CHAPTER 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 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 resort 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. 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 a difference is present, 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 a variety of one-sample tests, and Chapter 10 developed several two-sample tests. Chapter 11 discussed the analysis of variance (ANOVA) that 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, as well as the hypothesis of *independence* in the joint responses to two categorical variables. 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 (INDEPENDENT SAMPLES)

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 2.4) to display the frequency of occurrence of successes and failures 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). In tabulating the responses to the single question "Are you likely to choose this hotel again?" 163 of 227 guests at the Beachcomber responded yes, and 154 of 262 guests at the Windsurfer responded yes. At the 0.05 level of significance, is there 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 has two rows and two columns and 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**
Successes	X_1	X_2	X
Failures	$n_1 - X_1$	$n_2 - X_2$	$n - X$
Totals	n_1	n_2	n

where

$$X_1 = \text{number of successes in group 1}$$

$$X_2 = \text{number of successes in group 2}$$

$$n_1 - X_1 = \text{number of failures in group 1}$$

$$n_2 - X_2 = \text{number of failures in group 2}$$

$$X = X_1 + X_2, \text{ the total number of successes}$$

$$n - X = (n_1 - X_1) + (n_2 - X_2), \text{ the total number of failures}$$

$$n_1 = \text{sample size in group 1}$$

$$n_2 = \text{sample size in group 2}$$

$$n = n_1 + n_2 = \text{total sample size}$$

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 (that is, success) or would not return to the hotel (that is, failure), 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 equality of 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 test statistic, shown in Equation (12.1).

χ^2 TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

The χ^2 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.

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

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

[1] *In general, in a contingency table, the degrees of freedom are equal to the (number of rows −1) times the (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 successes 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 successes for the two groups combined (that is, the total number of successes divided by the total sample size). The complement of $\bar{p}$, $1 - \bar{p}$, represents the estimated overall proportion of failures in the two groups. Using the notation presented in Table 12.1 on page 463, Equation (12.2) defines $\bar{p}$.

COMPUTING THE ESTIMATED OVERALL PROPORTION

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

To compute the expected frequency, f_e, for each cell pertaining to success (that is, 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 each cell pertaining to failure (that is, 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 test statistic shown in Equation (12.1) approximately follows a **chi-square 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 test statistic is greater than χ^2_U, the upper-tail critical value from the χ^2 distribution having 1 degree of freedom. Thus, the decision rule is

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

Figure 12.1 illustrates the decision rule.

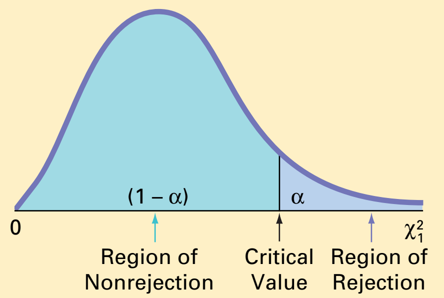


FIGURE 12.1
Regions of rejection and nonrejection when using the chi-square test for the difference between two proportions, with level of significance α

If the null hypothesis is true, the computed χ^2 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 statistic is expected to be large. However, what constitutes 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 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 and the corresponding contingency table displayed in Table 12.2 on page 463. 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 respective 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 463.

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				
CHOOSE HOTEL AGAIN?	Beachcomber Observed	Beachcomber Expected	Windsurfer Observed	Windsurfer 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 test statistic given by Equation (12.1) on page 464. Table 12.4 presents the calculations.

TABLE 12.4
Computation of χ^2 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 distribution** is a right-skewed distribution whose shape depends solely on the number of degrees of freedom. You find the critical value of the χ^2 test statistic from Table E.4, a portion of which is presented as Table 12.5.

TABLE 12.5
Finding the χ^2 Critical Value from the Chi-Square Distribution with 1 Degree of Freedom, Using the 0.05 Level of Significance

	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 statistic is greater than 3.841 (see Figure 12.2). Because $9.05 > 3.841$, you reject H_0. You conclude that there is a difference in the proportion of guests who would return to the Beachcomber and 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

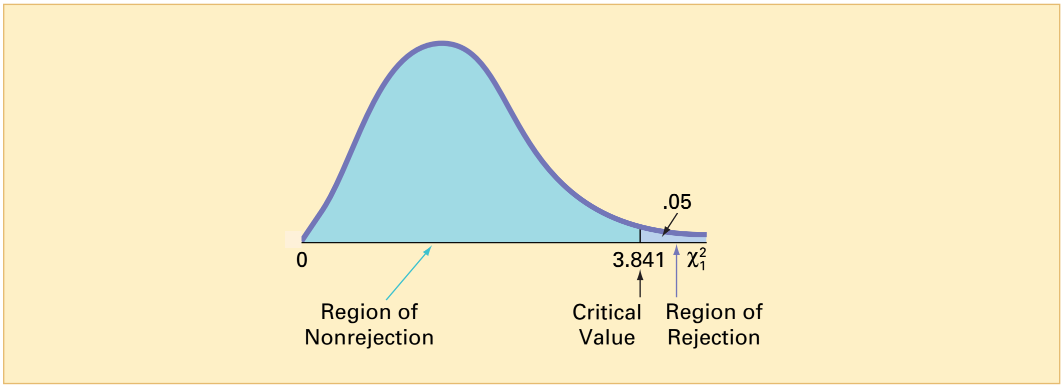


Figure 12.3 represents a Microsoft Excel worksheet for the guest satisfaction contingency table (Table 12.2 on page 463).

FIGURE 12.3

Microsoft Excel worksheet for the hotel guest satisfaction data

See Section E12.1 to create this. Figure E12.1 in that section includes the formulas for the first 15 rows of the worksheet that are the basis for computing the chi-square statistic in cell B25.

	A	B	C	D	
1	**Guest Satisfaction Analysis**				
2					
3	**Observed Frequencies**				
4		**Hotel**			
5	**Choose Again?**	**Beachcomber**	**Windsurfer**	**Total**	
6	**Yes**	**163**	**154**	**317**	
7	**No**	**64**	**108**	**172**	
8	**Total**	**227**	**262**	**489**	
9					
10	Expected Frequencies				
11		Hotel			
12	Choose Again?	Beachcomber	Windsurfer	Total	
13	Yes	147.1554	169.8446	317	
14	No	79.8446	92.1554	172	
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			
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", "Do not reject the null hypothesis")**
28					
29	***Expected frequency assumption***				
30	***is met.***				**=IF(OR(B13 < 5, C13 < 5, B14 < 5, C14 < 5), " is violated.", " is met.")**

This worksheet includes the expected frequencies, χ^2 test statistic, degrees of freedom, and p-value. The χ^2 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 of 0.0026 is the probability of observing sample proportions as different as or more different than the actual difference ($0.718 - 0.588 = 0.13$ 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 the guest is likely to return to the hotel again. An examination of Table 12.3 on page 466 indicates that a greater proportion of guests at the Beachcomber are likely to return than at the Windsurfer.

For the χ^2 test to give accurate results for a 2×2 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 test statistic is the square of the Z test statistic. For instance, in the guest satisfaction study, the computed Z test statistic is +3.0088 and the computed χ^2 test statistic is 9.0526. Except for rounding error, this latter value is the square of +3.0088 [that is, $(+3.0088)^2 \cong 9.0526$]. Also, if you compare the critical values of the test statistics from the two distributions, at the 0.05 level of significance, the χ_1^2 value of 3.841 is the square of the Z value of ±1.96 (that is, $\chi_1^2 = Z^2$). Furthermore, the p-values for both tests are equal. Therefore, when testing the null hypothesis of equality of proportions:

$$H_0: \pi_1 = \pi_2$$

against the alternative that the population proportions are not equal:

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

the Z test and the χ^2 test are equivalent methods. However, 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. $\alpha = 0.01$, $n = 16$
b. $\alpha = 0.025$, $n = 11$
c. $\alpha = 0.05$, $n = 8$

12.2 Determine the critical value of χ^2 with 1 degree of freedom in each of the following circumstances:
a. $\alpha = 0.05$, $n = 28$
b. $\alpha = 0.025$, $n = 21$
c. $\alpha = 0.01$, $n = 5$

PH Grade ASSIST **12.3** Use the following contingency table:

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

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

PH Grade ASSIST **12.4** Use the following contingency table:

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

a. Find the expected frequency for each cell.
b. Find the χ^2 statistic for this contingency table. 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:

	GENDER		
ENJOY SHOPPING FOR CLOTHING	**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 10.33 (a) through (c) on page 395.

12.6 Is good gas mileage a priority for car shoppers? A survey conducted by Progressive Insurance asked this question of both men and women shopping for new cars. The data were reported as percentages, and no sample sizes were given:

	GENDER	
GAS MILEAGE A PRIORITY?	**Men**	**Women**
Yes	76%	84%
No	24%	16%

Source: Extracted from "Snapshots," **usatoday.com**, *June 21, 2004.*

a. Assume that 50 men and 50 women were included in the survey. At the 0.05 level of significance, is there a difference between males and females in the proportion who make gas mileage a priority?
b. Assume that 500 men and 500 women were included in the survey. At the 0.05 level of significance, is there a difference between males and females in the proportion who make gas mileage a priority?
c. Discuss the effect of sample size on the chi-square test.

12.7 The results of a yield improvement study at a semiconductor manufacturing facility provided defect data for a sample of 450 wafers. The following contingency table presents a summary of the responses to two questions: "Was a particle found on the die that produced the wafer?" and "Is the wafer good or bad?"

	QUALITY OF WAFER		
PARTICLES	**Good**	**Bad**	**Totals**
Yes	14	36	50
No	320	80	400
Totals	334	116	450

Source: Extracted from S. W. Hall, "Analysis of Defectivity of Semiconductor Wafers by Contingency Table," Proceedings Institute of Environmental Sciences, *Vol. 1 (1994), pp. 177–183.*

a. At the 0.05 level of significance, is there a difference between the proportion of good and bad wafers that have particles?
b. Determine the p-value in (a) and interpret its meaning.
c. What conclusions can you draw from this analysis?
d. Compare the results of (a) and (b) to those of Problem 10.35 on page 395.

SELF Test **12.8** According to an Ipsos poll, the perception of unfairness in the U.S. tax code is spread fairly evenly across income groups, age groups, and education levels. In an April 2006 survey of 1,005 adults, Ipsos reported that almost 60% of all people said the code is unfair, while slightly more that 60% of those making more than \$50,000 viewed the code as unfair (Extracted from "People Cry Unfairness," *The Cincinnati Enquirer*, April 16, 2006, p. A8). Suppose that the following contingency table represents the specific breakdown of responses:

	INCOME LEVEL		
U.S. TAX CODE	**Less Than \$50,000**	**More Than \$50,000**	**Total**
Fair	225	180	405
Unfair	280	320	600
Total	505	500	1,005

a. At the 0.05 level of significance, is there evidence of a difference in the proportion of adults who think the U.S. tax code is unfair between the two income groups?
b. Determine the p-value in (a) and interpret its meaning.
c. Compare the results of (a) and (b) to those of Problem 10.36 on page 396.

12.9 Where people turn to for news is different for various age groups. Suppose that a study conducted on this issue (extracted from P. Johnson, "Young People Turn to the Web for News," *USA Today*, March 23, 2006, p. 9D) was based on 200 respondents who were between the ages of 36 and 50 and 200 respondents who were above age 50. Of the 200 respondents who were between the ages of 36 and 50, 82 got their news primarily from newspapers. Of the 200 respondents who were above age 50, 104 got their news primarily from newspapers.
a. Construct a 2×2 contingency table.
b. Is there evidence of a significant difference in the proportion who get their news primarily from newspapers between those 36 to 50 years old and those above 50 years old? (Use $\alpha = 0.05$.)
c. Determine the p-value in (a) and interpret its meaning.
d. Compare the results of (b) and (c) to those of Problem 10.39 on page 396.

12.10 An experiment was conducted to study the choices made in mutual fund selection. Undergraduate and MBA students were presented with different S&P 500 index funds that were identical except for fees. Suppose 100 undergraduate students and 100 MBA students were selected. Partial results are shown on page 470:

	STUDENT GROUP	
FUND	**Undergraduate**	**MBA**
Highest-cost fund	27	18
Not highest-cost fund	73	82

Source: Extracted from J. Choi, D. Laibson, and B. Madrian, "Why Does the Law of One Practice Fail? An Experiment on Mutual Funds," **www.som.yale.edu/faculty/jjc83/fees.pdf**.

a. At the 0.05 level of significance, is there evidence of a difference between undergraduate and MBA students in the proportion who selected the highest-cost fund?
b. Determine the p-value in (a) and interpret its meaning.
c. Compare the results of (a) and (b) to those of Problem 10.38 on page 396.

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 464:

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

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 464, 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

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

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 464 approximately follows a chi-square distribution, with degrees of freedom equal to the number of rows in the contingency table minus 1, times the number of columns in the table minus 1. For a **$2 \times c$ contingency table**, there are $c - 1$ degrees of freedom:

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

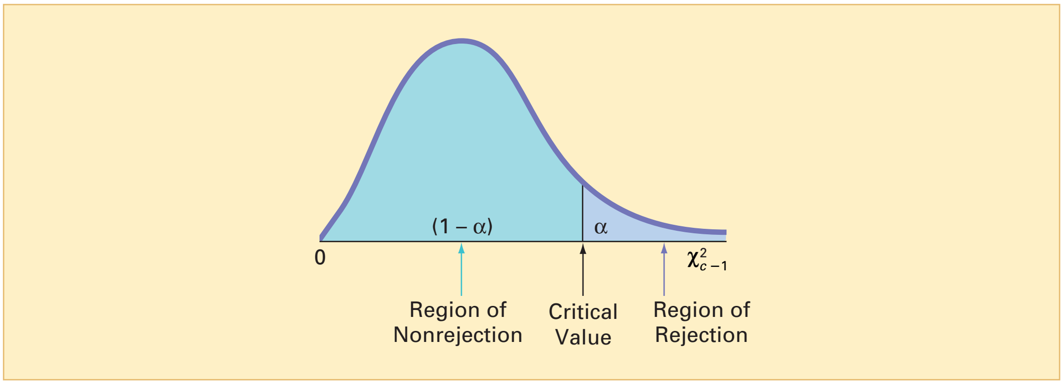
Using the level of significance α, you reject the null hypothesis if the computed χ^2 test statistic is greater than χ^2_U, the upper-tail critical value from a chi-square distribution having $c - 1$ degrees of freedom. Therefore, the decision rule is

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

Figure 12.4 illustrates the 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 concerning T.C. Resort Properties. A similar survey was recently conducted on a different island on which T.C. Resort Properties has three different hotels. Table 12.6 presents the responses to a question concerning whether guests would be likely to choose this hotel again.

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 taken at 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 on page 471 and the expected frequencies from Table 12.7 to compute the χ^2 test statistic (given by Equation (12.1) on page 464) displayed in Table 12.8.

TABLE 12.8
Computation of χ^2 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 three hotels are evaluated, 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. Because the computed test statistic ($\chi^2 = 40.23$) is greater than this critical value, you reject the null hypothesis (see Figure 12.5). Microsoft Excel (see Figure 12.6) also reports the *p*-value. Likewise, because the *p*-value is approximately 0.0000, which is less than $\alpha = 0.05$, you reject the null hypothesis. Further, this *p*-value indicates that there is virtually no chance to see 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 12.5
Regions of rejection and nonrejection when testing for differences in three proportions at the 0.05 level of significance, with 2 degrees of freedom

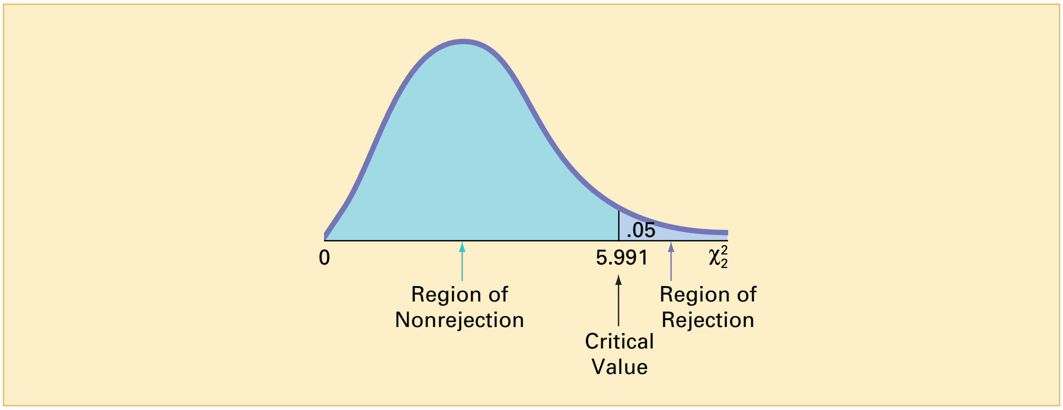


FIGURE 12.6
Microsoft Excel worksheet for the guest satisfaction data of Table 12.6

See Section E12.2 to create this.

	A	B	C	D	E	
1	**Guest Satisfaction (3-Hotels) Analysis**					
2						
3		**Observed Frequencies**				
4		**Hotel**				
5	**Choose Again?**	**Golden Palm**	**Palm Royale**	**Palm Princess**	**Total**	
6	**Yes**	**128**	**199**	**186**	**513**	
7	**No**	**88**	**33**	**66**	**187**	
8	**Total**	**216**	**232**	**252**	**700**	
9						
10		Expected Frequencies				
11		Hotel				
12	Choose Again?	Golden Palm	Palm Royale	Palm Princess	Total	
13	Yes	158.2971	170.0229	184.68	513	
14	No	57.7029	61.9771	67.32	187	
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", "Do not reject the null hypothesis")			
28						
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.")			

For the χ^2 test to give accurate results when dealing with $2 \times c$ contingency tables, all expected frequencies must be large. There is much debate among statisticians about the definition of *large*. Some statisticians (see reference 5) have found that the test gives accurate results as long as all expected frequencies equal or exceed 0.5. Other statisticians, more conservative in their approach, require that no more than 20% of the cells contain expected frequencies less than 5 and no cells have expected frequencies less than 1 (see reference 3). A reasonable compromise between these points of view is to make sure that each expected frequency is at least 1. To accomplish this, you may need to collapse two or more low-expected-frequency categories into one category in the contingency table before performing the test. Such merging of categories usually results in expected frequencies sufficiently large to conduct the χ^2 test accurately. If combining categories is undesirable, alternative procedures are available (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. But *which* of the proportions differ? Because the result of the χ^2 test for equality of proportions does not specifically answer this question, you need to use a multiple comparisons procedure such as the Marascuilo procedure.

The **Marascuilo procedure** enables you to make comparisons between all pairs of groups. First, compute the observed differences $p_j - p_{j'}$ (where $j \neq j'$) among all $c(c-1)/2$ pairs. Then, use Equation (12.4) to compute the corresponding critical ranges for the Marascuilo procedure.

CRITICAL RANGE FOR THE MARASCUILO PROCEDURE

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

You need to compute a different critical range for each pairwise comparison of sample proportions. In the final step, 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. Because there are three hotels, there are $(3)(3-1)/2 = 3$ pairwise comparisons. From Table 12.6 on page 471, the three sample proportions are

$$p_1 = \frac{X_1}{n_1} = \frac{128}{216} = 0.593$$

$$p_2 = \frac{X_2}{n_2} = \frac{199}{232} = 0.858$$

$$p_3 = \frac{X_3}{n_3} = \frac{186}{252} = 0.738$$

Using Table E.4 and an overall level of significance of 0.05, the upper-tail critical value of the χ^2 test statistic for a chi-square distribution having $(c-1) = 2$ degrees of freedom is 5.991. Thus,

$$\sqrt{\chi^2_U} = \sqrt{5.991} = 2.448$$

Next, you compute the three pairs of absolute differences in sample proportions and their corresponding critical ranges. If the absolute difference is greater than its critical range, the proportions are significantly different:

Absolute Difference in Proportions	Critical Range
$\lvert p_j - p_{j'} \rvert$	$2.448\sqrt{\dfrac{p_j(1-p_j)}{n_j} + \dfrac{p_{j'}(1-p_{j'})}{n_{j'}}}$
$\lvert p_1 - p_2 \rvert = \lvert 0.593 - 0.858 \rvert = 0.265$	$2.448\sqrt{\dfrac{(0.593)(0.407)}{216} + \dfrac{(0.858)(0.142)}{232}} = 0.0992$
$\lvert p_1 - p_3 \rvert = \lvert 0.593 - 0.738 \rvert = 0.145$	$2.448\sqrt{\dfrac{(0.593)(0.407)}{216} + \dfrac{(0.738)(0.262)}{252}} = 0.1063$
$\lvert p_2 - p_3 \rvert = \lvert 0.858 - 0.738 \rvert = 0.120$	$2.448\sqrt{\dfrac{(0.858)(0.142)}{232} + \dfrac{(0.738)(0.262)}{252}} = 0.0880$

These computations are shown in worksheet format in Figure 12.7.

FIGURE 12.7

Microsoft Excel Marascuilo procedure worksheet

See the "Using the Marascuilo Worksheets" part of Section E12.2 to learn more about this worksheet.

	A	B	C	D
1	**Marascuilo Procedure**			
2	**Guest Satisfaction (3-Hotels) Analysis**			
3	Level of Significance	0.05		
4	Square Root of Critical Value	2.4477		
5				
6	Sample Proportions			
7	Group 1	0.5926		
8	Group 2	0.8578		
9	Group 3	0.7381		
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**

With 95% confidence, 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 management should study the reasons for these differences and, in particular, should try to determine why satisfaction is significantly lower at the Golden Palm than at the other two hotels.

PROBLEMS FOR SECTION 12.2

Learning the Basics

PH Grade ASSIST **12.11** Consider a contingency table with two rows and five columns.

a. Find the degrees of freedom.
b. Find the critical value for $\alpha = 0.05$.
c. Find the critical value for $\alpha = 0.01$.

PH Grade ASSIST **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 the χ^2 statistic for this contingency table. Is it significant at $\alpha = 0.05$?
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which groups are different.

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 the χ^2 statistic for this contingency table. 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 A survey was conducted in five countries. The percentages of respondents who said that they eat out once a week or more are as follows:

Germany	10%
France	12%
United Kingdom	28%
Greece	39%
United States	57%

Source: Adapted from M. Kissel, "Americans Are Keen on Cocooning," The Wall Street Journal*, July 22, 2003, p. D3.*

Suppose that the survey was based on 1,000 respondents in each country.

a. At the 0.05 level of significance, determine whether there is a significant difference in the proportion of people who eat out at least once a week in the various countries.
b. Find the p-value in (a) and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which countries are different. Discuss your results.

12.15 Is the degree to which students withdraw from introductory business statistics courses the same for online courses and traditional courses taught in a classroom? Professor Constance McLaren at Indiana State University collected data for five semesters to investigate this question. The following table cross-classifies introductory business statistics students by the type of course (classroom or online) and student persistence (active, dropped, or vanished):

	STUDENT PERSISTENCE		
TYPE OF COURSE	**Active**	**Dropped**	**Vanished**
Classroom	127	8	4
Online	81	51	20

Source: Extracted from C. McLaren, "A Comparison of Student Persistence and Performance in Online and Classroom Business Statistics Experiences," Decision Sciences Journal of Innovative Education*, Spring 2004, 2(1), pp. 1–10. Published by the Decision Sciences Institute, headquartered at Georgia State University, Atlanta, GA.*

a. Is there evidence of a difference in student persistence (active, dropped, or vanished) based on type of course? (Use $\alpha = 0.05$.)
b. Compute the p-value and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which 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 the day of the week a person does the majority of grocery shopping dependent on age? A study cross-classified grocery shoppers by age and major shopping day ("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%

Source: Extracted from "Major Shopping by Day," Progressive Grocer Annual Report*, April 30, 2002.*

Assume that 200 shoppers for each age category 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. Discuss your results.
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) through (b) of Problem 12.16, assuming that only 50 shoppers for each age category were surveyed. Discuss the implications of sample size on the χ^2 test for differences among more than two populations.

12.18 An experiment was conducted by James Choi, David Laibson, and Brigitte Madrian to study the choices made in fund selection. When presented with four S&P 500 index funds that were identical except for their fees, undergraduate and MBA students chose the funds as follows (in percentages):

	FUND			
STUDENT GROUP	**Lowest Cost**	**Second-Lowest Cost**	**Third-Lowest Cost**	**Highest Cost**
Undergraduates	19	37	17	27
MBA	19	40	23	18

Source: Extracted from J. Choi, D. Laibson, and B. Madrian, "Why Does the Law of One Practice? An Experiment in Mutual Funds," **www.som.yale.edu/faculty/jjc83/fees.pdf**.

a. Determine whether there is a difference in the fund selection (lowest cost, second-lowest cost, third-lowest cost, highest cost) based on the student group. (Use $\alpha = 0.05$.)
b. Determine the p-value and interpret its meaning.
c. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which groups are different.

12.19 An article (Extracted from P. Kitchen, "Retirement Plan: To Keep Working," *Newsday*, September 24, 2003) discussed the results of a sample of 2,001 Americans ages 50 to 70 who were employed full time or part time. The results were as follows:

	PLANS					
GENDER	Not Work for Pay	Start Own Business	Work Full Time	Work Part Time	Don't Know	Other
Male	257	115	103	457	27	42
Female	359	87	49	436	34	35

a. Is there evidence of a significant difference among the plans for retirement with respect to gender? (Use $\alpha = 0.05$.)
b. Determine the p-value in (a) and interpret its meaning.

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 (that is, there is no relationship between them).
H_1: The two categorical variables are dependent (that is, there is a relationship between them).

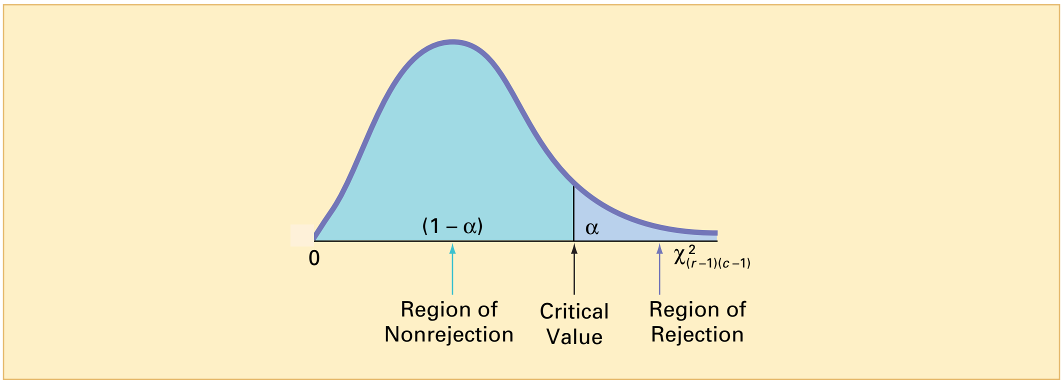
Once again, you use Equation (12.1) on page 464 to compute the test statistic:

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

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

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

FIGURE 12.8
Regions of rejection and nonrejection when testing for independence in an $r \times c$ contingency table, using the χ^2 test



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 stated 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 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. Nevertheless, there is a fundamental difference between the two types of tests. The major difference is 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 sample group or level are classified into two categories, such as *success* and *failure*. The objective is to make comparisons and evaluate differences between the proportions of *success* 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, a second question was asked of all respondents who indicated that they were not likely to return. These guests were asked to indicate the primary reason for their response. Table 12.9 presents the resulting 4×3 contingency table.

TABLE 12.9

Contingency Table of Primary Reason for Not Returning and Hotel

	HOTEL			
PRIMARY REASON FOR NOT RETURNING	**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

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 471, there were 88 guests in the Golden Palm, 33 guests in the Palm Royale, and 66 guests in the Palm Princess who were not planning 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.

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 464 to compute the test statistic:

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

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 were true

To compute the expected frequency, f_e, in any cell, use the multiplication rule for independent events discussed on page 163 [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(Price) and P(Golden Palm). Here, the proportion of reasons that are due to price, P(Price), is 67/187 = 0.3583, and the proportion of all responses from the Golden Palm, P(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

row total = sum of all the frequencies in the row

column total = sum of all the frequencies in the column

n = overall sample size

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

	HOTEL			
PRIMARY REASON FOR NOT RETURNING	**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 test statistic shown in Equation (12.1) on page 464. Here, the test statistic approximately follows a chi-square distribution, with degrees of freedom equal to the number of rows in the contingency table minus 1, times the number of columns in the table minus 1:

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

Table 12.11 illustrates the computations for the χ^2 test statistic.

TABLE 12.11
Computation of χ^2 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 level of significance $\alpha = 0.05$, the upper-tail critical value from the chi-square distribution with 6 degrees of freedom is 12.592 (see Table E.4). Because the computed test statistic $\chi^2 = 27.41 > 12.592$, you reject the null hypothesis of independence (see Figure 12.9). Similarly, you can use the Microsoft Excel worksheet in Figure 12.10. Because the p-value $= 0.0001 < 0.05$, you reject the null hypothesis of independence. This p-value indicates that there is virtually no chance of having a relationship this large or larger between hotels and 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.

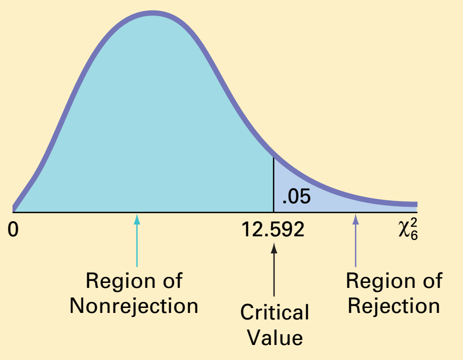


FIGURE 12.9
Regions of rejection and nonrejection when testing for independence in the hotel guest satisfaction survey example at the 0.05 level of significance, with 6 degrees of freedom

FIGURE 12.10

Microsoft Excel worksheet for the 4 × 3 contingency table for primary reason for not returning and hotel

See Section E12.3 to create this. Formulas not shown in Figure 12.10 (in rows 3 through 19) are similar to formulas shown in Figure E12.1.

	A	B	C	D	E	
1	**Cross-Classification Hotel Analysis**					
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.")

Examination of the observed and expected frequencies (see Table 12.11) reveals that price is underrepresented as a reason for not returning to the Golden Palm (that is, $f_o = 23$ and $f_e = 31.53$) but is overrepresented at the Palm Princess. Guests are more satisfied with the price at the Golden Palm compared to 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 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 on page 471, all expected frequencies should be at least 1. For cases in which one or more expected frequencies are less than 1, you can use the test after collapsing two or more low-frequency rows into one row (or collapsing two or more low-frequency columns into one column). Merging of rows or columns usually results in expected frequencies sufficiently large to conduct the χ^2 test accurately.

PROBLEMS FOR SECTION 12.3

Learning the Basics

PH Grade ASSIST

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

PH Grade ASSIST

12.21 When performing a χ^2 test for independence in a contingency table with r rows and c columns, determine the upper-tail critical value of the χ^2 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 During the Vietnam War, a lottery system was instituted to choose males to be drafted into the military. Numbers representing days of the year were "randomly" selected; men born on days of the year with low numbers were drafted first; those with high numbers were not drafted. The table on page 452 shows how many low

(1–122), medium (123–244), and high (245–366) numbers were drawn for birth dates in each quarter of the year:

	QUARTER OF YEAR				
NUMBER SET	**Jan.–Mar.**	**Apr.–Jun.**	**Jul.–Sep.**	**Oct.–Dec.**	**Total**
Low	21	28	35	38	122
Medium	34	22	29	37	122
High	36	41	28	17	122
Total	91	91	92	92	366

a. Is there evidence that the numbers selected were significantly related to the time of year? (Use $\alpha = 0.05$.)
b. Would you conclude that the lottery drawing appears to have been random?
c. What are your answers to (a) and (b) if the frequencies are

23	30	32	37
27	30	34	31
41	31	26	24

12.23 *USA Today* reported on preferred types of office communication by different age groups ("Talking Face to Face vs. Group Meetings," *USA Today*, October 13, 2003, p. A1). Suppose the results were based on a survey of 500 respondents in each age group. The results are cross-classified in the following table:

	TYPE OF COMMUNICATION PREFERRED				
AGE GROUP	**Group Meetings**	**Face-to-face Meetings with Individuals**	**Emails**	**Other**	**Total**
Generation Y	180	260	50	10	500
Generation X	210	190	65	35	500
Boomer	205	195	65	35	500
Mature	200	195	50	55	500
Total	795	840	230	135	2,000

Source: Extracted from "Talking Face to Face vs. Group Meetings," USA Today*, October 13, 2003, p. A1.*

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

PH Grade ASSIST 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 assembly-line workers reveals the following:

	STRESS LEVEL			
COMMUTING TIME	**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 to for news is different for various age groups. A study indicated where different age groups primarily get their news:

	AGE GROUP		
MEDIA	**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.

PH Grade ASSIST **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 were cross-classified in the following table:

	TYPE OF HOUSEHOLD		
WHEN DECISION MADE	**One Adult/ No Children**	**Adult/ 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: Extracted from "What's for Dinner," USA Today*, 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 test for the difference between two proportions. These tests require that the samples are independent from one another. However, sometimes when you are testing differences between two proportions, the data are from repeated measurements or matched samples, and therefore the samples are related. Such situations arise often in marketing 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 of two related samples, you can use the **McNemar test**. If you are doing a two-tail test, you could use a test statistic that follows a chi-square distribution or one that approximately follows the normal distribution. However, if you are carrying out a one-tail test, you need to use the test statistic that approximately follows the normal distribution.

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

$$\pi_1 = \text{proportion in the population who would answer yes to condition 1}$$

$$\pi_2 = \text{proportion in the population who would answer yes to condition 2}$$

Equation (12.6) presents the McNemar test statistic used to test H_0: $\pi_1 = \pi_2$.

McNEMAR TEST

$$Z = \frac{B - C}{\sqrt{B + C}} \quad (12.6)$$

where the test statistic Z is approximately normally distributed.

To illustrate the McNemar test, suppose that a consumer panel of $n = 600$ participants is selected for a marketing study and the panel members are initially asked to state their preferences for two competing cell phone providers, Sprint and Verizon. Suppose that, initially, 282 panelists say they prefer Sprint and 318 say they prefer Verizon. After exposing the entire panel to an intensive marketing campaign strategy for Verizon, suppose the same 600 panelists are again asked to state their preferences, with the following results: Of the 282 panelists who previously preferred Sprint, 246 maintain their brand loyalty, but 36 switch to Verizon. Of the 318 panelists who initially preferred Verizon, 306 remain brand loyal, but 12 switch to Sprint. The results are displayed Table 12.13.

TABLE 12.13
Brand Loyalty for Cell Phone Providers

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

You use the McNemar test for these data because you have repeated measurements from the same set of panelists. Each panelist 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.

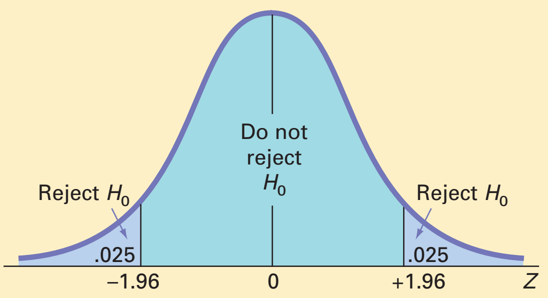
To determine whether the intensive marketing campaign was effective, you want to investigate whether there is a difference between the population proportion who favor Verizon before the campaign, π_1, versus the proportion who favor Verizon after the campaign, π_2. The null and alternative hypotheses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

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

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 $Z = 3.4641 > 1.96$, 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 proportion who preferred Verizon before the intensive marketing campaign is different from the proportion who prefer Verizon after exposure to the intensive marketing campaign. In fact, from Table 12.13, observe that more panelists actually preferred Verizon over Sprint after exposure to the intensive marketing campaign.

FIGURE 12.12
Microsoft Excel results for the McNemar test for brand loyalty of cell phone providers

See Section E12.4 to create this.

	A	B	C	D
1	McNemar Test			
2				
3	Observed Frequencies			
4		Column variable		
5	Row variable	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 2		
GROUP 1	**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

12.28 A market researcher wanted to determine whether the proportion of coffee drinkers who preferred Brand A increased as the 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 favor 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 COMPLETION OF 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 prefer Brand *A* is 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 effects of recent implementation of Six Sigma management on customer satisfaction. A random sample of 100 patients is selected from a list of thousands 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	≥ 5 Days Absent	Total
<5 days absent	32	4	36
≥ 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 5 days was lower in year 1 than in year 2?
b. Compute the *p*-value in (a) and interpret its meaning.

12.5 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 the Wilcoxon rank sum test that does not depend on the assumption of normality for the two populations.
- Use the pooled-variance t test, following some *normalizing transformation* on the data (see reference 9).

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 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 when dealing with studies 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 the average of the ranks that otherwise would have been assigned had there been no ties.

For convenience, 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 samples, either sample may be selected for determining T_1.) For any integer value n, the sum of the first n consecutive integers is $n(n + 1)/2$. Therefore, the test statistic 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 (12.7) to check the accuracy of your rankings.

CHECKING THE RANKINGS

$$T_1 + T_2 = \frac{n(n+1)}{2} \qquad \textbf{(12.7)}$$

The Wilcoxon rank sum test can be either a two-tail test or a one-tail test, depending on whether you are testing whether the two population medians are *different* or whether one median is *greater than* the other median:

Two-Tail Test	One-Tail Test	One-Tail Test
$H_0: M_1 = M_2$ $H_1: M_1 \neq M_2$	$H_0: M_1 \geq M_2$ $H_1: M_1 < M_2$	$H_0: M_1 \leq M_2$ $H_1: M_1 > M_2$

where

$$M_1 = \text{median of population 1}$$

$$M_2 = \text{median of population 2}$$

When both samples n_1 and n_2 are ≤ 10, you use Table E.8 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 12.13) if the computed value of T_1 equals or is greater than 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 $H_1: M_1 < M_2$, 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 12.13). 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 12.13).

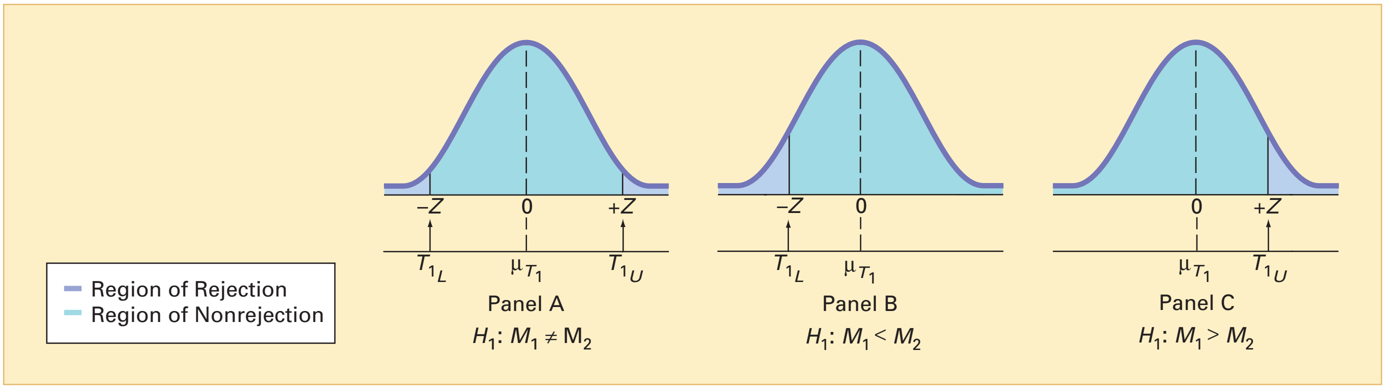


FIGURE 12.13 Regions of rejection and nonrejection using the Wilcoxon rank sum test

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 (12.8) defines the standardized Z test statistic.

LARGE SAMPLE WILCOXON RANK SUM TEST

$$Z = \frac{T_1 - \dfrac{n_1(n+1)}{2}}{\sqrt{\dfrac{n_1 n_2(n+1)}{12}}} \tag{12.8}$$

where the test statistic Z approximately follows a standardized normal distribution.

You use Equation (12.8) for testing the null hypothesis when the sample sizes are outside the range of Table E.8. Based on α, the level of significance selected, you reject the null hypothesis if the computed Z value 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 two locations: normal shelf display and end-aisle location (see page 370). If you do not think that the populations are normally distributed, you can use the Wilcoxon rank sum test for evaluating possible differences in the median sales for the two display locations.[1] The data (stored in the file cola.xls) and the combined ranks are shown in Table 12.14.

[1] 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 location (that is, the medians).

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

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

To perform the Wilcoxon rank sum test, you compute the rankings for the sales from the $n_1 = 10$ stores with a normal shelf display and the $n_2 = 10$ stores with an end-aisle display. Table 12.14 provides the combined rankings.

TABLE 12.14
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 372.

The next step is 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 identify 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 (12.7) on page 487 to show that the sum of the first $n = 20$ integers in the combined ranking is equal to $T_1 + T_2$:

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

To test the null hypothesis that there is no difference between the median sales of the two populations, you use Table E.8 to determine the lower- and upper-tail critical values for the test statistic T_1. From Table 12.15, a portion of Table E.8, observe that for a level of significance of 0.05, the critical values are 78 and 132. The decision rule is

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

TABLE 12.15
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**	**8** (Lower, Upper)	**9**	$\boxed{10}$
	.05	.10	16,40	24,51	33,63	43,76	54,90	66,105	
9	.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	
	.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.8.

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 displays. Because the sum of the ranks is higher for the end-aisle display, you conclude that median sales are higher for the end-aisle display. From the Microsoft Excel worksheet in Figure 12.14, observe that the p-value is 0.0126, which is less than $\alpha = 0.05$. The 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.

FIGURE 12.14

Microsoft Excel Wilcoxon rank sum test worksheet for the BLK cola sales example

See Section E12.5 to create this.

	A	B	
1	**Display Location Analysis**		
2			
3	**Data**		
4	**Level of Significance**	**0.05**	
5			
6	**Population 1 Sample**		
7	**Sample Size**	**10**	
8	**Sum of Ranks**	**72**	
9	**Population 2 Sample**		
10	**Sample Size**	**10**	
11	**Sum of Ranks**	**138**	
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")**

Table E.8 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 (12.8) on page 488]. 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 (12.8),

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

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

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

Because $Z = -2.4946 < -1.96$, the critical value of Z at the 0.05 level of significance, you reject H_0.

PROBLEMS FOR SECTION 12.5

Learning the Basics

PH Grade ASSIST **12.33** Using Table E.8, 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?

12.34 Using Table E.8, 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. $\alpha = 0.005$, $n_1 = 6$, $n_2 = 8$

12.35 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: $M_1 = M_2$?

PH Grade ASSIST **12.36** In Problem 12.35, what are the lower- and upper-tail critical values for the test statistic T_1 from Table E.8 if you use a 0.05 level of significance and the alternative hypothesis is H_1: $M_1 \neq M_2$?

12.37 In Problems 12.35 and 12.36, what is your statistical decision?

12.38 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 (12.7) on page 487 to demonstrate that

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

PH Grade ASSIST **12.39** From Problem 12.38, 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: $M_1 < M_2$.

12.40 In Problems 12.38 and 12.39, what is your statistical decision?

Applying the Concepts

12.41 A vice president for marketing recruits 20 college graduates for management training. The 20 individuals are 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.xls**):

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 $\alpha = 0.05$.)

12.42 Wine experts Gaiter and Brecher use a six-category scale when rating wines: Yech, OK, Good, Very Good, Delicious, and Delicious! (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 the **cabernet.xls** file, 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 $\alpha = 0.05$.)

12.43 In intaglio printing, a design or figure is carved beneath the surface of hard metal or stone. Suppose that an experiment is designed to compare differences in surface hardness of steel plates used in intaglio printing (measured in indentation numbers), based on two different surface conditions—untreated and treated by lightly polishing with emery paper. In the experiment, 40 steel plates are randomly assigned—20 that are untreated, and 20 that are treated. The data are shown in the following table and are stored in the file **intaglio.xls**:

Untreated		Treated	
164.368	177.135	158.239	150.226
159.018	163.903	138.216	155.620
153.871	167.802	168.006	151.233
165.096	160.818	149.654	158.653
157.184	167.433	145.456	151.204
154.496	163.538	168.178	150.869
160.920	164.525	154.321	161.657
164.917	171.230	162.763	157.016
169.091	174.964	161.020	156.670
175.276	166.311	167.706	147.920

a. Is there evidence of a difference in the median surface hardness between untreated and treated steel plates? (Use $\alpha = 0.05$.)
b. What assumptions must you make in (a)?
c. Compare the results of (a) with those of Problem 10.18 (a) on page 380.

SELF Test **12.44** Management of a hotel was concerned with 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 the file **luggage.xls**.
a. Is there evidence of a difference in the median delivery time in the two wings of the hotel? (Use $\alpha = 0.05$.)
b. Compare the results of (a) with those of Problem 10.74 on page 409.

12.45 The director of training for an electronic equipment manufacturer wants to determine whether different training methods have an effect on the productivity of assembly-line employees. She randomly assigns 42 recently hired employees to two groups of 21. The first group receives a computer-assisted, individual-based training program, and the other group receives a team-based training program. Upon completion of the training, the employees are evaluated on the time (in seconds) it takes to assemble a part. The results are in the data file **training.xls**.
a. Using a 0.05 level of significance, is there evidence of a difference in the median assembly times (in seconds) between employees trained in a computer-assisted, individual-based program and those trained in a team-based program?
b. What assumptions must you make in order to do (a) of this problem?
c. Compare the results of Problem 10.20 (a) on page 380 with the results of (a) in this problem. Discuss.

12.46 Nondestructive evaluation is a method that is used to describe the properties of components or materials without causing any permanent physical change to the units. It includes the determination of properties of materials and the classification of flaws by size, shape, type, and location. This method is most effective for detecting surface flaws and characterizing surface properties of electrically conductive materials. Recently, 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. Do the components classified as unflawed have a smaller median crack size than components classified as flawed? The results in terms of crack size (in inches) are in the data file **crack.xls** (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 less for unflawed components than for flawed components?
b. What assumptions must you make in (a)?
c. Compare the results of Problem 10.21 (a) on page 380 with the results of (a) in this problem. Discuss.

12.47 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 waiting time (defined as the time elapsed from when the customer enters the line until he or she reaches the teller window) of all customers during this hour is recorded over a period of 1 week. A random sample of 15 customers is selected (and stored in the file **bank1.xls**), and 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.xls**), and 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 of Problem 10.14 (a) on page 379 with the results of (a) in this problem. Discuss.

12.48 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 data in the file **phone.xls** represent 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 these problems between the two offices? (Use $\alpha = 0.05$.)

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

c. Compare the results of Problem 10.16 (a) on page 379 with the results of (a) in this problem. Discuss.

12.6 KRUSKAL-WALLIS RANK TEST: NONPARAMETRIC ANALYSIS FOR THE ONE-WAY ANOVA

If the normality assumption of the one-way ANOVA F test is not met, 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 12.5. 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

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

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 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, $\bar{X}_j$, against the grand mean, $\bar{\bar{X}}$, 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. 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.

Equation (12.9) defines the Kruskal-Wallis test statistic, H.

KRUSKAL-WALLIS RANK TEST FOR DIFFERENCES AMONG *c* MEDIANS

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right] - 3(n+1) \tag{12.9}$$

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

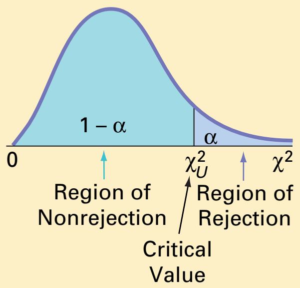
T_j^2 = square of the sum of the ranks assigned to the jth sample

c = number of groups

As the sample sizes in each group get large (that is, greater than 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 χ^2_U upper-tail critical value (see Figure 12.15). Therefore, the decision rule is

$$\text{Reject } H_0 \text{ if } H > \chi^2_U;$$
$$\text{otherwise, do not reject } H_0.$$

FIGURE 12.15
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, concerning the strength of parachutes. If you cannot assume that the tensile strength 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 12.16 presents the data (stored in the file **parachute.xls**), along with the corresponding ranks.

TABLE 12.16
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 12.16 that the third parachute for Supplier 1 has the lowest tensile strength, 17.2. It is given 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, they are 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:

Rank sums: $T_1 = 27 \quad T_2 = 76.5 \quad T_3 = 62 \quad T_4 = 44.5$

As a check on the rankings, recall from Equation (12.7) on page 487 that for any integer n, the sum of the first n consecutive integers is $\frac{n(n+1)}{2}$. Therefore

$$T_1 + T_2 + T_3 + T_4 = \frac{n(n+1)}{2}$$

$$27 + 76.5 + 62 + 44.5 = \frac{(20)(21)}{2}$$

$$210 = 210$$

Using Equation (12.9) on page 493 to test the null hypothesis of equal population medians,

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

$$= \left\{ \frac{12}{(20)(21)} \left[\frac{(27)^2}{5} + \frac{(76.5)^2}{5} + \frac{(62)^2}{5} + \frac{(44.5)^2}{5} \right] \right\} - 3(21)$$

$$= \left(\frac{12}{420} \right)(2{,}481.1) - 63 = 7.8886$$

The statistic H approximately follows a chi-square distribution with $c - 1$ degrees of freedom. Using a 0.05 level of significance, χ^2_U, the upper-tail critical value of the chi-square distribution with $c - 1 = 3$ degrees of freedom is 7.815 (see Table 12.17). 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 not all the suppliers are the same with respect to median tensile strength. The same conclusion is reached by using the p-value approach. In Figure 12.16, observe that the p-value $= 0.0484 < 0.05$.

TABLE 12.17

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

	Upper-Tail Area									
Degrees of Freedom	**.995**	**.99**	**.975**	**.95**	**.90**	**.75**	**.25**	**.10**	**.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
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	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 12.16

Microsoft Excel worksheet for the Kruskal-Wallis rank test for differences among the four medians in the parachute example

	A	B	C	D	E	F	G
1	Tensile-Strength Analysis						
2							
3	Data						
4	Level of Significance	0.05					
5				Group	Sample Size	Sum of Ranks	Mean Ranks
6	Intermediate Calculations			1	5	27	5.4
7	Sum of Squared Ranks/Sample Size	2481.1		2	5	76.5	15.3
8	Sum of Sample Sizes	20		3	5	62	12.4
9	Number of Groups	4		4	5	44.5	8.9
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",			
16				"Do not reject the null hypothesis")			
17		Also					
18		Cell B7:		=(G6 * F6) + (G7 * F7) + (G8 * F8) + (G9 * F9)			
19		Cell B8:		=SUM(E6:E9)			

See Section E12.6 to create this.

You reject the null hypothesis and conclude that there is evidence of a significant difference among the suppliers with respect to the median tensile strength. At this point, you could simultaneously compare all pairs of suppliers to determine which ones differ (see reference 2).

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. However, to use the F test, you must assume that the c samples are from underlying 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 12.6

Learning the Basics

12.49 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?

12.50 Using the results of Problem 12.49,

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

12.51 Periodically, *The Wall Street Journal* has conducted a stock-picking contest. The last one was conducted in March 2001. In this experiment, three different methods were used to select stocks that were expected to perform well during the next five months. Four Wall Street professionals, considered experts on picking stocks, each selected one stock. Four randomly chosen readers of *The Wall Street Journal* each selected one stock. Finally, four stocks were selected by flinging darts at a table containing a list of stocks. The returns of the selected stocks for March 20, 2001, to August 31, 2001 (in percentage return), are given in the following table and stored in the file **contest2001.xls**. Note that during this period, the Dow Jones Industrial Average gained 2.4% (extracted from G. Jasen, "In Picking Stocks, Dartboard Beats the Pros," *The Wall Street Journal*, September 27, 2001, pp. C1, C10).

Experts	Readers	Darts
+39.5	−31.0	+39.0
−1.1	−20.7	+31.9
−4.5	−45.0	+14.1
−8.0	−73.3	+5.4

a. Is there evidence of a significant difference in the median return for the three categories? (Use $\alpha = 0.05$.)

b. Compare the results of (a) with those of Problem 11.8 (a) on page 434.

c. Which assumptions do you think are more appropriate, those of Problem 11.8 (a) or those of part (a) of this problem? Explain.

12.52 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 (measured from check-in to when the patient was called into the clinic area) was measured. The results are stored in the file **erwaiting.xls**.

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 435.

12.53 The following data (stored in the file **cdyield.xls**) represent the nationwide highest yield of different types of accounts (extracted from Bankrate.com, January 24, 2006):

Money Market	6-Month CD	1-Yr CD	2.5-Yr CD	5-Yr CD
4.55	4.75	4.94	4.95	5.05
4.50	4.70	4.90	4.91	5.05
4.40	4.69	4.85	4.85	5.02
4.38	4.65	4.85	4.82	5.00
4.38	4.65	4.85	4.80	5.00

a. At the 0.05 level of significance, is there evidence of a difference in the median yields of the different accounts?

b. Compare the results of (a) with those of Problem 11.11 (a) on page 435.

12.54 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 the file **pen.xls**) 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 11.12 (a) on page 435.

c. Which assumptions do you think are more appropriate, those of Problem 11.12 (a) or those of part (a) of this problem? Explain.

12.55 A sporting goods manufacturing company wanted to compare the distance traveled by golf balls produced using each of four different designs. Ten balls were manufactured with each design and were 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 the file **golfball.xls**:

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 436.

12.56 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 data file **trashbags.xls** 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 11.10 on page 435.

12.7 (*CD-ROM Topic*) CHI-SQUARE TEST FOR A VARIANCE OR STANDARD DEVIATION

When analyzing numerical data, sometimes you need to make conclusions about a population variance or standard deviation. For further discussion, see Section 12.7 on the Student CD-ROM that accompanies this book.

SUMMARY

Figure 12.17 presents a roadmap for this chapter. First, you used hypothesis testing for analyzing categorical response data from two samples (independent and related) 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 applied these methods to the surveys conducted by T.C. Resort Properties. You concluded that a greater proportion of guests are willing to return to the Beachcomber Hotel than to the Windsurfer; that the Golden Palm, Palm Royale, and Palm Princess hotels are different with respect to the proportion of guests who are likely to return; and that the reasons given for not returning to a hotel are dependent on the hotel the guests visited. These inferences will allow T.C. Resort Properties to improve the quality of service it provides.

In addition to the chi-square tests, 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 were violated.

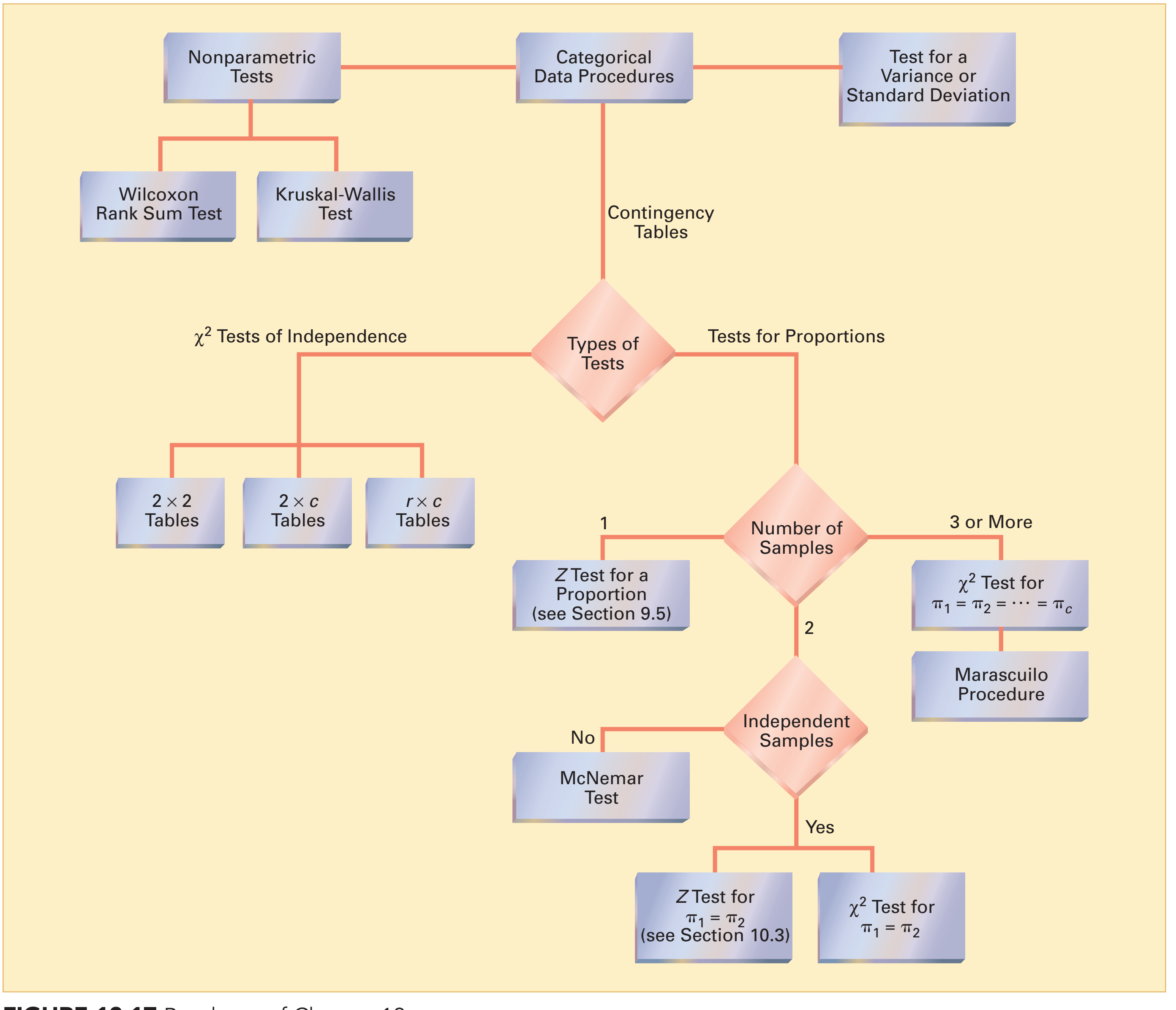


FIGURE 12.17 Roadmap of Chapter 12

KEY EQUATIONS

χ^2 Test for the Difference Between Two Proportions

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

Computing the Estimated Overall Proportion

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

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

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

Computing the Expected Frequency

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

McNemar Test

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

Checking the Rankings

$$T_1 + T_2 = \frac{n(n+1)}{2} \tag{12.7}$$

Large-Sample Wilcoxon Rank Sum Test

$$Z = \frac{T_1 - \dfrac{n_1(n+1)}{2}}{\sqrt{\dfrac{n_1 n_2 (n+1)}{12}}} \tag{12.8}$$

Kruskal-Wallis Rank Test for Differences Among c Medians

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}\right] - 3(n+1) \tag{12.9}$$

KEY TERMS

CHAPTER REVIEW PROBLEMS

Checking Your Understanding

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

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

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

12.60 Under what conditions should you use the McNemar test?

12.61 What is a nonparametric procedure?

12.62 Under what conditions should you use the Wilcoxon rank sum test?

12.63 Under what conditions should you use the Kruskal-Wallis rank test?

Applying the Concepts

12.64 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 on having 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

The survey also 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 relationship between a student's gender and his or her pizzeria selection?
b. What is your answer to (a) if nine of the male students selected Pizza Hut and nine selected other?
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.
e. If appropriate, use the Marascuilo procedure and α = 0.05 to determine which prices are different in terms of pizzeria preference.

12.65 A 2004 study by the American Society for Quality investigated executives' views toward quality. Top executives were asked whether they view quality as a profession in the way law, medicine, engineering, and accounting are viewed, or whether they see practicing quality more as the ability to understand and use a variety of tools and techniques to produce a result. Table (1) provides the responses to this question, cross-classified by the type of industry with which the executive is involved. A second question asked whether the executives' companies actually measure the impact of process improvement initiatives designed to raise the quality of their products or services. Table (2) provides the results to this question.

(1) Do you believe that quality is a profession?

	Manufacturing	**Service**	**Health Care**
Yes	108	88	49
No	72	132	50

(2) Does your company measure the impact of process improvement initiatives?

	Manufacturing	**Service**	**Health Care**
Yes	132	129	54
No	48	91	46

Source: Adapted from G. Weiler, "What Do CEOs Think About Quality?" Quality Progress*, May 2004, 37(5), pp. 52–56.*

a. Is there a significant difference among the three industries with respect to the proportion of top executives who believe quality is a profession? (Use $\alpha = 0.05$.)
b. If appropriate, apply the Marascuilo procedure to (a), using $\alpha = 0.05$.
c. Is there a significant difference among the different industries with respect to the proportion of companies that measure the impact of process improvement initiatives? (Use $\alpha = 0.05$.)
d. If appropriate, apply the Marascuilo procedure to (c), using $\alpha = 0.05$.

12.66 A company is considering an organizational change by adopting 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 shown on page 501.

	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?

12.67 A company that produces and markets videotaped continuing education programs for the financial industry has traditionally mailed sample tapes that contain previews of the programs to prospective customers. Customers then agree to purchase the program tapes or return the sample tapes. A group of sales representatives studied how to increase sales and found that many prospective customers believed it was difficult to tell from a sample tape alone whether the educational programs would meet their needs. The sales representatives performed an experiment to test whether sending the complete program tapes for review by customers would increase sales. They selected 80 customers from the mailing list and randomly assigned 40 to receive the sample tapes and 40 to receive the full-program tapes for review. They then determined the number of tapes that were purchased and returned in each group. The results of the experiment are as follows:

	TYPE OF VIDEOTAPE RECEIVED		
ACTION	**Sample**	**Full**	**Total**
Purchased	6	14	20
Returned	34	26	60
Total	40	40	80

a. At the 0.05 level of significance, is there evidence of a difference in the proportion of tapes purchased on the basis of the type of tape sent to the customer?

b. On the basis of the results of (a), which tape do you think a representative should send in the future? Explain the rationale for your decision.

The sales representatives also wanted to determine which of three initial sales approaches result in the most sales: (1) a video sales-information tape mailed to prospective customers, (2) a personal sales call to prospective customers, and (3) a telephone call to prospective customers. A random sample of 300 prospective customers was selected, and 100 were randomly assigned to each of the three sales approaches. The results, in terms of purchases of the full-program tapes, are as follows:

	SALES APPROACH			
ACTION	**Videotape**	**Personal Sales Call**	**Telephone**	**Total**
Purchase	19	27	14	60
Don't purchase	81	73	86	240
Total	100	100	100	300

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

d. If appropriate, use the Marascuilo procedure and $\alpha = 0.05$ to determine which sales approaches are different.

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

12.68 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 respondents:

	PREFERENCE AFTER TASTE TEST		
PREFERENCE BEFORE 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?

12.69 A market researcher was interested in studying the effect of advertisements on brand preference of new car buyers. Prospective purchasers of new cars were first asked whether they preferred Toyota or GM and then watched video advertisements of comparable models of the two manufacturers. After viewing the ads, the prospective customers again indicated their preferences. The results are summarized in the following table:

PREFERENCE BEFORE ADS	PREFERENCE AFTER ADS: Toyota	GM	Total
Toyota	97	3	100
GM	11	89	100
Total	108	92	200

a. Is there evidence of a difference in the proportion of respondents who prefer Toyota 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.

PREFERENCE	MANUFACTURER: Toyota	GM	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 Toyota 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?

12.70 Researchers studied the goals and outcomes of 349 work teams from various manufacturing companies in Ohio. In the first table, teams are categorized as to whether they had specified environmental improvements as a goal and also according to one of four types of manufacturing processes that best described their workplace. The following three tables indicate different outcomes the teams accomplished, based on whether the team had specified cost cutting as one of the team goals.

TYPE OF MANUFACTURING PROCESS	ENVIRONMENTAL GOAL: Yes	No	Total
Job shop or batch	2	42	44
Repetitive batch	4	57	61
Discrete process	15	147	162
Continuous process	17	65	82
Total	38	311	349

OUTCOME	COST-CUTTING GOAL: Yes	No	Total
Improved environmental performance	77	52	129
Environmental performance not improved	91	129	220
Total	168	181	349

OUTCOME	COST-CUTTING GOAL: Yes	No	Total
Improved profitability	70	68	138
Profitability not improved	98	113	211
Total	168	181	349

OUTCOME	COST-CUTTING GOAL: Yes	No	Total
Improved morale	67	55	122
Morale not improved	101	126	227
Total	168	181	349

Source: Extracted from M. Hanna, W. Newman, and P. Johnson, "Linking Operational and Environmental Improvement Thru Employee Involvement," International Journal of Operations and Production Management*, 2000, 20, pp. 148–165.*

a. At the 0.05 level of significance, determine whether there is evidence of a significant relationship between the presence of environmental goals and the type of manufacturing process.
b. Determine the p-value in (a) and interpret its meaning.
c. At the 0.05 level of significance, is there evidence of a difference in improved environmental performance for teams with a specified goal of cutting costs?
d. Determine the p-value in (c) and interpret its meaning.
e. At the 0.05 level of significance, is there evidence of a difference in improved profitability for teams with a specified goal of cutting costs?
f. Determine the p-value in (e) and interpret its meaning.
g. At the 0.05 level of significance, is there evidence of a difference in improved morale for teams with a specified goal of cutting costs?
h. Determine the p-value in (g) and interpret its meaning.

Team Project

The data file **Mutual Funds.xls** contains information regarding nine variables from a sample of 838 mutual funds. The variables are:

Category—Type of stocks comprising the mutual fund (small cap, mid cap, or large cap)
Objective—Objective of stocks comprising the mutual fund (growth or value)
Assets—In millions of dollars
Fees—Sales charges (no or yes)

Expense ratio—Ratio of expenses to net assets in percentage
2005 return—Twelve-month return in 2005
Three-year return—Annualized return, 2003–2005
Five-year return—Annualized return, 2001–2005
Risk—Risk-of-loss factor of the mutual fund (low, average, or high)

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

12.72 **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 mutual fund and whether there is a fee?

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

12.74 **a.** Construct a 3×3 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 mutual fund and its perceived risk?

Student Survey Database

12.75 Problem 1.27 on page 15 describes a survey of 50 undergraduate students (see the file **undergradsurvey.xls**). 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.

12.76 Problem 1.27 on page 15 describes a survey of 50 undergraduate students (see the file **undergradsurvey.xls**).
a. Select a sample of 50 undergraduate students at your school and conduct a similar survey for those students.
b. For the data collected in (a), repeat Problem 12.75.
c. Compare the results of (b) to those of Problem 12.75.

12.77 Problem 1.28 on page 15 describes a survey of 50 MBA students (see the file **gradsurvey.xls**). 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.

12.78 Problem 1.28 on page 15 describes a survey of 50 MBA students (see the file **gradsurvey.xls**).
a. Select a sample of 50 graduate students in your MBA program and conduct a similar survey for those students.
b. For the data collected in (a), repeat Problem 12.77.
c. Compare the results of (b) to those of Problem 12.77.

Managing the *Springville Herald*

Phase 1

Reviewing the results of its research, the marketing department concluded that a segment of Springville households might be interested in a discounted trial home subscription to the *Herald*. 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 newspaper. Subscribers would pay \$4.50 per week for the newspaper during the 90-day trial period.
2. Moderate discount for the newspaper. Subscribers would pay \$4.00 per week for the newspaper during the 90-day trial period.
3. Substantial discount for the newspaper. Subscribers would pay \$3.00 per week for the newspaper 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 Springville 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 *Herald* after the trial period. Table SH12.1 summarizes the results.

TABLE SH12.1 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	34	37	38	61	170
No	66	63	62	39	230
Total	100	100	100	100	400

EXERCISE

SH12.1 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.

DO NOT CONTINUE UNTIL THE PHASE 1 EXERCISE HAS BEEN COMPLETED.

Phase 2

The marketing department team discussed the results of the survey presented in Chapter 8, on pages 320–321. The team realized that the evaluation of individual questions was providing only limited information. In order to further understand the market for home-delivery subscriptions, the data were organized in the following cross-classification tables:

HOME DELIVERY	READ OTHER NEWSPAPER: Yes	No	Total
Yes	61	75	136
No	77	139	216
Total	138	214	352

HOME DELIVERY	RESTAURANT CARD: Yes	No	Total
Yes	26	110	136
No	40	176	216
Total	66	286	352

INTEREST IN TRIAL SUBSCRIPTION	MONDAY–SATURDAY PURCHASE BEHAVIOR: Every Day	Most Days	Occasionally or Never	Total
Yes	29	14	3	46
No	49	81	40	170
Total	78	95	43	216

INTEREST IN TRIAL SUBSCRIPTION	SUNDAY PURCHASE BEHAVIOR: Every Sunday	2–3 Times/Month	No More Than Once/Month	Total
Yes	35	10	1	46
No	103	44	23	170
Total	138	54	24	216

WHERE PURCHASED	INTEREST IN TRIAL SUBSCRIPTION: Yes	No	Total
Convenience store	12	62	74
Newsstand/candy store	15	80	95
Vending machine	10	11	21
Supermarket	5	8	13
Other locations	4	9	13
Total	46	170	216

SUNDAY PURCHASE BEHAVIOR	MONDAY–SATURDAY PURCHASE BEHAVIOR: Every Day	Most Days	Occasionally or Never	Total
Every Sunday	55	65	18	138
2–3 times/month	19	23	12	54
Once/month	4	7	13	24
Total	78	95	43	216

EXERCISE

SH12.2 Analyze the results of the cross-classification tables. Write a report for the marketing department team and discuss the marketing implications of the results for the *Springville Herald*.

Web Case

Apply your knowledge of testing for the difference between two proportions in this Web 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% percent of the customers preferred its "Concierge Class" travel reward program over the T.C. Resorts "TCPass Plus" program. Visit the SunLow Web site, **www.prenhall.com/Springville/SunLowHome.htm** (or open the Web page file in the Web Case folder on the Student CD-ROM), and examine the survey data. Then answer the following:

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 factors 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 2007* (Redmond, WA: Microsoft Corp., 2007).
9. Winer, B. J., D. R. Brown, and K. M. Michels, *Statistical Principles in Experimental Design*, 3rd ed. (New York: McGraw-Hill, 1989).

Excel Companion

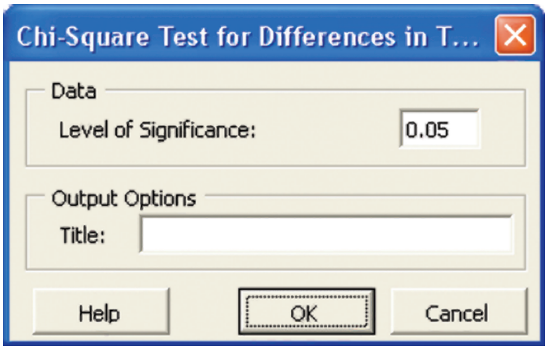
to Chapter 12

E12.1 USING THE CHI-SQUARE TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS

You conduct a chi-square test for the difference between two proportions by either selecting the PHStat2 Chi-Square Test for Differences in Two Proportions procedure or by making entries in the Chi-Square.xls workbook.

Using PHStat2 Chi-Square Test for Differences in Two Proportions

Select **PHStat ➔ Two-Sample Tests ➔ Chi-Square Test for Differences in Two Proportions**. In the procedure's dialog box (shown below), enter the **Level of Significance**, enter a title as the **Title**, and click **OK**.



PHStat2 creates a worksheet in which you enter contingency table data, such as Table 12.2 on page 463, into the rows 4 through 7 Observed Frequencies area. (You can also enter custom row and column labels for your data.)

Before you enter contingency table data, many worksheet cells display the message #DIV/0!. This is not an error.

Using Chi-Square.xls

Open to the **ChiSquare2P** worksheet of the Chi-Square.xls workbook. This worksheet (see Figure 12.3 on page 467) uses the function **CHIINV(*level of significance, degrees of freedom*)** to compute the critical value and the function **CHIDIST(χ^2 *test statistic, degrees of freedom*)** to compute the *p*-value for the Section 12.1 hotel guest satisfaction example. To adapt this worksheet to other problems, enter the problem's contingency table data into the rows 4 through 7 Observed Frequencies area, edit the title in cell A1, and change the level of significance in cell B18, if necessary.

Figure E12.1 shows the cell formulas for rows 6 through 15, not shown in Figure 12.3. Formulas in cells B11, A12, A13, A14, B12, and C12 display the row and column labels entered in the Observed Frequencies area. The results of the formulas in cells F13:G14 are summed in cell B25 to compute the χ^2 test statistic. (The other formulas shown compute row or column totals or the expected frequencies.)

Not shown in Figure E12.1 is the A30 formula. This formula uses an IF function to complete the phrase "Expected frequency assumption . . . ". The *comparison* part of this IF function (see Section E9.1 on page 364), **OR(B13 < 5, C13 < 5, B14 < 5, C14 < 5)**, uses the OR function to make sure that none of the expected frequencies is less than 5, an assumption that is necessary for the chi-square test to be accurate.

	A	B	C	D	E	F	G
3		Observed Frequencies					
4		Hotel				Calculations	
5	Choose Again?	Beachcomber	Windsurfer	Total		fo-fe	
6	Yes	163	154	=SUM(B6:C6)		=B6 - B13	=C6 - C13
7	No	64	108	=SUM(B7:C7)		=B7 - B14	=C7 - C14
8	Total	=SUM(B6:B7)	=SUM(C6:C7)	=SUM(B8:C8)			
9							
10		Expected Frequencies					
11		=B4					
12	=A5	=B5	=C5	Total		(fo-fe)^2/fe	
13	=A6	=D6 * B8/D8	=D6 * C8/D8	=SUM(B13:C13)		=F6^2/B13	=G6^2/C13
14	=A7	=D7 * B8/D8	=D7 * C8/D8	=SUM(B14:C14)		=F7^2/B14	=G7^2/C14
15	Total	=SUM(B13:B14)	=SUM(C13:C14)	=SUM(B15:C15)			

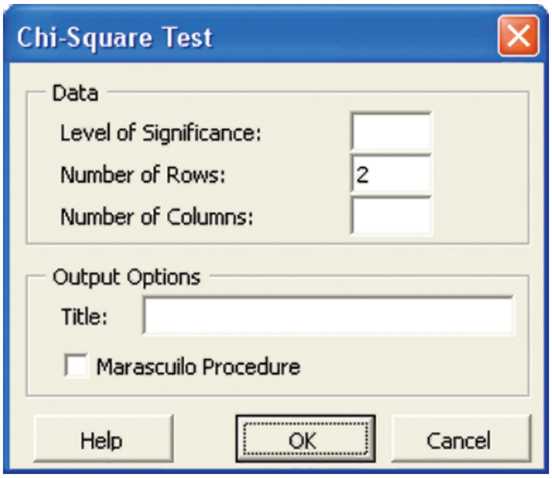
FIGURE E12.1 ChiSquare2P rows 3 through 15

E12.2 USING THE CHI-SQUARE TEST FOR THE DIFFERENCES AMONG MORE THAN TWO PROPORTIONS

You conduct a chi-square test for the differences among more than two proportions by either selecting the PHStat2 Chi-Square Test procedure or by making entries in the **Chi-Square Worksheets.xls** workbook.

Using PHStat2 Chi-Square Test

Select **PHStat ➔ Multiple-Sample Tests ➔ Chi-Square Test**. In the procedure's dialog box (shown below), enter the **Level of Significance**, enter **2** as the **Number of Rows**, and enter the **Number of Columns**. Enter a title as the **Title** and click **OK**. If you want to select the Marascuilo procedure, select **Marascuilo Procedure** before clicking **OK**.



PHStat2 creates a worksheet in which you enter your $2 \times c$ contingency table data, such as Table 12.6 on page 471, in the Observed Frequencies area that begins in row 4. (You can also enter custom row and column labels for your data.) Before you enter contingency table data, many worksheet cells display the message #DIV/0!. This is not an error.

Using Chi-Square Worksheets.xls

Open the **Chi-Square Worksheets.xls** workbook to the worksheet that contains the appropriate $2 \times c$ observed frequency table for your problem. For example, for the guest satisfaction data in Table 12.6 on page 471 that requires a 2×3 table, open to the **ChiSquare2x3** worksheet. Worksheets with empty observed frequencies tables display the message #DIV/0! in many cells. This is not an error, and these messages disappear when you enter your contingency table data.

All $2 \times c$ worksheets contain formulas similar to those in the ChiSquare2P worksheet discussed in Section E12.1. All worksheets have their level of significance set to 0.05, but you can change this value. The **Chi-Square Worksheets.xls** workbook includes the ChiSquare2x3Formulas worksheet, which allows you to examine the formulas of the ChiSquare2x3 worksheet in formatted, formulas view. You should note that the formula in cell A30 verifies whether all expected frequencies are at least 1, an assumption of the Chi-Square test.

Using the Marascuilo Worksheets

Each $2 \times c$ chi square worksheet is linked to a companion Marascuilo worksheet in the **Chi-Square Worksheets.xls** workbook. Marascuilo worksheet names echo the chi-square worksheet to which they are linked, so that **Marascuilo2x3** is linked to **ChiSquare2x3**. (The Marascuilo2x3 worksheet is shown in Figure 12.7 on page 475.)

The **Marascuilo2x3Formulas** worksheet allows you to examine the formulas of the Marascuilo2x3 worksheet in formatted, formulas view. If you examine this worksheet, you see that the formulas for the level of significance, the square root of the critical value, the sample proportions, and the critical range all use one or more values from the ChiSquare2x3 worksheet.

All Marascuilo worksheets compare the absolute differences and critical range values for each pair of groups and display either Significant or Not Significant in column D.

E12.3 USING THE CHI-SQUARE TEST OF INDEPENDENCE

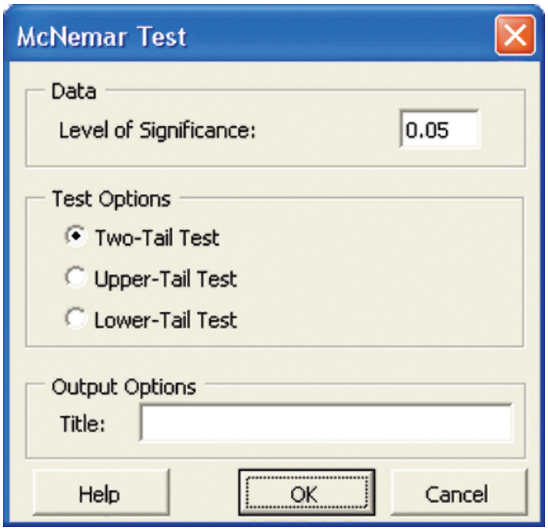
Adapt the instructions of the previous section to the chi-square test of independence. If you use PHStat2, enter your number of rows as the **Number of Rows**. If you use the **Chi-Square Worksheets.xls** workbook, open to either the chi-square worksheets for 3×4, 4×3, 7×3, or 8×3, or open one of the $2 \times c$ worksheets mentioned in Section E12.2.

E12.4 USING THE McNEMAR TEST

You conduct a McNemar test by either selecting the PHStat2 McNemar Test procedure or by making entries in the **McNemar.xls** workbook.

Using PHStat2 McNemar Test

Select **PHStat ➔ Two-Sample Tests ➔ McNemar Test**. In the procedure's dialog box (see top of page 508), enter the **Level of Significance**, click a test option, enter a title as the **Title**, and click **OK**.



PHStat2 creates a worksheet in which you enter the observed frequencies, such as Table 12.13 on page 484, in the Observed Frequencies area that begins in row 3. (You can also enter custom row and column labels for your data.) Before you enter data, many worksheet cells display the message #DIV/0!. This is not an error.

Using McNemar.xls

You open and use either the **McNemar_TT** or the **McNemar_All** worksheets of the **McNemar.xls** workbook to use the McNemar test. These worksheets use the **NORMSINV(*P*<*X*)** function to determine the lower and upper critical values and use the **NORMSDIST(Z *value*)** function to compute the *p*-values from the *Z* value calculated in cell B16. To understand how messages get displayed in these worksheets, read "About the IF Function" on page 364.

The **McNemar_TT** worksheet (see Figure 12.12 on page 485) uses the two-tail test for the Section 12.4 consumer preference example. The **McNemar_All** worksheet also includes the one-tail tests (these additions are shown in Figure E12.2). To adapt these worksheets to other problems, change the observed frequency table data and labels in rows 4 through 7 and (if necessary) the level of significance in cell B11.

	A	B	
24	Lower-Tail Test		
25	Lower Critical Value	-1.6449	=NORMSINV(B11)
26	p-Value	0.9997	=NORMSDIST(B16)
27	Do not reject the null hypothesis		=IF(B26 < B11, "Reject the null hypothesis", "Do not reject the null hypothesis")
28			
29	Upper-Tail Test		
30	Upper Critical Value	1.6449	=NORMSINV(1-B11)
31	p-Value	0.0003	=1 - NORMSDIST(B16)
32	Reject the null hypothesis		=IF(B31 < B11, "Reject the null hypothesis", "Do not reject the null hypothesis")

FIGURE E12.2 McNemar_All worksheet one-tail tests

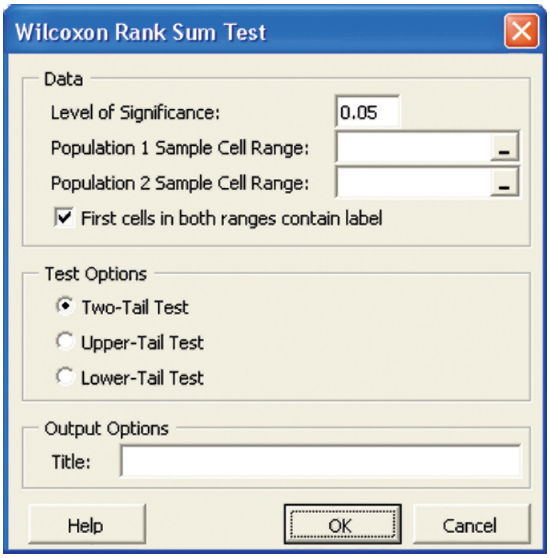
If you want the **McNemar_All** worksheet to show only one of the one-tail tests, first make a copy of that worksheet (see the Excel Companion to Chapter 1). For a lower-tail-test-only worksheet, select and delete rows 29 through 32 and then select and delete rows 18 through 23. For an upper-tail-test-only worksheet, select and delete rows 18 through 28.

E12.5 USING THE WILCOXON RANK SUM TEST

You conduct a Wilcoxon rank sum test by either selecting the PHStat2 Wilcoxon Rank Sum Test procedure or by making entries in the **Wilcoxon.xls** workbook. The PHStat2 procedure uses unsummarized, unstacked data, while the workbook uses summarized data. (See "Counting and Summing Ranks" on page 509 if you have unsummarized, unstacked data and you wish to use the workbook.)

Using PHStat2 Wilcoxon Rank Sum Test

Open to the worksheet that contains the unsummarized, unstacked data for the two independent populations. Select **PHStat ➔ Two-Sample Tests ➔ Wilcoxon Rank Sum Test**. In the procedure's dialog box (shown below), enter the **Level of Significance** and the cell ranges for the **Population 1 Sample Cell Range** and the **Population 2 Sample Cell Range**. Click **First cells in both ranges contain label**, click a test option, enter a title as the **Title**, and click **OK**.



PHStat creates a worksheet similar to the Figure 12.14 worksheet on page 490. However, PHStat2 uses the functions **COUNTIF** and **SUMIF** in formulas in cells B7, B8, B10, and B11 to compute the sample size and sum of the ranks for each population. To learn more about these functions, read "Counting and Summing Ranks," later in this section.

Using Wilcoxon.xls

You open and use either the **Wilcoxon_TT** or the **Wilcoxon_All** worksheets of the Wilcoxon.xls workbook to use the Wilcoxon rank sum test. These worksheets use the **NORMSINV(**P<X**)** function to determine the lower and upper critical values and use the **NORMSDIST(*Z value*)** function to compute the *p*-values from the *Z* value calculated in cell B18. To understand how messages get displayed in these worksheets, read "About the IF Function" on page 364.

The **Wilcoxon_TT** worksheet (see Figure 12.14 on page 490) uses the two-tail test for the Section 12.5 BLK cola sales example. The **Wilcoxon_All** worksheet also includes the one-tail tests (these additions are shown in Figure E12.3). To adapt these worksheets to other problems, change the title in cell A1 and (if necessary) the level of significance, sample sizes, and rank sum values in the tinted cells B4, B7, B8, B10, and B11.

	A	B	
25			
26	Lower-Tail Test		
27	Lower Critical Value	-1.6449	=NORMSINV(B4)
28	*p*-Value	0.0063	=NORMSDIST(B18)
29	Reject the null hypothesis		=IF(B28 < B4, "Reject the null hypothesis", "Do not reject the null hypothesis")
30			
31	Upper-Tail Test		
32	Upper Critical Value	1.6449	=NORMSINV(1 - B4)
33	*p*-Value	0.9937	=1 - NORMSDIST(B18)
34	Do not reject the null hypothesis		=IF(B33 < B4, "Reject the null hypothesis", "Do not reject the null hypothesis")

FIGURE E12.3 Wilcoxon one-tail tests

If you want the Wilcoxon_All worksheet to show only one of the one-tail tests, first make a copy of that worksheet (see the Excel Companion to Chapter 1). For a lower-tail-test-only worksheet, select and delete rows 31 through 34 and then select and delete rows 20 through 25. For an upper-tail-test-only worksheet, select and delete rows 20 through 30.

Counting and Summing Ranks

If you want to use the Wilcoxon.xls workbook but have unsummarized, unstacked data, you can use the **COUNTIF (*cell range for matching, value to be matched*)** function to compute the sample size and the **SUMIF(*cell range for matching, value to be matched, cell range for summing*)** function to compute the sum of the ranks of each population.

To use these functions, first sort your unstacked data worksheet in ascending order, using the column containing the values (not the population labels). Then add a column of ranks, breaking ties by using the method stated on page 487. With your data so arranged, you can use the formula **=COUNTIF(*cell range of all population labels,* "*population 1 name*")** to count the sample size of the population 1 sample and the formula **=SUMIF(*cell range of all population labels,* "*population 1 name*", *cell range of all sorted values*)** to sum the ranks of the population 1 sample. (The population 1 name must appear in a set of double quotation marks.) Create another pair of formulas and use the name of the second population as the ***value to be matched*** to count the sample size and sum the ranks of the population 2 sample.

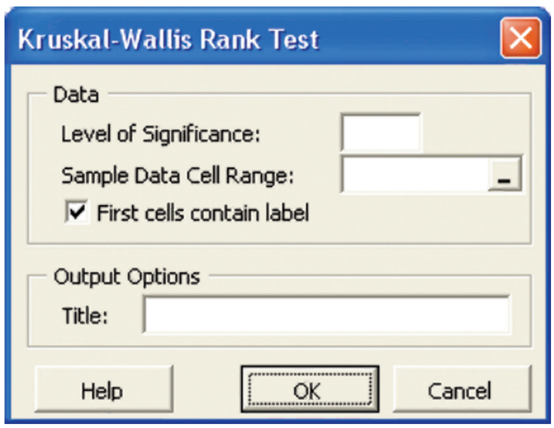
For example, the formulas **=COUNTIF(A1:A21, "EndAisle")** and **=SUMIF(A1:A21, "EndAisle", C1:C21)** would compute the sample size and the sum of ranks for the BLK cola end-aisle sales if placed in empty cells of the **ColaSortedStacked** worksheet of the cola.xls workbook. The formulas **=COUNTIF(A1:A21, "Normal")** and **=SUMIF(A1:A21, "Normal", C1:C21)** would do the same things for the normal display sample.

E12.6 USING THE KRUSKAL-WALLIS RANK TEST

You conduct a Kruskal-Wallis rank test by either selecting the PHStat2 Kruskal-Wallis Rank Test procedure or by making entries in the Kruskal-Wallis Worksheets.xls workbook. The PHStat2 procedure uses unsummarized, unstacked data, while the workbook uses summarized data.

Using PHStat2 Kruskal-Wallis Rank Test

Select **PHStat → Multiple-Sample Tests → Kruskal-Wallis Rank Test**. In the procedure's dialog box (shown below), enter the **Level of Significance** and enter the cell range of the unsummarized, unstacked data as the **Sample Data Cell Range**. Click **First cells contain label**, enter a title as the **Title**, and click **OK**.



Using Kruskal-Wallis.xls

Open the **Kruskal-Wallis Worksheets.xls** workbook to the worksheet that contains the appropriate number of groups (populations) for your problem. For example, for the Section 12.6 parachute example that contains four different groups, open to the **Kruskal-Wallis4** worksheet (shown in Figure 12.16 on page 495).

Kruskal-Wallis worksheets use the function **CHIINV(*level of significance, degrees of freedom*)** to compute the critical value and the function **CHIDIST(χ^2 *test statistic, degrees of freedom*)** to compute the *p*-value. When you open to a worksheet, enter the sample size, sum of ranks, and mean rank values in the tinted cells in columns E through G, the level of significance value in cell B4, and a title in cell A1 to complete the worksheet. #DIV/0! messages that may appear in several cells disappear after you enter data. This is not an error.

Forum Click on the ALTERNATIVE METHODS link to learn more about how the COUNTIF and SUMIF functions (see Section E12.4) could be used with Kruskal-Wallis worksheets.