th ed. (New York: McGraw-Hill, 1983).
4. Hollander, M., and D. A. Wolfe, *Nonparametric Statistical Methods*, 2nd ed. (New York: Wiley, 1999).
5. Lewontin, R. C., and J. Felsenstein, "Robustness of Homogeneity Tests in $2 \times n$ Tables," *Biometrics* 21 (March 1965): 19–33.
6. Marascuilo, L. A., "Large-Sample Multiple Comparisons," *Psychological Bulletin* 65 (1966): 280–290.
7. Marascuilo, L. A., and M. McSweeney, *Nonparametric and Distribution-Free Methods for the Social Sciences* (Monterey, CA: Brooks/Cole, 1977).
8. *Microsoft Excel 2010* (Redmond, WA: Microsoft Corp., 2010).
9. *Minitab Release16* (State College, PA: Minitab, Inc., 2010).
10. Winer, B. J., D. R. Brown, and K. M. Michels, *Statistical Principles in Experimental Design*, 3rd ed. (New York: McGraw-Hill, 1989).

CHAPTER 12 EXCEL GUIDE

EG12.1 CHI-SQUARE TEST for the DIFFERENCE BETWEEN TWO PROPORTIONS

PHStat2 Use **Chi-Square Test for Differences in Two Proportions** to perform this chi-square test. For example, to perform the Figure 12.3 test for the two-hotel guest satisfaction data on page 472, select **PHStat ➔ Two-Sample Tests (Summarized Data) ➔ Chi-Square Test for Differences in Two Proportions**. In the procedure's dialog box, enter **0.05** as the **Level of Significance**, enter a **Title**, and click **OK**. In the new worksheet:

1. Read the yellow note about entering values and then press the **Delete** key to delete the note.
2. Enter **Hotel** in cell **B4** and **Choose Again?** in cell **A5**.
3. Enter **Beachcomber** in cell **B5** and **Windsurfer** in cell **C5**.
4. Enter **Yes** in cell **A6** and **No** in cell **A7**.
5. Enter **163**, **64**, **154**, and **108** in cells **B6**, **B7**, **C6**, and **C7**, respectively.

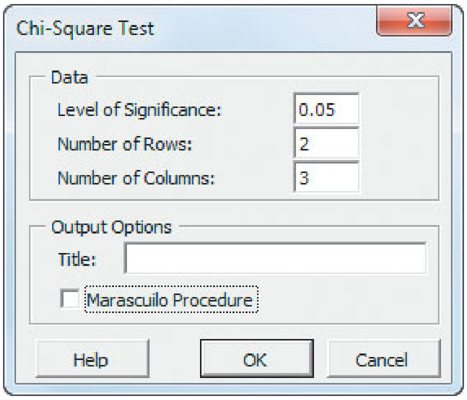
In-Depth Excel Use the **COMPUTE worksheet** of the **Chi-Square workbook**, shown in Figure 12.3 on page 472, as a template for performing this test. The worksheet contains the Table 12.3 two-hotel guest satisfaction data. Use the **CHIINV** and **CHIDIST** functions to help perform the chi-square test for the difference between two proportions. In cell B24, the worksheet uses **CHIINV(*level of significance, degrees of freedom*)** to compute the critical value for the test and in cell B26 uses **CHIDIST(*chi-square test statistic, degrees of freedom*)** to compute the *p*-value. Open to the **COMPUTE_FORMULAS worksheet** to examine the other formulas used in the worksheet.

For other problems, change the **Observed Frequencies** cell counts and row and column labels in rows 4 through 7.

EG12.2 CHI-SQUARE TEST for DIFFERENCES AMONG MORE THAN TWO PROPORTIONS

PHStat2 Use **Chi-Square Test** to perform the test for differences among more than two proportions. For example, to perform the Figure 12.6 test for the three-hotel guest satisfaction data on page 478, select **PHStat ➔ Multiple-Sample Tests ➔ Chi-Square Test**. In the procedure's dialog box (shown in the right column):

1. Enter **0.05** as the **Level of Significance**.
2. Enter **2** as the **Number of Rows**.
3. Enter **3** as the **Number of Columns**.
4. Enter a **Title** and click **OK**.



In the new worksheet:

5. Read the yellow note about entering values and then press the **Delete** key to delete the note.
6. Enter the Table 12.6 data on page 476, including row and column labels, in rows 4 through 7.

In-Depth Excel Use the **ChiSquare2x3 worksheet** of the **Chi-Square Worksheets workbook**, shown in Figure 12.6 on page 478, as a model for this chi-square test. The worksheet contains the data for Table 12.6 guest satisfaction data (see page 476). The worksheet uses formulas to compute the expected frequencies and the intermediate results for the chi-square test statistic in much the same way as the COMPUTE worksheet of the Chi-Square workbook discussed in the Section EG12.1 *In-Depth Excel* instructions and shown in Figure 12.3 on page 472. (Open to the **ChiSquare2x3 _FORMULAS worksheet** to examine all the formulas used in the worksheet.)

For other 2×3 problems, change the **Observed Frequencies** cell counts and row and column labels in rows 4 through 7. For 2×4 problems, use the **ChiSquare2x4 worksheet**. For 2×5 problems, use the **ChiSquare2x5 worksheet**. In either case, enter the contingency table data for the problem in the rows 4 through 7 Observed Frequencies area.

The Marascuilo Procedure

PHStat2 Modify the *PHStat2* instructions for the chi-square test to include the Marascuilo procedure to test for the difference among more than two proportions (see page 514). In step 4, enter a **Title**, check **Marascuilo Procedure**, and then click **OK**.

In-Depth Excel Use the Marascuilo worksheet linked to a particular chi-square 2 × *c* worksheet in the **Chi-Square Wooksheets workbook** to perform the Marascuilo procedure.

For example, Figure 12.7 on page 479 shows the **Marascuilo2x3 worksheet**, which is linked to the **ChiSquare2x3 worksheet**. This Marascuilo worksheet uses values from the ChiSquare2x3 worksheet to compute group sample proportions in cells B7 through B9 (shown in Figure 12.7) and to compute the critical range in rows 13, 14, and 16. In column D, the worksheet uses IF functions in the form **IF(*absolute difference > critical range, display "Significant", display "Not significant"*)** to indicate which pairs of groups are significantly different. (Open to the **Marascuilo2x3_ FORMULAS worksheet** to examine all the formulas used in the worksheet.)

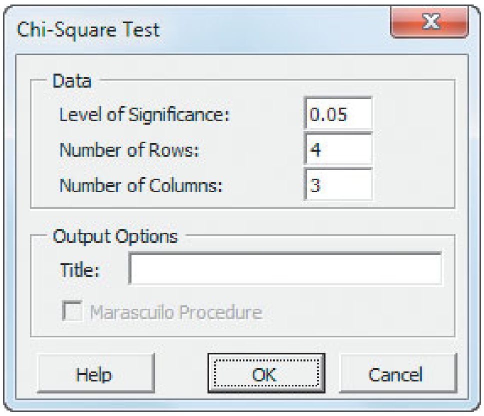
EG12.3 CHI-SQUARE TEST of INDEPENDENCE

PHStat2 Use **Chi-Square Test** to perform the chi-square test of independence. For example, to perform the Figure 12.10 test for the survey data concerning three hotels on page 485, select **PHStat → Multiple-Sample Tests → Chi-Square Test**. In the procedure's dialog box (shown below):

1. Enter **0.05** as the **Level of Significance**.
2. Enter **4** as the **Number of Rows**.
3. Enter **3** as the **Number of Columns**.
4. Enter a **Title** and click **OK**.

In the new worksheet:

5. Read the yellow note about entering values and then press the **Delete** key to delete the note.
6. Enter the Table 12.9 data on page 482, including row and column labels, in rows 4 through 9.



In-Depth Excel Use one of the *r* × *c* worksheets in the **Chi-Square worksheets workbook** to perform the chi-square test of independence. For example, Figure 12.10 on page 485 shows the **ChiSquare4x3 worksheet** that contains the data for Table 12.9 not-returning survey (see page 482). The worksheet computes the expected frequencies and the intermediate results for the chi-square test statistic in much the same way as the COMPUTE worksheet of the Chi-Square workbook discussed in the Section EG12.1 *In-Depth Excel* instructions.

For other 4 × 3 problems, change the **Observed Frequencies** cell counts and row and column labels in rows 4 through 9. For 3 × 4 problems, use the **ChiSquare3x4 worksheet.** For 4 × 3 problems, use the **ChiSquare4x3 worksheet**. For 7 × 3 problems, use the **ChiSquare7x3 worksheet**. For 8 × 3 problems, use the **ChiSquare8x3 worksheet**. In each case, enter the contingency table data for the problem in the Observed Frequencies area.

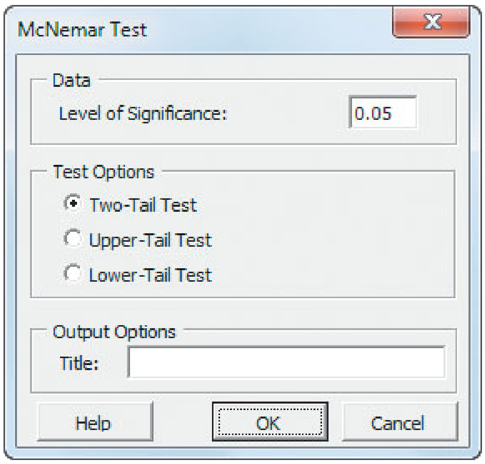
EG12.4 McNEMAR TEST for the DIFFERENCE BETWEEN TWO PROPORTIONS (RELATED SAMPLES)

PHStat2 Use **McNemar Test** to perform the McNemar test. For example, to perform the Figure 12.12 test for the brand loyalty of cell phone providers (see page 489), select **PHStat → Two-Sample Tests (Summarized Data) → McNemar Test**. In the procedure's dialog box (shown below):

1. Enter **0.05** as the **Level of Significance**.
2. Click **Two-Tail Test**.
3. Enter a **Title** and click **OK**.

In the new worksheet:

4. Read the yellow note about entering values and then press the **Delete** key to delete the note.
5. Enter the Table 12.13 data (see page 488), including row and column labels, in rows 4 through 7.



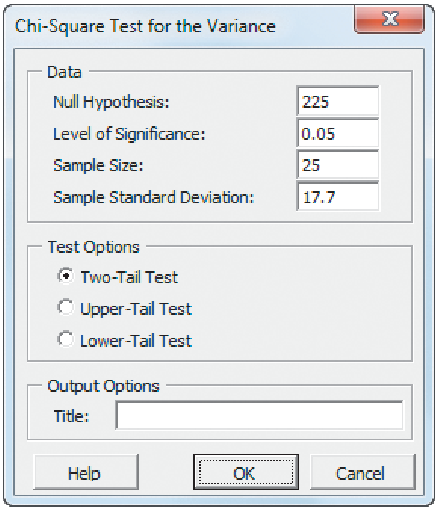
In-Depth Excel Use the **COMPUTE worksheet** of the **McNemar workbook**, shown in Figure 12.12 on page 489, as a template for performing the McNemar test. The worksheet contains the data of Table 12.13 concerning brand loyalty for cell phone providers (see page 488). In cells B20 and B19, respectively, the worksheet uses the expressions **NORMSINV((1 − *level of significance*) / 2)** and **NORMSINV(*level of significance* / 2)** to compute the upper and lower critical values. In cell B21, the expression **2 * (1 − NORMSDIST(*absolute value of the Z test statistic*))** computes the *p*-value.

To perform the McNemar two-tail test for other problems, change the row 4 through 7 entries in the **Observed Frequencies** area and enter the level of significance for the test in cell B11. For one-tail tests, change the Observed Frequencies area and level of significance in the **COMPUTE_ALL worksheet** in the **McNemar workbook**. (Open to the **COMPUTE_ALL_FORMULAS worksheet** to examine the formulas used in the worksheet.)

EG12.5 CHI-SQUARE TEST for the VARIANCE or STANDARD DEVIATION

PHStat2 Use **Chi-Square Test for the Variance** to perform this chi-square test. For example, to perform the test for the Section 12.5 cereal-filling process example, select **PHStat → One-Sample Tests → Chi-Square Test for the Variance**. In the procedure's dialog box (shown below):

1. Enter **225** as the **Null Hypothesis**.
2. Enter **0.05** as the **Level of Significance**.
3. Enter **25** as the **Sample Size.**
4. Enter **17.7** as the **Sample Standard Deviation**.
5. Select **Two-Tail Test**.
6. Enter a **Title** and click **OK**.



The procedure creates a worksheet similar to Figure 12.14 on page 492.

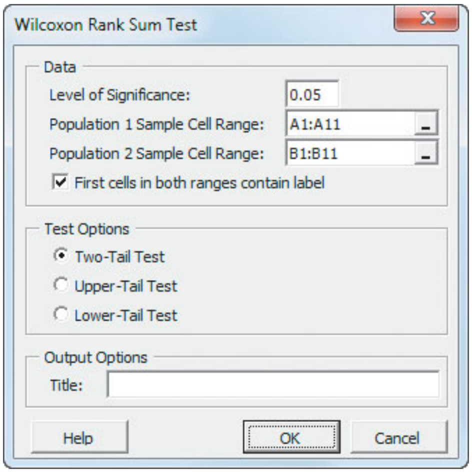
In-Depth Excel Use the **COMPUTE worksheet** of the **Chi-Square Variance workbook**, shown in Figure 12.14 on page 492, as a template for performing the chi-square test. The worksheet contains the data for the cereal-filling process example. In cells B15 and B16, respectively, the worksheet uses the expressions **CHIINV(1 − *half area, degrees of freedom*)** and enter **CHIINV(*half area, degrees of freedom*)** to compute the lower and upper critical values. In B17, the expression **CHIDIST(χ^2 *test statistic, degrees of freedom)*** helps to compute the *p*-value.

To perform the test for other problems, change the null hypothesis, level of significance, sample size, and sample standard deviation in the cell range B4:B7. (Open to the **COMPUTE_ FORMULAS worksheet** to examine the details of all formulas used in the COMPUTE worksheet.)

EG12.6 WILCOXON RANK SUM TEST: NONPARAMETRIC ANALYSIS for TWO INDEPENDENT POPULATIONS

PHStat2 Use **Wilcoxon Rank Sum Test** to perform the Wilcoxon rank sum test. For example, to perform the Figure 12.16 test for the BLK Cola sales data on page 498, open to the **DATA worksheet** of the **Cola workbook**. Select **PHStat → Two-Sample Tests (Unsummarized Data) → Wilcoxon Rank Sum Test**. In the procedure's dialog box (shown below):

1. Enter **0.05** as the **Level of Significance**.
2. Enter **A1:A11** as the **Population 1 Sample Cell Range**.
3. Enter **B1:B11** as the **Population 2 Sample Cell Range**.
4. Check **First cells in both ranges contain label**.
5. Click **Two-Tail Test**.
6. Enter a **Title** and click **OK**.



The procedure creates a worksheet that contains sorted ranks in addition to the worksheet shown in Figure 12.16. Both of these worksheets are discussed in the following *In-Depth Excel* instructions.

In-Depth Excel Use the **COMPUTE worksheet** of the **Wilcoxon workbook**, shown in Figure 12.16 on page 498, as a template for performing the two-tail Wilcoxon rank sum test. The worksheet contains data and formulas to use the unsummarized data for the BLK Cola sales example. In cells B22 and B21, respectively, the worksheet uses **NORMSINV ((1 − *level of significance*) / 2)** and **NORMSINV(*level of significance* / 2)** to compute the upper and lower critical values. In cell B23, **2 * (1 − NORMSDIST(*absolute value of the Z test statistic*))** computes the *p*-value.

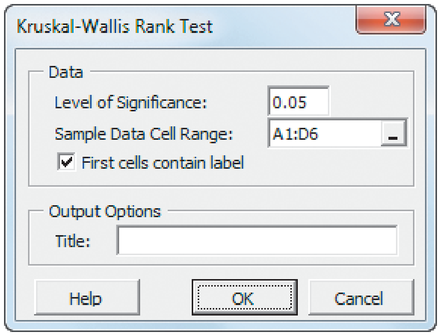
For other problems, use the COMPUTE worksheet with either unsummarized or summarized data. For summarized data, overwrite the formulas that compute the **Sample Size** and **Sum of Ranks** in cells B7, B8, B10, and B11, with the values for these statistics.

For unsummarized data, first open to the **SortedRanks worksheet** and enter the sorted values for both groups in stacked format. Use column A for the sample names and column B for the sorted values. Assign a rank for each value and enter the ranks in column C of the same worksheet. Then open to the COMPUTE worksheet (or the similar COMPUTE_ALL worksheet, if performing a one-tail test) and edit the formulas in cells B7, B8, B10, and B11. Enter **COUNTIF(*cell range for all sample names*, *sample name to be matched*)** functions to compute the sample size for a sample. Enter **SUMIF(*cell range for all sample names*, *sample name to be matched, cell range in which to select cells for summing*)** functions to compute the sum of ranks for a sample. For example, in the current COMPUTE worksheet, the formula **=COUNTIF(SortedRanks!A2:A21, "Normal")** in cell B7 counts the number of occurrences of the sample name **"Normal"** in column A to compute the sample size of the **Population 1 Sample**. The formula **=SUMIF(SortedRanks!A2:A21, "Normal", C2:C21)** in cell B8 computes the sum of ranks for the **Population 1 Sample** by summing the column C ranks for rows in which the column A value is **Normal**.

EG12.7 KRUSKAL-WALLIS RANK TEST: NONPARAMETRIC ANALYSIS for the ONE-WAY ANOVA

PHStat2 Use **Kruskal-Wallis Rank Test** to perform the Kruskal-Wallis rank test. For example, to perform the Figure 12.18 Kruskal-Wallis rank test for differences among the four median tensile strengths of parachutes on page 503, open to the **DATA worksheet** of the **Parachute workbook**. Select **PHStat → Multiple-Sample Tests → Kruskal-Wallis Rank Test**. In the procedure's dialog box (shown in the right column):

1. Enter **0.05** as the **Level of Significance**.
2. Enter **A1:D6** as the **Sample Data Cell Range**.
3. Check **First cells contain label**.
4. Enter a **Title** and click **OK**.



The procedure creates a worksheet that contains sorted ranks in addition to the worksheet shown in Figure 12.18 on page 503. Both of these worksheets are discussed in the following *In-Depth Excel* instructions.

In-Depth Excel Use the **KruskalWallis4 worksheet** of the **Kruskal-Wallis Worksheets workbook**, shown in Figure 12.18 on page 503, as a model for performing the Kruskal-Wallis rank test. The worksheet contains the data and formulas to use the unsummarized data for the Section 12.7 four-supplier parachute example. In cell B13, the worksheet uses **CHIINV(*level of significance*, *number of groups* - 1)** to compute the critical value and, in cell B14, **CHIDIST(*H test statistic, number of groups* - 1)** computes the *p*-value.

For other problems with four groups, use the KruskalWallis4 worksheet with either unsummarized or summarized data. For summarized data, overwrite the formulas that display the group names and compute the **Sample Size** and **Sum of Ranks** in columns D, E, and F with the values for these statistics.

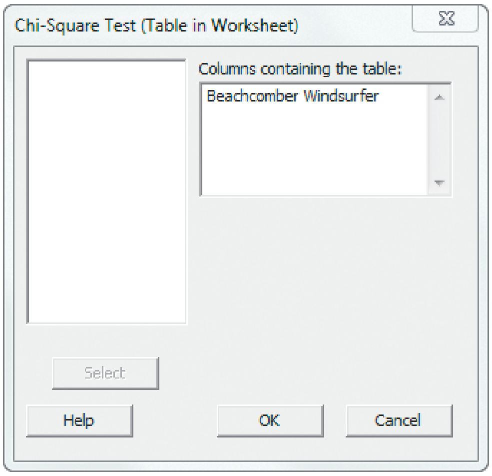
For unsummarized data, first open to the **SortedRanks worksheet** and enter the sorted values for both groups in stacked format. Use column A for the sample names and column B for the sorted values. Assign ranks for each value and enter the ranks in column C of the same worksheet. Also paste your unsummarized stacked data in columns, starting with Column E. (The row 1 cells, starting with cell E1, are used to identify each group.) Then open to the KruskalWallis4 worksheet (or the similar **KruskalWallis3 worksheet**, if using three groups) and edit the formulas in columns E and F. Enter **COUNTIF(*cell range for all group names*, *group name to be matched*)** functions to compute the sample size for a group. Enter **SUMIF(*cell range for all group names*, *group name to be matched, cell range in which to select cells for summing*)** functions to compute the sum of ranks for a group. (Open to the **Kruskal Wallis4_FORMULAS worksheet** to examine all current formulas.)

CHAPTER 12 MINITAB GUIDE

MG12.1 CHI-SQUARE TEST for the DIFFERENCE BETWEEN TWO PROPORTIONS

Use **Chi-Square Test (Two-Way Table in Worksheet)** to perform the chi-square test with summarized data. For example, to perform the Figure 12.3 test for the two-hotel guest satisfaction data on page 472, open to the **Two-Hotel Survey worksheet**. Select **Stat → Tables → Chi-Square Test (Two-Way Table in Worksheet)**. In the Chi-Square Test (Table in Worksheet) dialog box (shown below):

1. Double-click **C2 Beachcomber** in the variables list to add **Beachcomber** to the **Columns containing the table** box.
2. Double-click **C3 Windsurfer** in the variables list to add **Windsurfer** to the **Columns containing the table** box.
3. Click **OK**.



Minitab can also perform a chi-square test for the difference between two proportions using unsummarized data. Use the Section MG2.2 instructions for using **Cross Tabulation and Chi-Square** to create contingency tables (see page 87), replacing step 4 with these steps 4 though 7:

4. Click **Chi-Square**.

In the Cross Tabulation - Chi-Square dialog box:

5. Select **Chi-Square analysis**, **Expected cell counts**, and **Each cell's contribution to the Chi-Square statistic**.
6. Click **OK**.
7. Back in the original dialog box, click **OK**.

MG12.2 CHI-SQUARE TEST for DIFFERENCES AMONG MORE THAN TWO PROPORTIONS

Use **Chi-Square Test (Two-Way Table in Worksheet)** to perform the chi-square test with summarized data. Use modified Section MG2.2 instructions on the page 87 for using **Cross Tabulation and Chi-Square** to perform the chi-square test with unsummarized data. See Section MG12.1 for detailed instructions.

To perform the Figure 12.6 test for the guest satisfaction data concerning three hotels on page 478, open to the **Three-Hotel Survey worksheet**, select **Stat → Tables → Chi-Square Test (Two-Way Table in Worksheet)**, and add the names of columns 2 through 4 to the **Columns containing the table** box.

The Marascuilo Procedure

There are no Minitab Guide instructions for this section.

MG12.3 CHI-SQUARE TEST of INDEPENDENCE

Again, as in Section MG12.2, use either **Chi-Square Test (Two-Way Table in Worksheet)** for summarized data or the modified instructions for using **Cross Tabulation and Chi-Square** for unsummarized data to perform this test.

MG12.4 MCNEMAR TEST for the DIFFERENCE BETWEEN TWO PROPORTIONS (RELATED SAMPLES)

There are no Minitab Guide instructions for this section.

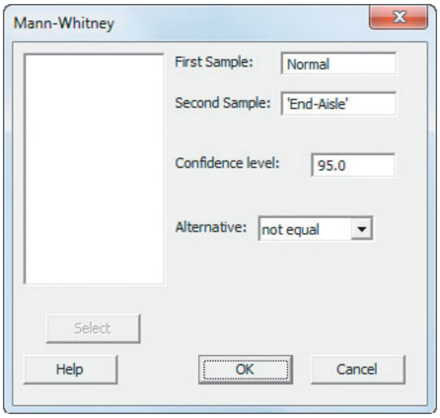
MG12.5 CHI-SQUARE TEST for the VARIANCE or STANDARD DEVIATION

There are no Minitab Guide instructions for this section.

MG12.6 WILCOXON RANK SUM TEST: NONPARAMETRIC ANALYSIS for TWO INDEPENDENT POPULATIONS

Use **Mann-Whitney** to perform a test numerically equivalent to the Wilcoxon rank sum test. For example, to perform the Figure 12.16 test for the BLK Cola sales data on page 498, open to the **Cola worksheet**. Select **Stat ➔ Nonparametrics ➔ Mann-Whitney**. In the Mann-Whitney dialog box (shown below):

1. Double-click **C1 Normal** in the variables list to add **Normal** in the **First Sample** box.
2. Double-click **C2 End-Aisle** in the variables list to add **'EndAisle'** in the **Second Sample** box.
3. Enter **95.0** in the **Confidence level** box.
4. Select **not equal** in the **Alternative** drop-down list.
5. Click **OK**.



MG12.7 KRUSKAL-WALLIS RANK TEST: NONPARAMETRIC ANALYSIS for the ONE-WAY ANOVA

Use **Kruskal-Wallis** to perform the Kruskal-Wallis rank test. For example, to perform the Figure 12.18 Kruskal-Wallis rank test for differences among the four median tensile strengths of parachutes on page 503, open to the **ParachuteStacked worksheet**. Select **Stat ➔ Nonparametrics ➔ Kruskal-Wallis**. In the Kruskal-Wallis dialog box (shown below):

1. Double-click **C2 Strength** in the variables list to add **Strength** in the **Response** box.
2. Double-click **C1 Supplier** in the variables list to add **Supplier** in the **Factor** box.
3. Click **OK**.