Nonparametric Statistical Tests: Chi-Square

15

LEARNING OBJECTIVES

- Differentiate parametric tests from nonparametric tests.
- Calculate and interpret a chi-square goodness-of-fit test.
- Calculate and interpret a chi-square test of independence.
- Know when to use a Spearman rank-order correlation coefficient and a Mann–Whitney *U* test.

CHAPTER OVERVIEW

So far, all the tests covered in this text have had two things in common. One is that no matter the test—whether for a *z, t, F*, or *r*—the outcome variable has always been measured at the interval or ratio level. The other is that for each test, the outcome variable was supposed to be normally distributed in the population.

But, sometimes a researcher takes on a study where the outcome variable is ordinal or nominal. Sometimes, the outcome variable isn't normally distributed. Tests for these situations, what are called nonparametric tests, are the subject of this chapter.

15.1 Introduction to Nonparametric Tests

z, t, F, and *r* all come from a family of tests called *parametric tests.* **Parametric tests** should only be used when assumptions about the population, about the parameters, are met. In contrast, **nonparametric tests** don't have to meet the same assumptions. There are two situations where nonparametric tests are used:

1. If the outcome variable is nominal or ordinal, then a nonparametric test is planned from the outset.
2. If a researcher is planning to use a parametric test, but a nonrobust assumption is violated, then the researcher can "fall back" to a nonparametric test.

Nonparametric tests are desirable because they are less restricted by assumptions. However, this comes at a cost—nonparametric statistical tests are usually less powerful than parametric tests. This is a problem because a test with less power is less likely to succeed in rejecting the null hypothesis, the usual goal of hypothesis testing. Also, nonparametric tests work with nominal- or ordinal-level data, not interval/ratio. Nominal- and ordinal-level numbers contain less information than do interval- and ratio-level numbers and this can make it harder to find an effect.

To see how less information means less power, imagine a medication that is only slightly effective in reducing fever. This small effect would be more evident if the temperature were measured to a hundredth of a degree than simply measuring whether or not a person has a fever. The more precisely a researcher can measure the outcome, the greater the ability to find an effect.

Statisticians prefer parametric tests because of their greater power. But, when parametric tests can't be used, when their assumptions are not met, or when the outcome variable is nominal or ordinal, it is time for a nonparametric test. Our primary focus in this chapter will be two different versions of the most commonly used nonparametric test, the chi-square. (The "chi" in chi-square is abbreviated with an uppercase Greek letter, χ, pronounced "kai," so chi-square is written χ^2.) Then, at the end of the chapter, the nonparametric version of a Pearson *r* and nonparametric version of an independent-samples *t* test are explored.

15.2 The Chi-Square Goodness-of-Fit Test

The **chi-square goodness-of-fit test** is a nonparametric, single-sample test that can be used with a nominal dependent variable or a higher-level variable treated as categorical, such as turning actual scores on an exam into letter grades. This type of test is called a **single-sample test** because it compares the results from a single sample to a specific value, usually a population value. The chi-square goodness-of-fit test is the nonparametric version of the single-sample *z* test and single-sample *t* test.

Here's an example of how the chi-square goodness-of-fit test might be used. Suppose a university administrator wants to survey a sample of students to get their opinion about a planned decrease in the intramural sports program. The administrator wants to make sure that the sample represents the university population, particularly in terms of gender. His sample is 59% female. From the registrar, he learns that 52% of the students at the university are female. A chi-square goodness-of-fit test could be used to compare the percentage of women in the sample to the percentage of women in the population. This allows the researcher to determine if obtaining a sample that is 59% female is a common occurrence if the population is 52% female. If it is a common occurrence, then he'll decide that the sample may be representative of the population on this variable and he'll be more likely to trust the results. If it is an uncommon occurrence, then the sample may be an odd one and its results unrepresentative.

To learn how to calculate and interpret the chi-square goodness-of-fit test, imagine this example from a small community where teenagers believe that the local police single them out more often than adult drivers for traffic stops. To investigate this, a traffic researcher, Dr. Koenig, randomly selected 72 tickets from all the tickets issued during a calendar year. As the age of the driver was recorded on each ticket, Dr. Koenig determined that 11 of the tickets went to teen drivers and 61 went to adults. The results are displayed in two cells, shown in Table 15.1.

TABLE 15.1 Ages of 72 Ticketed Drivers

Teenagers	Adults
11	61

11 of the 72 ticketed drivers (15.28%) were teenagers. 61 of the 72 ticketed drivers (84.72%) were adults.

Teenagers received 11 of the 72 traffic tickets (15.28%) in the sample. Dr. Koenig found, from the Department of Motor Vehicles, that 8% of licensed drivers in the population of the town are teenagers. She reasoned that if teens were treated the same as adults, 8% of the traffic tickets should go to teens. Yet, the teens received more than 15% of the tickets. Do teens get more than their fair share of traffic tickets? Or, can the difference between what is expected (8%) and what is observed (15.28%) be explained by sampling error?

"**T**om **a**nd **H**arry **d**espise **c**rabby **i**nfants" is the mnemonic for the six steps of hypothesis testing for a chi-square goodness-of-fit test: (1) Pick the **test,** (2) check the **assumptions,** (3) list the **hypotheses,** (4) set the **decision rule,** (5) **calculate** the test statistic, and (6) **interpret** the results.

Step 1 Pick a Test

This scenario calls for a single-sample test because a sample value, 15.28%, is being compared to a specific value, 8%. It may seem as if there are two samples (a sample of teens and a sample of adults), but it is just one sample (a sample of ticketed drivers) where the age status of each case is being measured (teen or adult) as the outcome variable. The outcome variable (the category of who gets a ticket, teen or adult) is a nominal-level variable, so the appropriate test is a chi-square goodness-of-fit test. Goodness-of-fit is abbreviated GOF.

Step 2 Check the Assumptions

The chi-square goodness-of-fit test has three assumptions:

1. *Random sample*. The sample should be a random sample from the population. This is a robust assumption, so it can be—and usually is—violated. As with parametric tests, if the random sample assumption is violated, then the researcher needs to be careful about generalizing the results from the sample. For the ticket data, there is a random sample of tickets, so the first assumption is not violated. But, the sample is only from tickets given in this town, so generalizability is limited.

2. *Independence of observations*. The cases in the sample should be independent of each other. This means that the observations don't influence each other. This assumption is not robust, so one can't proceed with the planned test if it is violated. With the ticket data, random sampling was used, so the cases within the sample aren't connected to each other. In addition, no case was in the sample more than once. The independence of observations assumption is not violated.

3. *Expected frequencies*. In order to conduct a chi-square test, all cells must have expected frequencies of at least 5. The chi-square test only works if there are enough cases in each cell. We will cover the method for calculating expected frequencies later in this chapter. Just know that the expected frequencies assumption is not robust, so the chi-square can't be calculated if each cell doesn't have enough cases. (For this example, expected frequencies will be large enough and this assumption isn't violated.)

Step 3 List the Hypotheses

The hypotheses for the chi-square goodness-of-fit test are easier to express in words than in mathematical symbols:

- The null hypothesis states that the proportion of each category in the population matches specified values. This means that since the sample comes from the population, the proportions in the sample should match the specified values. But, as sampling error exists, one shouldn't expect the sample to match the specified values exactly.
- The alternative hypothesis says that the distribution of the characteristic in the population is different from the specified values. The alternative hypothesis means that the difference between the percentages in the sample and the specified values is too large to be explained by sampling error.

H_0: In the population, the percentage of tickets received by teenage drivers = 8%.

H_1: In the population, the percentage of tickets received by teenage drivers ≠ 8%.

Step 4 Set the Decision Rule

The decision rule for a chi-square goodness-of-fit test follows the same format as the decision rules for *z, t, F,* and *r:*

- Find the critical value of chi-square, χ^2_{cv}, using a table of critical values of chi-square (Appendix Table 9).
  - The critical value is the value that separates the rare zone from the common zone in the sampling distribution of chi-square.
  - The sampling distribution of chi-square is the distribution of chi-square values that would occur if the null hypothesis were true.
- Compare χ^2, the observed value of the chi-square statistic to χ^2_{cv}.
  - If $\chi^2 \geq \chi^2_{cv}$, then the observed value of chi-square falls in the rare zone, and reject the null hypothesis.
  - If $\chi^2 < \chi^2_{cv}$, then the observed value of chi-square falls in the common zone, so fail to reject the null hypothesis.

A portion of Appendix Table 9, the table of critical values of chi-square, is shown in Table 15.2. There are three characteristics to note about the table of critical values of chi-square:

1. The chi-square is always a two-tailed test. Because chi-square is calculated using squared values, it is always positive. If the results are statistically significant, the researcher will need to look at the direction of the difference in the sample to figure out the direction of the difference in the population.
2. There are three alpha levels: .01, .05, and .10. The alpha level most commonly used, the column where $\alpha = .05$, appears in bold. That alpha level represents a 5% chance of making a Type I error. (Type I error occurs when the null hypothesis is rejected by mistake.)
3. The critical value of chi-square also depends on how many degrees of freedom there are. Each row in the table of critical values represents a different number of

degrees of freedom. The critical value of chi-square is found at the intersection of the column for the desired alpha level with the row for the correct number of degrees of freedom.

TABLE 15.2 Critical Values of Chi-Square (Appendix Table 9)

	Alpha Level		
df	.10	.05	.01
1	2.706	**3.841**	10.828
2	4.605	**5.991**	13.816
3	6.251	**7.815**	16.266
4	7.779	**9.488**	18.467
5	9.236	**11.070**	20.515
6	10.645	**12.592**	22.458
7	12.017	**14.067**	24.322

The critical value of χ^2 is found at the intersection of the row with the correct number of degrees of freedom and the column with the desired alpha level. The most commonly used alpha level is $\alpha = .05$, so those values are in bold.

The next step is to find the degrees of freedom, which can be calculated using Equation 15.1.

Equation 15.1 Degrees of Freedom for a Chi-Square Goodness-of-fit Test

$$df = k - 1$$

where df = degrees of freedom
k = number of categories

Equation 15.1 says to calculate degrees of freedom (df) for a chi-square goodness-of-fit test as the number of categories, k, minus 1. For the ticket data, the variable is age of drivers and it has two categories—adults and teens. This can be seen in the two cells in Table 15.1. With $k = 2$, degrees of freedom is calculated as follows:

$$\begin{aligned} df &= k - 1 \\ &= 2 - 1 \\ &= 1 \end{aligned}$$

Dr. Koenig is willing to have a 5% chance of making a Type I error, so she sets alpha at .05. Looking in the table of critical values of chi-square at the intersection of the row with $\alpha = .05$ and the row with $df = 1$, she finds $\chi^2_{cv} = 3.841$. The sampling distribution of chi-square, with the rare and common zones marked, is seen in Figure 15.1.

Now that χ^2_{cv} is known, the decision rule can be written:

- If $\chi^2 \geq 3.841$, reject H_0.
- If $\chi^2 < 3.841$, fail to reject H_0.

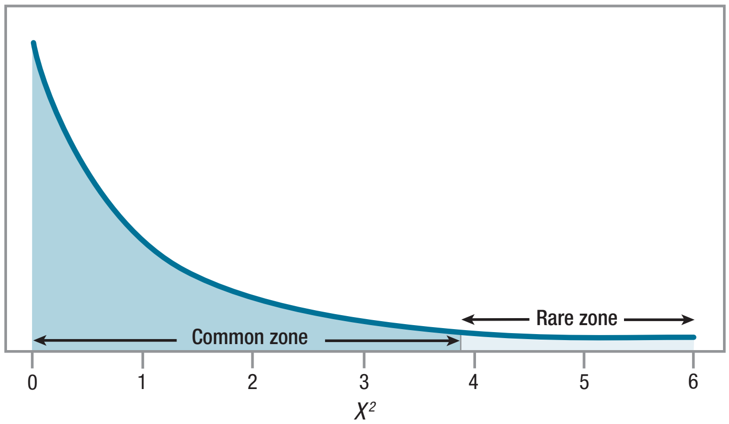

Figure 15.1 Sampling Distribution of Chi-Square with 1 Degree of Freedom When $df = 1$, the critical value of chi-square is 3.841 if $\alpha = .05$.

Step 5 Calculate the Test Statistic

Calculating the test statistic, χ^2, involves comparing the *observed* frequency of each category of the dependent variable in the sample to the frequency that would be *expected* if the null hypothesis were true:

- If the differences between observed and expected are small enough to be accounted for by sampling error, then the χ^2 value will be small, it will land in the common zone of the sampling distribution, and the null hypothesis won't be rejected.
- If the differences between observed and expected are too large to be accounted for by sampling error, then the χ^2 value will be large, it will land in the rare zone, and the null hypothesis will be rejected.

In order to calculate chi-square, one needs to know the observed frequency for each category and the expected frequency for each category. In Chapter 2, f was introduced as the abbreviation for frequency. Now f_{Observed} will be used as the abbreviation for observed frequency and f_{Expected} for expected frequency.

The observed frequencies for each cell (category) of the dependent variable in the sample are already known—there were 11 teens with tickets and 61 adults with tickets. Those are the observed frequencies and **Table 15.3** organizes them with each observed frequency in a cell.

The next step is to calculate the expected frequencies for each cell. It is already known that 8% of drivers are teens. The total percentage of drivers is 100%. So, 100% − 8%, or 92%, of drivers must be adults. These two percentages, 8% for teens

TABLE 15.3 Observed Frequencies for the Ages of Ticketed Drivers ($N = 72$)

	Teenagers	Adults
Observed Frequency	11	61

In the random sample of 72 traffic tickets, 11 were issued to teenagers and 61 to adults.

and 92% for adults, are the expected *percentages*, abbreviated $\%_{\text{Expected}}$, for the two cells. If the null hypothesis is true, one would expect that 8% of tickets would go to teens and 92% to adults. These expected percentages are used to calculate the expected frequencies, as shown in Equation 15.2.

Equation 15.2 Formula for Calculating Expected Frequency

$$f_{\text{Expected}} = \frac{\%_{\text{Expected}} \times N}{100}$$

where f_{Expected} = expected frequency for a cell/category
$\%_{\text{Expected}}$ = expected percentage for a cell/category
N = total number of cases in the sample

To calculate the expected frequency of traffic tickets for teenagers in this sample where $\%_{\text{Expected}} = 8\%$ and $N = 72$, Dr. Koenig uses Equation 15.2:

$$\begin{aligned} f_{\text{Expected}} &= \frac{\%_{\text{Expected}} \times N}{100} \\ &= \frac{8 \times 72}{100} \\ &= \frac{576.0000}{100} \\ &= 5.7600 \\ &= 5.76 \end{aligned}$$

This means that if the null hypothesis were true and teens received their fair share of the tickets, which is 8%, one would expect 5.76 of these 72 tickets to have been issued to teenagers. Don't be bothered by the fact that there are fractional tickets. Expected frequencies don't have to be whole numbers.

Equation 15.2 is also used by Dr. Koenig to calculate the expected frequency for adult tickets. The sample size, N, is still 72, but the expected percentage is now 92%:

$$\begin{aligned} f_{\text{Expected}} &= \frac{\%_{\text{Expected}} \times N}{100} \\ &= \frac{92 \times 72}{100} \\ &= \frac{6{,}624.0000}{100} \\ &= 66.2400 \\ &= 66.24 \end{aligned}$$

If the null hypothesis were true, one would expect the vast majority of the 72 tickets in the sample, 66.24, to go to adults. Only 5.76 are expected to go to teens. Table 15.4 shows the observed frequencies and the expected frequencies. Note that all expected frequencies are greater than 5, so the third assumption was not violated.

TABLE 15.4 Observed Frequencies and Expected Frequencies for the Ages of Ticketed Drivers

	Teenagers	Adults	
Observed Frequency	11	61	$N = 72$
Expected Frequency	5.76	66.24	$\Sigma = 72.00$

The sum of the expected frequencies is the same as the sample size.

Notice something interesting in Table 15.4. If the expected frequencies for all the categories are added up, the total is the same as the original sample size: $5.76 + 66.24 = 72.00$. This will always be the case and is a good way to check that the math was done correctly in calculating expected frequencies.

A Common Question

Q Do I have to use Equation 15.2 to calculate expected frequencies for each cell?

A No. You can take advantage of the fact that the expected frequencies add up to *N* and only calculate expected frequencies for as many categories as there are degrees of freedom. Then subtract the calculated frequencies from *N* to find the missing frequency.

Now that the observed frequencies and expected frequencies are known, they will be used to calculate the test statistic, chi-square. The formula for chi-square is shown in Equation 15.3.

Equation 15.3 Formula for Calculating Chi-Square (χ^2)

$$\chi^2 = \Sigma \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}}$$

where χ^2 = chi-square value
f_{Observed} = observed frequency for a cell/category
f_{Expected} = expected frequency for a cell/category

To use Equation 15.3 to calculate chi-square, follow these four steps:

1. For each cell/category, subtract the expected frequency from the observed frequency.
2. Square each difference.
3. Divide each squared difference by its respective expected frequency to yield a quotient.
4. Sum all the quotients to obtain the chi-square value.

Here are the calculations for the ticket data, where it is found that $\chi^2 = 5.18$:

$$\begin{aligned}
\chi^2 &= \sum \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}} \\
&= \frac{(11 - 5.76)^2}{5.76} + \frac{(61 - 66.24)^2}{66.24} \\
&= \frac{5.2400^2}{5.76} + \frac{-5.2400^2}{66.24} \\
&= \frac{27.4576}{5.76} + \frac{27.4576}{66.24} \\
&= 4.7669 + 0.4145 \\
&= 5.1814 \\
&= 5.18
\end{aligned}$$

Step 6 Interpret the Results

In interpreting the results of a chi-square goodness-of-fit test, there are two questions to be addressed: (1) Was the null hypothesis rejected? (2) If so, what is the direction of the results?

Was the Null Hypothesis Rejected?

To answer this question, refer back to the decision rule generated in Step 4. Decide which of the two decision rules is true:

- Is $5.18 \geq 3.841$? If so, reject H_0 and call the results statistically significant.
- Is $5.18 < 3.841$? If so, fail to reject H_0 and call the results not statistically significant.

With the ticket data, the chi-square value (5.18) is greater than the critical value (3.841), so the first statement is true and the null hypothesis is rejected. This means the difference between the percentages observed in the sample and those found in the population is statistically significant. The differences are too large to be explained by sampling error, so we conclude that the distribution of the dependent variable in the population differs from the specified value in the null hypothesis. Here, this means that the percentage of tickets issued to teens in the population is probably not 8%.

Writing results in APA format for chi-square calls for six pieces of information: (1) what test was done, (2) how many degrees of freedom there were, (3) what the sample size was, (4) what the value of the test statistic was, (5) what alpha level was selected, and (6) whether the null hypothesis was rejected. For the traffic ticket data, Dr. Koenig would write

$$\chi^2(1, N = 72) = 5.18, p < .05$$

1. χ^2 reveals that the test statistic is a chi-square value.
2. The 1 in the parentheses states the degrees of freedom.
3. $N = 72$ gives the sample size.
4. 5.18 is the value of the test statistic that was calculated.

5. .05 indicates that alpha was set at the .05 level.
6. $p < .05$ says that the null hypothesis was rejected and that the value of the test statistic (5.18) is a rare one if the null hypothesis is true.

What Is the Direction of the Results?

If the results were statistically significant, the researcher needs to comment on the direction of the difference. This can be done by comparing what was observed to what was expected. If the results were not statistically significant, then there's not enough evidence to say a difference exists so there's no need to worry about the direction of the difference.

With the traffic ticket data, there are just two categories, so it is easy to tell the direction of the difference. Teenagers account for 8% of the drivers, but they received 15.28% of the tickets. The researcher can say that 15.28% is statistically higher than 8%. Dr. Koenig can conclude that teens in this town get statistically significantly more tickets than expected for the number of teens who are drivers.

Putting It All Together

A four-point interpretation (What was done? What was found? What does it mean? What suggestions are there for future research?) can be completed for a chi-square goodness-of-fit test. Here's an interpretation for the traffic ticket data:

> A traffic researcher analyzed a sample of traffic tickets from a town in order to determine if teen drivers received proportionally more tickets than adult drivers. 8% of the drivers in the town were teens, but 15.28% of traffic tickets issued went to teens. The difference was statistically significant [$\chi^2(1, N = 72) = 5.18$, $p < .05$]. Teenage drivers in this town get more than their fair share of traffic tickets, almost twice as many as expected. Future research should extend the study to other municipalities. It would also be worthwhile to explore whether the over-ticketing of teen drivers is deserved because they are worse drivers or because they are being unfairly targeted by the police.

Worked Example 15.1

Here's another example for a chi-square goodness-of-fit test. Courtney had not done well in intro psych and tried to make the case that the instructor, Dr. Wald, was an unfairly harsh grader. As evidence, she counted the number of A's, B's, C's, D's, and F's given as final grades for the 84 students in Dr. Wald's two sections of intro psych (Table 15.5). Courtney then found out, from the registrar, the percentages of A's, B's, C's, D's, and F's given in all other sections of intro psych that semester. In this table, D's and F's are combined into one category for reasons that will be explained shortly. This distribution, the information from the registrar, is shown in gray in Figure 15.2. Courtney then transformed the distribution of grades for Dr. Wald's two sections of intro psych into percentages, shown in dark blue in Figure 15.2. As Courtney complained to the school dean, "It is clear in my graph—Professor Wald gives fewer A's and B's than the other instructors, but more C's and more D's and F's. Dr. Wald is a harsh and unfair grader." Is he? To try to clear his name, Dr. Wald is going to need a statistical test.

TABLE 15.5 Observed Distribution of Final Grades in Dr. Wald's Sections of Intro Psych and Percentages of Different Grades Awarded in All Sections of Intro Psych

	A	B	C	D or F	
Number of grades awarded in Dr. Wald's sections	10	24	41	9	$N = 84$
Percentage of grades awarded in all sections of intro psych	19%	33%	42%	6%	$\Sigma = 100\%$

Grades of D and F are counted together because there are so few of them. This will serve to keep the expected frequency for this cell above 5.

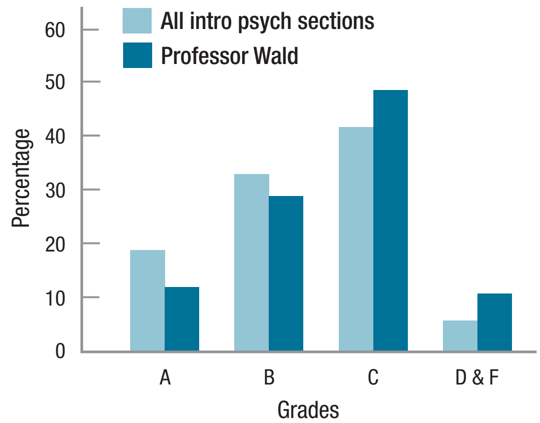

Figure 15.2 Final Grades in Professor Wald's Intro Psych Classes Compared to All Intro Psych Instructors Professor Wald gives fewer high grades and more low grades than do the other instructors. Is the difference a statistically significant one? Or, is it explained by normal variability in sampling?

Step 1 Pick a Test. The question being asked is whether the distribution of grades in Dr. Wald's class is different from the distribution in the larger population. In other words, is the difference between his classes and those of the other professors small enough to be due to the random error associated with sampling? A chi-square goodness-of-fit test addresses this question. Here, letter grades, which are ordinal, are treated as categorical. The appropriateness of this will become clear when two grade categories are merged into one.

Step 2 Check the Assumptions. The three assumptions for a chi-square goodness-of-fit test are (1) random sample, (2) independence of observations, and (3) adequate expected frequencies:

1. *Random sample*. With this example, the importance of the random sample assumption is crystal clear. Comparing this sample of grades from two sections to the larger population only makes sense if these two sections have students similar to the students in the other sections. If, for example, both of Dr. Wald's sections were 8 A.M. sections, it is possible that the students who register for an early morning section differ from those who register for classes that meet at a more reasonable time.

The ideal scenario, from an experimenter's standpoint, would be that students indicate a desire to take intro psych and then are randomly assigned to sections. But, of course, that didn't happen, so the random samples assumption is violated. The objective of the assumption, though, is to study a sample that is representative of the population. When the random samples assumption is violated, a researcher can proceed with the study if he or she makes the case that the sample is representative. As there was nothing unusual about his two sections, Dr. Wald is willing to continue with the chi-square goodness-of-fit test.

2. *Independence of observations.* The independence of observations assumption is not violated as each student was in only one section.
3. *Adequate expected frequencies.* All cells must have expected frequencies of at least 5. As will be seen later, each cell has an expected frequency greater than 5, so this assumption is not violated. To achieve this, Dr. Wald put together the small number of D's and F's in one cell.

Step 3 List the Hypotheses. The null hypothesis will state that the distribution of outcomes in the population is the same as is specified. The alternative hypothesis will state that the distribution of outcomes in the population is different from the specified values. The specified values, as percentages, are found in Table 15.5.

H_0: In the population, the distribution of A's, B's, C's, and D's/F's is, respectively, 19%, 33%, 42%, and 6%.

H_1: In the population, the distribution of A's, B's, C's, and D's/F's is not, respectively, 19%, 33%, 42%, and 6%.

Step 4 Set the Decision Rule. Finding the critical value of chi-square depends on what is set as an acceptable risk of Type I error and the number of degrees of freedom. Dr. Wald follows the convention for Type I error and uses $\alpha = .05$. Degrees of freedom are calculated with Equation 15.1, which involves knowing k, the number of categories of outcome that are possible. There are four categories for grades (A, B, C, and D or F), so $k = 4$. To calculate degrees of freedom, apply Equation 15.1:

$$\begin{aligned} df &= k - 1 \\ &= 4 - 1 \\ &= 3 \end{aligned}$$

Looking in Appendix Table 9, the table of critical values of χ^2, at the intersection of the column for $\alpha = .05$ and the row for $df = 3$, one finds that $\chi^2_{cv} = 7.815$. Here is the decision rule:

- If $\chi^2 \geq 7.815$, reject H_0.
- If $\chi^2 < 7.815$, fail to reject H_0.

Step 5 Calculate the Test Statistic. To calculate the chi-square value, a researcher needs to know the observed and expected frequencies. The observed frequencies were shown in Table 15.5, which also displays the percentages for the expected frequencies.

To calculate the expected frequency for A's, use Equation 15.2, where $\%_{\text{Expected}} = 19\%$ and $N = 84$:

$$
\begin{aligned}
f_{\text{Expected}} &= \frac{\%_{\text{Expected}} \times N}{100} \\
&= \frac{19 \times 84}{100} \\
&= \frac{1{,}596.0000}{100} \\
&= 15.9600 \\
&= 15.96
\end{aligned}
$$

(The value 19% can be found in Table 15.5.)

Dr. Wald gave only 10 A's, but if his grade distribution was exactly the same as that found in the entire population, he would have distributed 15.96 A's. The expected frequency for B's is

$$f_{\text{Expected}} = \frac{\%_{\text{Expected}} \times N}{100} = \frac{33 \times 84}{100} = 27.72$$

The expected frequency for C is

$$f_{\text{Expected}} = \frac{\%_{\text{Expected}} \times N}{100} = \frac{42 \times 84}{100} = 35.28$$

To find the expected frequency for the last cell, Dr. Wald takes advantage of the fact that there are 3 degrees of freedom, and once the expected frequencies for three cells are known, the fourth can be determined. The sum of the expected frequencies for all the cells will be the same as N, which in this case is 84. So, Dr. Wald subtracts the expected frequencies for the A, B, and C cells from 84, to find the remainder, which is the expected frequency for the D or F cell:

$$84 - 15.96 - 27.72 - 35.28 = 5.04$$

Table 15.6 shows the observed frequencies and the expected frequencies for the four categories. To calculate the chi-square statistic, Dr. Wald uses Equation 15.3:

$$
\begin{aligned}
\chi^2 &= \sum \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}} \\
&= \frac{(10 - 15.96)^2}{15.96} + \frac{(24 - 27.72)^2}{27.72} + \frac{(41 - 35.28)^2}{35.28} + \frac{(9 - 5.04)^2}{5.04} \\
&= \frac{-5.9600^2}{15.96} + \frac{-3.72^2}{27.72} + \frac{5.7200^2}{35.28} + \frac{3.96^2}{5.04} \\
&= \frac{35.5216}{15.96} + \frac{13.8384}{27.72} + \frac{32.7184}{35.28} + \frac{15.6816}{5.04} \\
&= 2.2257 + 0.4992 + 0.9274 + 3.1114 \\
&= 6.7637 \\
&= 6.76
\end{aligned}
$$

For these data, $\chi^2 = 6.76$.

TABLE 15.6 Observed Frequencies and Expected Frequencies for Students in Two Sections of Intro Psych

	A	B	C	D or F	
$f_{Observed}$	10	24	41	9	$\Sigma = 84$
$f_{Expected}$	15.96	27.72	35.28	5.04	$\Sigma = 84.00$

Once the observed frequencies and expected frequencies are known for each category, Equation 15.3 can be used to calculate the chi-square value.

Step 6 Interpret the Results. To decide if the null hypothesis is rejected, Dr. Wald uses the decision rule:

- Is $6.76 \geq 7.815$? If so, reject the null hypothesis and call the results statistically significant.
- If $6.76 < 7.815$? If so, fail to reject the null hypothesis and the results are called not statistically significant.

The second statement is true: 6.76 is less than 7.815. As seen in Figure 15.3, the observed value of chi-square, 6.76, falls in the common zone. Dr. Wald has failed to reject the null hypothesis. There is not enough evidence to conclude that the percentages of different grades in the population differ from what was indicated by the null hypothesis. With a sample size of 84, the discrepancy of the observed frequencies from the expected frequencies was small enough that sampling error could account for it. There is insufficient evidence to conclude that Dr. Wald grades more harshly than do the other intro psych faculty.

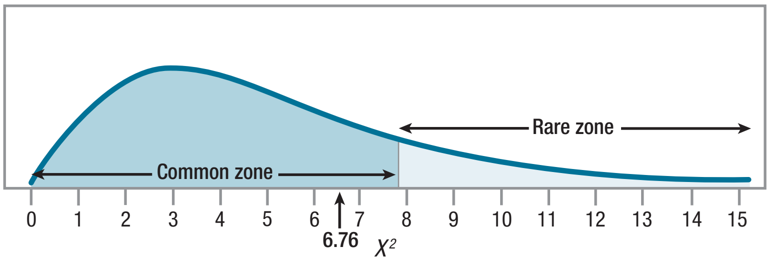

Figure 15.3 Is Dr. Wald Harsh and Unfair? Comparing an Observed Chi-Square Value to the Expected Distribution of Chi-Square Values The observed chi-square value, 6.76, falls in the common zone, meaning insufficient evidence exists to justify rejecting the null hypothesis. There is not enough evidence to conclude that this professor is harsh and unfair.

The results in Figure 15.2 show C's, and D's and F's, turning up more often in Dr. Wald's grades than among the grades of the other instructors. So, it may not seem correct to conclude that insufficient evidence exists to say that Dr. Wald gives more low grades. However, remember that hypothesis testing works like the American legal system: finding a defendant not guilty is different from saying the defendant is innocent. Saying that the evidence is insufficient is not the same as saying Dr. Wald gives the same grades as the other instructors.

In APA format, the results would be reported as

$$\chi^2(3, N = 84) = 6.76, p > .05$$

- χ^2 says that the test is a chi-square test.
- 3 is the degrees of freedom.
- $N = 84$ gives the sample size.
- 6.76 is the calculated value of chi-square.
- .05 indicates that alpha was set at the .05 level.
- $p > .05$ says that the null hypothesis was not rejected. The value of the test statistic (6.76) is a common one. It happens more than 5% of the time when the null hypothesis is true.

Here is a four-point interpretation for these results:

A psychology professor gave more C's, D's, and F's than the other instructors. A chi-square goodness-of-fit test was used to see if he was a harsher grader than the others. The results showed that his grade distribution was not statistically different [$\chi^2(3, N = 84) = 6.76, p > .05$]. This means there is not enough evidence to say that this professor grades differently from the other instructors. If one were to replicate this study, it would be desirable to use a larger sample and to measure characteristics like sex and GPA in order to make sure that the students in his classes are similar to students in the other classes.

Practice Problems 15.1

15.01 People in the world are classified in terms of their natural hair color: black, blonde, brown, red, auburn, chestnut, grey, or white. A hair salon owner keeps track of the natural color of her clients. If she wants to use a chi-square goodness-of-fit test to compare the distribution of hair colors in her salon to the worldwide distribution, how many degrees of freedom will she have?

15.02 If $\alpha = .01$ and $df = 2$, what is χ^2_{cv}?

15.03 Current worldwide estimates are that 51.69% of births are boys and 48.31% are girls. A demographer in the United States obtained a random sample of 5,873 births from all 50 states, determined the sex of each child, and planned to use a chi-square goodness-of-fit test to compare the sex of children born in the United States to the world rate. What are his expected frequencies?

15.04 A researcher is conducting a chi-square goodness-of-fit test on a variable that has two categories. The observed frequencies are 75 for Cell A and 82 for Cell B. The expected frequencies are 62.80 for Cell A and 94.20 for Cell B. Calculate χ^2.

15.05 If $N = 73$, $df = 1$, $\alpha = .05$, and $\chi^2 = 4.72$, write the results in APA format.

15.3 Calculating the Chi-Square Test of Independence

There's a second chi-square test to learn about, the *chi-square test of independence*. Like the chi-square goodness-of-fit test, it uses nominal data (or higher-level data treated as categorical), but it is not a single-sample test. The **chi-square test of**

independence is a nonparametric test that answers the question of whether two or more samples of cases differ on some nominal-level variable. For example, comparing a sample of boys to a sample of girls to see if the nominal variable of having an eating disorder differs between the two sexes calls for a chi-square test of independence. The chi-square test of independence is also known as the chi-square test of association or the chi-square test for contingency tables.

This chi-square test is called a test of "independence" because it answers the question of whether two variables are related to each other or are independent. The chi-square test of independence functions like a difference test ("Do the sexes differ in the prevalence of eating disorders?") and like a relationship test ("Is there a relationship between sex and the presence of an eating disorder?"). This shows that though relationship tests and difference tests may look different, at their core they are the same.

To learn how to conduct a hypothesis test with a chi-square test of independence, imagine the following example: An educational psychologist, Dr. Pradesh, wants to explore whether a student reads a textbook before or after class has an impact on how well he or she does in that class. Dr. Pradesh puts together a random sample of 50 students from introductory psychology classes at her university, then randomly assigns 26 to read the textbook chapters before class and the other 24 to read them after the lectures. At the end of the semester, the students' grades are classified as high (A or B) or low (C, D, or F). The question she asks can be phrased as a relationship question ("Is there a *relationship* between when one reads the text and how well one does in the class?") or a difference question ("Does class performance *differ* depending on when one reads the text?"). These two questions are really the same and both are appropriate for a chi-square test of independence.

Table 15.7 shows the results of Dr. Pradesh's study. The matrix in Table 15.7 is called a **contingency table**, because it shows how the values of the cases on the dependent variable depend on the category of the independent variable. This table illustrates the degree to which students' grades are *contingent on* when they read the text. It is also called a cross-tabulation table because it indicates how the levels of one variable intersect with the levels of the other variable.

There are four cells in this table and each student fits in only one cell. Following convention, the independent variable (when the text is read) is the row variable and

TABLE 15.7 Observed Frequencies for Read the Text Before Lecture vs. Read the Text After Lecture Study

	High Grade	Low Grade	
Read text before lecture	A 20	B 6	26
Read text after lecture	C 12	D 12	24
	32	18	$N = 50$

This contingency table has four cells (A through D), one for each combination of the independent variable (when the text is read) and the dependent variable (grade). The numbers in the far right column (24 and 26) are the frequencies for the rows and the numbers in the bottom row (32 and 18) are the frequencies for the columns.

the dependent variable (high or low grade) is the column variable. Here is what each of the four cells tallies:

- Cell A counts the students who were in the "read before" group and received a high grade.
- Cell B counts the students who were in the "read before" group and received a low grade.
- Cell C counts the students who were in the "read after" group and received a high grade.
- Cell D counts the students who were in the "read after" group and received a low grade.

Contingency tables allow researchers to calculate and compare percentages. 20/26 students who read the text before class (76.92%,) received high grades compared to only 12/24 who read the text after class (50.00%). This difference, 76.92% vs. 50.00%, suggests that reading the text before class makes it more likely one will receive a high grade. However, is the difference statistically significant? To answer that question, Dr. Pradesh needs to use a hypothesis test.

Step 1 Pick a Test

This scenario involves comparing two groups (read text before class vs. read text after class) to see if a difference exists for a variable used to categorize people as good performers vs. poor performers. This calls for a chi-square test of independence.

Step 2 Check the Assumptions

The three assumptions for the chi-square test of independence are the same as for the chi-square goodness-of-fit test:

1. *Random samples.* The samples should be random samples from their populations. The random samples assumption is a robust assumption. If it is violated, one can proceed with the chi-square test of independence, but must be careful about the population to which the results are generalized. In the example, the participants in the read before class vs. read after class study are a random sample from introductory psychology classes who are then randomly assigned to the two groups. The random samples assumption was not violated and the results can be generalized to intro psych classes at this university.
2. *Independence of observations.* The cases in the sample should be independent of each other. That is, the same case can't be in the sample twice. The independence of observations assumption is not robust, so one can't proceed with the planned test if it is violated. In the example, who was in which group was up to chance thanks to random sampling and random assignment, and each student fits in only one cell. The independence of observations assumption was not violated.
3. *Expected frequencies.* All cells must have expected frequencies of at least 5. This assumption is not robust, so the chi-square can't be calculated if the expected frequencies in the cells are small. Until expected frequencies are calculated, this assumption can't be evaluated.

Step 3 List the Hypotheses

The chi-square test of independence tests to see if two variables are independent. Two variables are independent of each other when no relationship exists between them. So, the null hypothesis for a chi-square test of independence states that the two variables are independent of each other *in the population.* The alternative hypothesis for a chi-square test of independence states that, in the population, the two variables are not independent of each other. Notice several things about the alternative hypothesis:

- It doesn't say what direction the relationship goes, whether one variable has a positive or negative impact on the other.
- It doesn't say whether the relationship between the two variables is small or large.
- It just says there is something other than a zero relationship between the two variables in the population.

The simplest way to express the null and alternative hypotheses for the chi-square test of independence is the same way they were expressed for the Pearson *r,* using ρ, the abbreviation for the population value of a correlation, to say whether there is a relationship between the variables:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Step 4 Set the Decision Rule

Setting the decision rule for a chi-square test depends on the alpha level and the number of degrees of freedom (*df*). For the read before class vs. read after class experiment, Dr. Pradesh was willing to have a 5% chance of making a Type I error, so she set $\alpha = .05$.

Determining *df* for a chi-square test of independence depends on how many rows (R) and how many columns (C) are in the contingency table. Look back to Table 15.7:

- There are two rows, one for the students who read the text before class and one for the students who read the text after class, so $R = 2$.
- There are two columns, one for students with high grades and one for students with low grades, so $C = 2$.

Equation 15.4 uses R (the number of rows in the contingency table) and C (the number of columns in the contingency table) to calculate the degrees of freedom for a chi-square test of independence.

Equation 15.4 Formula for Degrees of Freedom (*df*) for a Chi-Square Test of Independence

$$df = (R - 1) \times (C - 1)$$

where df = degrees of freedom
R = number of rows in the contingency table
C = number of columns in the contingency table

For the textbook reading data, there is 1 degree of freedom:

$$\begin{aligned} df &= (R - 1) \times (C - 1) \\ &= (2 - 1) \times (2 - 1) \\ &= 1 \times 1 \\ &= 1 \end{aligned}$$

Now that the degrees of freedom are known, Dr. Pradesh can find the critical value of chi-square in Appendix Table 9. The intersection of the column where $\alpha = .05$ and the row where $df = 1$ shows that $\chi^2_{cv} = 3.841$. If the value of chi-square calculated for the data in the sample is greater than or equal to this critical value, then the results will fall in the rare zone, the null hypothesis is rejected, and the results are called statistically significant. If χ^2 is less than χ^2_{cv}, then χ^2 falls in the common zone, the null hypothesis is not rejected, and the results are called not statistically significant. Here is the decision rule:

- If $\chi^2 \geq 3.841$, reject H_0.
- If $\chi^2 < 3.841$, fail to reject H_0.

Step 5 Calculate the Test Statistic

χ^2 for a chi-square test of independence is calculated with the same formula (Equation 15.3) used to calculate a chi-square goodness-of-fit test. Here it is again.

Equation 15.3 Formula for Calculating Chi-Square (χ^2)

$$\chi^2 = \Sigma \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}}$$

where χ^2 = chi-square value
f_{Observed} = observed frequency for a category
f_{Expected} = expected frequency for a category

To apply the formula, two values are needed for each cell: (1) the observed frequency, f_{Observed}, and (2) the expected frequency, f_{Expected}. The contingency table in Table 15.7 gives the observed frequencies for the cells, so those are known. Finding the expected frequencies takes Equation 15.5.

Equation 15.5 Formula for Calculating Cell Expected Frequencies (f_{Expected})

$$f_{\text{Expected}} = \frac{N_{\text{Row}} \times N_{\text{Column}}}{N}$$

where f_{Expected} = the expected frequency for a cell
N_{Row} = number of cases in the row with that cell
N_{Column} = number of cases in the column with that cell
N = total number of cases in the contingency table

This formula says that the expected frequency for a cell is found by multiplying together the N for the row that contains the cell by the N for the column that contains the cell. This product is then divided by the total sample size.

Dr. Pradesh applies Equation 15.5 to the read before class vs. read after class data. From Table 15.7, the following is known:

- For the first row, $N_{\text{Row}} = 26$. For the second row, $N_{\text{Row}} = 24$.
- For the first column, $N_{\text{Column}} = 32$. For the second column, $N_{\text{Column}} = 18$.
- $N = 50$.

Using Equation 15.5 to calculate the expected frequency for Cell A finds:

$$
\begin{aligned}
f_{\text{Expected}A} &= \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} \\
&= \frac{26 \times 32}{50} \\
&= \frac{832.0000}{50} \\
&= 16.6400 \\
&= 16.64
\end{aligned}
$$

Continuing to use Equation 15.5, the expected frequencies for the other three cells are found in a similar way.

$$
\begin{aligned}
f_{\text{Expected}B} &= \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} \\
&= \frac{26 \times 18}{50} \\
&= \frac{468.0000}{50} \\
&= 9.3600 \\
&= 9.36
\end{aligned}
$$

$$
\begin{aligned}
f_{\text{Expected}C} &= \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} \\
&= \frac{24 \times 32}{50} \\
&= \frac{768.0000}{50} \\
&= 15.3600 \\
&= 15.36
\end{aligned}
$$

$$
\begin{aligned}
f_{\text{Expected}D} &= \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} \\
&= \frac{24 \times 18}{50} \\
&= \frac{432.0000}{50} \\
&= 8.6400 \\
&= 8.64
\end{aligned}
$$

The four expected frequencies are shown in Table 15.8. There are five things to note:

1. Expected frequencies don't have to be whole numbers. Don't be bothered by saying, for example, that 16.64 cases are expected to fall in the read before/get high grades cell.
2. The row frequencies for the *expected* frequencies (Table 15.8) are exactly the same as those found for the *observed* frequencies (Table 15.7).
3. Similarly, the column frequencies for the expected frequencies are the same as the column frequencies for the observed frequencies.
4. Finally, the total number of cases in the expected frequency cells is the same as the total number of cases in the observed frequency cells.
5. All the expected frequencies were at least 5, so Dr. Pradesh now knows that the third assumption was not violated.

TABLE 15.8 Expected Frequencies for the Read Text Before Lecture vs. Read Text After Lecture Study

	High Grade	Low Grade	
Read text before lecture	A 16.64	B 9.36	26
Read text after lecture	C 15.36	D 8.64	24
	32	18	$N = 50$

Equation 15.16 is used to calculate the expected frequencies for a contingency table. Note that the row frequencies and column frequencies found in this table are the same as those in Table 15.7, the contingency table showing observed frequencies for these data.

A Common Question

Q Does one have to use Equation 15.5 to calculate an expected frequency for each cell?

A No. Equation 15.5 is only needed to calculate expected frequencies for as many cells as there are degrees of freedom. For the read before class vs. read after class study, where $df = 1$, once the first cell is known, the other three can be figured out. If the frequency for the first row is 26 and $f_{\text{Expected}} = 16.64$ for Cell A, then Cell B must have an expected frequency of $26 - 16.64$, or 9.36.

Now that the expected frequencies have been found, Equation 15.3 may be used to calculate χ^2. Using the observed frequencies for the read before class vs. read after class data (see Table 15.7) and the expected frequencies (see Table 15.8), the formula finds $\chi^2 = 3.93$:

$$
\begin{aligned}
\chi^2 &= \sum \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}} \\
&= \frac{(20 - 16.64)^2}{16.64} + \frac{(6 - 9.36)^2}{9.36} + \frac{(12 - 15.36)^2}{15.36} + \frac{(12 - 8.64)^2}{8.64} \\
&= \frac{3.3600^2}{16.64} + \frac{-3.3600^2}{9.36} + \frac{-3.3600^2}{15.36} + \frac{3.3600^2}{8.64} \\
&= \frac{11.2896}{16.64} + \frac{11.2896}{9.36} + \frac{11.2896}{15.36} + \frac{11.2896}{8.64} \\
&= 0.6785 + 1.2062 + 0.7350 + 1.3067 \\
&= 3.9264 \\
&= 3.93
\end{aligned}
$$

The value of the test statistic, χ^2, for the text-reading data is 3.93. After some more practice with the first five steps, we'll move on to the sixth step of hypothesis testing, interpretation.

Worked Example 15.2

Imagine an elementary school teacher, Mr. Conaway, who is a follower of Carl Rogers and believes that unconditional positive regard leads to psychological health and positive behavior. He obtains a simple random sample of 40 children from other teachers' classrooms at his school and, through home observation, he categorizes each child as receiving unconditional positive regard (1) frequently, (2) sometimes, or (3) rarely. He then has each child's teacher classify the child as a behavior problem or not.

The contingency table for the relationship between the two variables—the grouping variable of frequency of unconditional positive regard and the dependent variable of being a behavior problem—is shown in Table 15.9:

- Five of the 15 frequent recipients of unconditional positive regard (33.33%) were behavior problems.

TABLE 15.9 Frequencies for the Relationship Between Frequency of Unconditional Positive Regard and Being a Behavior Problem, with Row Frequencies and Column Frequencies Added

	Behavior Problem	Not a Behavior Problem	
Frequently receives unconditional positive regard	A 5	B 10	15
Sometimes receives unconditional positive regard	C 5	D 8	13
Rarely receives unconditional positive regard	E 8	F 4	12
	18	22	$N = 40$

This contingency table appears to show that the likelihood of being a behavior problem increases for those students who receive unconditional positive regard less frequently. A chi-square test of independence will be necessary to determine if the effect is a statistically significant one.

- Five of the 13 sometimes recipients of unconditional positive regard (38.46%) were behavior problems.
- Eight of the 12 students who rarely received unconditional positive regard (66.67%) were behavior problems.

These results are graphed in Figure 15.4, which shows that the likelihood of being a behavior problem increases as the frequency of receiving unconditional positive regard decreases. To determine if this is a real effect—or can be explained by sampling error—Mr. Conaway will need a statistical test.

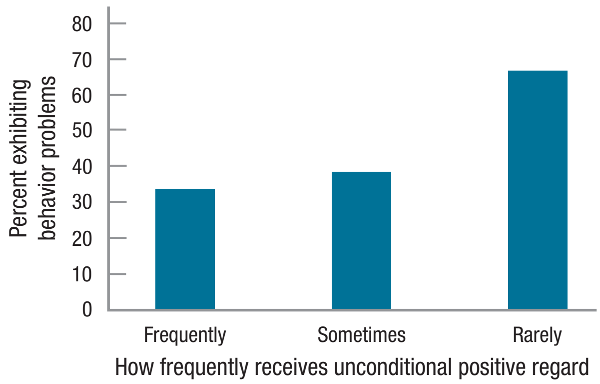

Figure 15.4 Relationship Between Being a Behavior Problem in School and Frequency of Receiving Unconditional Positive Regard These data show that the likelihood of being a behavior problem goes up as the frequency of receiving unconditional positive regard goes down. Whether the effect is a statistically significant one or can be explained by sampling error will require the completion of a chi-square test of independence.

Step 1 Pick a Test. A chi-square test of independence is appropriate to determine if there is a relationship between three categories of one variable (frequency of receiving unconditional positive regard) and two categories of another variable (behavior problem or not a behavior problem) in a sample of elementary school children.

Step 2 Check the Assumptions. There are three assumptions for the chi-square test of independence:

1. *Random samples.* The cases come from a number of classrooms at the school, but it is not stated that the sample is a random sample. The random samples assumption is violated, but it is robust so the teacher can still proceed with the chi-square test of independence. However, he should be careful about generalizing the results beyond this sample.
2. *Independence of observations.* Each child is in the sample only once, so this assumption is not violated.
3. *Adequate expected frequencies.* All cells must have expected frequencies of at least 5. Until the expected frequencies are calculated, one can't be sure that this assumption is met. If it turns out not to be met, one could collapse some

categories and, for example, the frequent recipients of unconditional positive regard could be compared to a combined group of the sometimes and rare recipients of unconditional positive regard.

Step 3 List the Hypotheses. The hypotheses are listed below. The null hypothesis says that no relation exists in the population between the amount of unconditional positive regard one receives and the likelihood of being a behavior problem. The alternative hypothesis states there is some relationship in the population.

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Step 4 Set the Decision Rule. To set the decision rule, a researcher needs to know the alpha and degrees of freedom. Mr. Conaway has set alpha at .05. He uses Equation 15.4 to calculate the degrees of freedom:

$$\begin{aligned} df &= (R - 1) \times (C - 1) \\ &= (3 - 1) \times (2 - 1) \\ &= 2 \times 1 \\ &= 2 \end{aligned}$$

Looking in the table of critical values of chi-square, Appendix Table 9, at the intersection of the column for $\alpha = .05$ and the row for $df = 2$, he finds $\chi^2_{cv} = 5.991$. The decision rule is:

- If $\chi^2 \geq 5.991$, reject H_0.
- If $\chi^2 < 5.991$, fail to reject H_0.

Step 5 Calculate the Test Statistic. The first step in calculating a chi-square test for independence (χ^2) is using Equation 15.5 to calculate expected frequencies. This contingency table has six cells and it is a bit tedious to perform Equation 15.5 six times. So, Mr. Conaway will use a bit of logic and just use Equation 15.5 twice. Why twice? Because the contingency table has 2 degrees of freedom, and once two cells are known, the rest can be calculated. The two to be calculated have to be chosen a little carefully, so Mr. Conaway has picked Cells A and C.

$$\begin{aligned} f_{\text{Expected}A} &= \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} \\ &= \frac{18 \times 15}{40} \\ &= \frac{270.0000}{40} \\ &= 6.7500 \\ &= 6.75 \end{aligned}$$

$$f_{\text{Expected}C} = \frac{N_{\text{Row}} \times N_{\text{Column}}}{N} = \frac{18 \times 13}{40} = \frac{234.0000}{40} = 5.8500 = 5.85$$

Table 15.10 contains the results so far. Now comes the logic.

- If the first row has a total of 15 cases and Cell A has 6.75 of them, then Cell B contains 15 – 6.75 cases, or 8.25 cases.
- The same logic applies to Cell D: 13 – 5.85 = 7.15 cases.
- To find the expected frequency for Cell E, look at the total number of cases in the column, 18. Subtracting the expected frequencies for Cells A and C from this yields the expected frequency of 5.40 for Cell E.
- Finally, for Cell F, subtract Cell E (5.40) from 12, the total number of cases in the row to find 6.60.

TABLE 15.10 Expected Frequencies for Two Cells of the Contingency Table for the Unconditional Positive Regard Frequency/Behavior Problem Data

	Behavior Problem	Not a Behavior Problem	
Frequently receives unconditional positive regard	A 6.75	B	15
Sometimes receives unconditional positive regard	C 5.85	D	13
Rarely receives unconditional positive regard	E	F	12
	18	22	$N = 40$

The row frequencies and column frequencies are set so the expected frequencies for the other cells in this contingency table can be determined once the expected frequencies for two of the cells are known.

The expected frequencies for all six cells are shown in Table 15.11. Note that all the expected frequencies are greater than 5. Mr. Conaway can finally determine that the remaining assumption, that all expected frequencies were large enough, was not violated. The expected frequencies, in conjunction with the observed frequencies (Table 15.9), are needed to calculate the value of the test statistic using Equation 15.3:

$$\chi^2 = \sum \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}}$$

$$= \frac{(5-6.75)^2}{6.75} + \frac{(10-8.25)^2}{8.25} + \frac{(5-5.85)^2}{5.85} + \frac{(8-7.15)^2}{7.15}$$

$$+ \frac{(8-5.40)^2}{5.40} + \frac{(4-6.60)^2}{6.60}$$

$$= \frac{-1.75000^2}{6.75} + \frac{1.75^2}{8.25} + \frac{-0.85^2}{5.85} + \frac{0.85^2}{7.15} + \frac{2.60^2}{5.40} + \frac{-2.60^2}{6.60}$$

$$= \frac{3.0625}{6.75} + \frac{3.0625}{8.25} + \frac{0.7225}{5.85} + \frac{0.7225}{7.15} + \frac{6.7600}{5.40} + \frac{6.7600}{6.60}$$

$$= 0.4537 + 0.3712 + 0.1235 + 0.1010 + 1.2519 + 1.0242$$

$$= 3.3255$$

$$= 3.33$$

TABLE 15.11 Expected Cell Frequencies for the Contingency Table for the Unconditional Positive Regard Frequency/Behavior Problem Data

	Behavior Problem	Not a Behavior Problem	
Frequently receives unconditional positive regard	A 6.75	B 8.25	15
Sometimes receives unconditional positive regard	C 5.85	D 7.15	13
Rarely receives unconditional positive regard	E 5.40	F 6.60	12
	18	22	$N = 40$

This table contains each cell's expected frequency. Note that the row frequencies and column frequencies here are the same as in Table 15.9, the contingency table with observed frequencies.

The chi-square test of independence value for the unconditional positive regard study is 3.33. Now it is time to learn how to interpret the results of a chi-square test of independence.

Practice Problems 15.2

15.06 Given a chi-square test of independence where the explanatory variable has four categories and the dependent variable has three, (a) How many degrees of freedom are there? (b) If $\alpha = .05$, what is χ^2_{cv}?

15.07 Given the cell frequencies below, calculate (a) the row totals and (b) the column totals.

5	10	12
5	8	10

15.08 The row and column totals are shown below. Calculate the expected frequencies for the cells.

A	B	39
C	D	49
28	60	

15.09 Given the information below, calculate χ^2.

$f_{\text{Observed}} = 50$, $f_{\text{Expected}} = 47.22$	$f_{\text{Observed}} = 36$, $f_{\text{Expected}} = 38.78$
$f_{\text{Observed}} = 45$, $f_{\text{Expected}} = 47.77$	$f_{\text{Observed}} = 42$, $f_{\text{Expected}} = 39.23$

15.4 Interpreting the Chi-Square Test of Independence

Now that χ^2 has been calculated for a chi-square test of independence, it is time to move on to the sixth step of hypothesis testing and interpret the results. To interpret a chi-square test of independence, there are three questions to answer: (1) Was the null hypothesis rejected? (2) If so, what is the direction of the difference? (3) How big is the effect?

Let's use the study by the educational psychologist Dr. Pradesh to see how to answer those questions for a chi-square test of independence. Dr. Pradesh's study compared the grades, high vs. low, for 50 students who were randomly assigned to read the text either before or after class.

Step 6 Interpret the Results

Was the Null Hypothesis Rejected?

For Dr. Pradesh's study, alpha was set at .05, there was 1 degree of freedom, and the critical value of chi-square was 3.841. Dr. Pradesh calculated $\chi^2 = 3.93$. Applying the decision rule $3.93 \geq 3.841$, so the null hypothesis is rejected, the alternative hypothesis is accepted, and the results are considered statistically significant. The results can be phrased two ways:

1. There is a relationship between when a student reads the text and what grade he or she receives in the class.
2. A statistically significant difference exists in grades between the students who read the text before class vs. those who read the text after class.

What Is the Direction of the Difference?

Both statements above are true, but neither is satisfying because neither reveals the direction of the difference, whether it is better to read the text before class or after class. With Dr. Pradesh's study, it is easy to determine the direction of the difference because there are only two groups.

By examining the results in the samples, Dr. Pradesh can draw a conclusion about the direction for the populations. Twenty of the 26 students in the read-before-class group (76.92%) received high grades compared to 12 of the 24 students (50.00%) of the read-after-class group. So, Dr. Pradesh concludes, for the larger population of introductory psychology students at this university, that completing the textbook readings before class leads to a better grade outcome than completing the reading after class.

APA Format

The results are reported in APA format for the chi-square test of independence the same way they are for a chi-square goodness-of-fit test. Here are the six pieces of information for the read before class vs. read after class study:

$$\chi^2(1, N = 50) = 3.93, p < .05$$

- χ^2 indicates that the test statistic was a chi-square.
- 1 gives the degrees of freedom.
- $N = 50$ tells how many cases there were.
- 3.93 is the value of the test statistic.
- .05 provides information about the alpha level selected.
- And, $p < .05$ means that the null hypothesis was rejected.

How Big Is the Effect?

Determining the size of the effect for a chi-square test of independence involves transforming the chi-square value into Cramer's *V*, a statistic that is a lot like a Pearson *r*:

- Cramer's *V* ranges from 0 to 1.
- As Cramer's *V* gets closer to 1, the effect is stronger.
- As Cramer's *V* gets closer to 0, the effect is weaker.

The formula for calculating Cramer's *V* is shown in Equation 15.7.

Equation 15.7 Formula for Calculating Cramer's *V*

$$V = \sqrt{\frac{\chi^2}{N \times df_{RC}}}$$

where V = Cramer's V

χ^2 = chi-square value, calculated via Equation 15.3

N = total number of cases in the contingency table

df_{RC} = $(R - 1)$ or $(C - 1)$, whichever is smaller

(R = number of rows in the contingency table;

C = number of columns in the contingency table)

Here's how to calculate *V* for the read before class vs. read after class study. It is already known that $\chi^2 = 3.93$ and $N = 50$. To calculate *V* for this equation, one needs to know the number of rows (2) and the number of columns (2) in the contingency table. The degrees of freedom for Equation 15.7 are either $R - 1$, which is $2 - 1$, or $C - 1$, which is also $2 - 1$, whichever is smaller. Both are the same, so $df_{RC} = 1$.

$$\begin{aligned} V &= \sqrt{\frac{\chi^2}{N \times df_{RC}}} \\ &= \sqrt{\frac{3.93}{50 \times 1}} \\ &= \sqrt{\frac{3.93}{50}} \\ &= \sqrt{0.0786} \\ &= .2804 \\ &= .28 \end{aligned}$$

Cramer's *V* for the text-reading data is .28. When there is only 1 degree of freedom, as here, Cramer's *V* is equivalent to a Pearson *r*. So, Dr. Pradesh could say that there is a .28 correlation between the independent variable (when the text is read) and the dependent variable (high grade or low grade). This is a medium-strength correlation. Guidelines for interpreting Cramer's *V* are given in Table 15.12.

- When $df_{RC} = 1$, Cramer's *V* is interpreted like a Pearson *r* in terms of what is a small, medium, and large effect size.
- As df_{RC} increases, the criteria for considering an effect size meaningful become more lenient.

For the read before class vs. read after class study, the effect, $V = .28$, would be considered a medium one.

TABLE 15.12 Guidelines for Interpreting Cramer's *V*

	Small Effect Size	Medium Effect Size	Large Effect Size
$df_{RC} = 1$	.10	.30	.50
$df_{RC} = 2$	.07	.21	.35
$df_{RC} = 3$	.06	.17	.29
$df_{RC} = 4$	.05	.15	.25
$df_{RC} = 5$	.05	.13	.22

These effect sizes are from Cohen (1988). Note that when $df_{RC} = 1$, the effect size for Cramer's *V* is the same as it is for a Pearson *r*. As degrees of freedom increase, a smaller *V* is interpreted as a larger effect.

Putting It All Together

Here is Dr. Pradesh's four-point interpretation. In it she (1) tells what the study was about, (2) gives the results, (3) explains what they mean, and (4) makes suggestions for future research.

> This study investigated whether completing readings before or after class has an impact on the grade a student receives. Fifty students in introductory psychology classes were randomly assigned to complete the readings before class or after it. Reading the text before attending class resulted in a significantly higher percentage of students receiving high grades: 77% vs. 50%, $\chi^2(1, N = 50) = 3.93$, $p < .05$. The impact on grades of when the reading is completed is more than small, but less than large. To increase the generalizability of results and certainty about the size of the effect, the study should be replicated in courses other than psychology and at other universities.

(Though the results do sound reasonable, please be aware that these data were manufactured for this example.)

Worked Example 15.3 For an example demonstrating how to interpret the chi-square test of independence results that are not statistically significant, a return to Mr. Conaway's study of the relationship between the frequency of receiving unconditional positive regard and being a behavior problem is in order. In that study, Mr. Conaway, an elementary school teacher, "measured" 40 children. The grouping variable was how frequently the child received unconditional positive regard (frequently, sometimes, or rarely) and the dependent variable was whether the child's teacher rated him or her as a behavior problem (yes or no). Alpha was set at .05, there were 2 degrees of freedom, and the critical value of chi-square was 5.991. Using Equation 15.3, Mr. Conaway found $\chi^2 = 3.33$.

Was the null hypothesis rejected? Here is the decision rule:

- If $\chi^2 \geq 5.991$, reject H_0.
- If $\chi^2 < 5.991$, fail to reject H_0.

$3.33 < 5.991$, so the results fall in the common zone (see Figure 15.5), the null hypothesis is not rejected, and the results are called "not statistically significant." There is not enough evidence to conclude that a relationship exists between the frequency with which elementary school children received unconditional positive regard at home and their being a behavior problem. Alternatively, one could say that there is not enough evidence to conclude that the rate of behavior problems differs among children who receive unconditional positive regard frequently, sometimes, or rarely. That is, the conclusion could be stated in either relationship test terms or difference test terms.

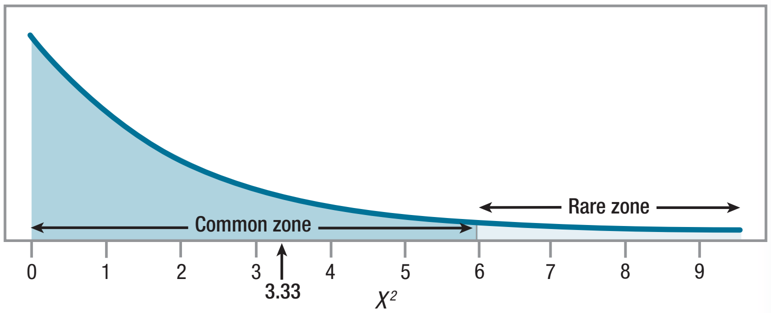

Figure 15.5 Chi-Square Results for the Effect of Frequency of Unconditional Positive Regard on Being a Behavior Problem The observed value of χ^2, 3.33, falls in the common zone of the sampling distribution of chi-square with $df = 2$ and $\alpha = .05$. This means that there is insufficient evidence to justify rejecting the null hypothesis.

Direction of the difference. Because not enough evidence is available to say that a difference exists, there is no need to comment on the direction of the difference.

APA format. In APA format, these results would be written as

$$\chi^2(2, N = 40) = 3.33, p > .05$$

- χ^2 indicates that the statistic calculated is a chi-square value.
- 2 indicates that there are 2 degrees of freedom.
- $N = 40$ tells the sample size.
- 3.33 is the chi-square value found in the sample.
- .05 indicates that alpha was set at .05.
- $p > .05$ indicates that the results fell in the common zone, that the null hypothesis was not rejected.

How big is the effect? Cramer's *V*, the effect size for a chi-square test of independence, should be calculated even when results are not statistically significant. Doing so allows a researcher to evaluate the likelihood of a Type II error. To calculate *V*, one needs the observed chi-square value, sample size, and degrees of freedom (Equation 15.7). For his data, Mr. Conaway already knows $\chi^2 = 3.33$ and $N = 40$. To find df_{RC}, he needs to subtract 1 from the number of rows or 1 from the number of columns, whichever is smaller. There are fewer columns (two) than rows (three), so $df_{RC} = 2 - 1 = 1$. Next, he plugs the numbers into Equation 15.7 to calculate V:

$$
\begin{aligned}
V &= \sqrt{\frac{\chi^2}{N \times df_{RC}}} \\
&= \sqrt{\frac{3.33}{40 \times 1}} \\
&= \sqrt{\frac{3.33}{40}} \\
&= \sqrt{.0833} \\
&= .2886 \\
&= .29
\end{aligned}
$$

According to Table 15.13, $V = .29$ is a medium effect size. This medium effect conflicts with the fact that there is insufficient evidence for an effect. In such situations, the first explanation that comes to mind is the possibility of a Type II error. Perhaps there really is an effect and, due to sampling error, Mr. Conaway failed to find it. This could have happened because the study was underpowered. To determine if an effect does exist, the study should be replicated with a larger sample size.

Putting it all together. Here is Mr. Conaway's four-point interpretation:

> Data from 40 elementary school students were collected to see if the frequency with which they received unconditional positive regard at home was related to their being considered a behavior problem at school. Though the likelihood of being a behavior problem increased as the likelihood of receiving unconditional positive regard decreased, there was not sufficient evidence to conclude that a relationship exists between these two variables [$\chi^2(2, N = 40) = 3.33, p > .05$]. Failure to find an effect could have been due to the small sample size. In order to have adequate power to determine if a relationship exists, it would be a good idea to replicate this study with a larger sample size.

Practice Problems 15.3

15.10 If $\chi^2 = 4.76$, $df = 1$, $N = 36$, and $\alpha = .05$, write the results in APA format.

15.11 If $N = 45$, $\chi^2 = 6.78$, and $df_{RC} = 2$, calculate V.

15.12 A human factors psychologist randomly assigned volunteers to drive on a twisty mountain road in a driving simulator. The control group drove in a normal car with a dashboard. The experimental group drove in a car with a heads-up display, where the dashboard information was projected onto the windshield. The dependent variable was whether the group crashed the car or not. Given the contingency table below, $\chi^2(1, N = 180) = 5.27$, $p < .05$, and Cramer's $V = .17$, write a four-point interpretation:

	Good Outcome: Did Not Crash	Bad Outcome: Crashed
Control Group	42	43
Experimental Group	63	32

15.5 Other Nonparametric Tests

In this brief section, two other nonparametric tests are introduced: the Spearman rank-order correlation coefficient and the Mann–Whitney *U*.

The Spearman Rank-Order Correlation Coefficient

The **Spearman rank-order correlation coefficient** (abbreviated Spearman *r* or r_s) is the nonparametric version of the Pearson correlation coefficient. The Spearman *r* examines the relationship between (1) two ordinal-level variables, (2) one ordinal-level variable and one interval/ratio-level variable, or (3) as a fallback option with two interval/ratio-level variables when the assumptions for the Pearson *r* have been violated.

One can think of the Spearman *r* as a Pearson *r* in which the correlation is calculated for data that are converted to ranks. This means that a researcher can directly apply the hypothesis-testing steps about hypotheses, decision rules, calculation, and interpretation from the Pearson *r* to the Spearman *r*—as long as he or she remembers that ordinal, or ranked, data are being used.

For an example of how the Spearman *r* works, imagine a psychologist who examined the relationship between romantic attractiveness and academic success in high school students. His measure of romantic attractiveness was an interval scale that ranged from 0 (low) to 10 (high). Class rank, an ordinal measure, was the measure of academic success. He obtained both of these measures on a random sample of 24 seniors, and Panel A in Figure 15.6 shows the relationship between them. It appears as if students who are doing less well academically are more romantically attractive.

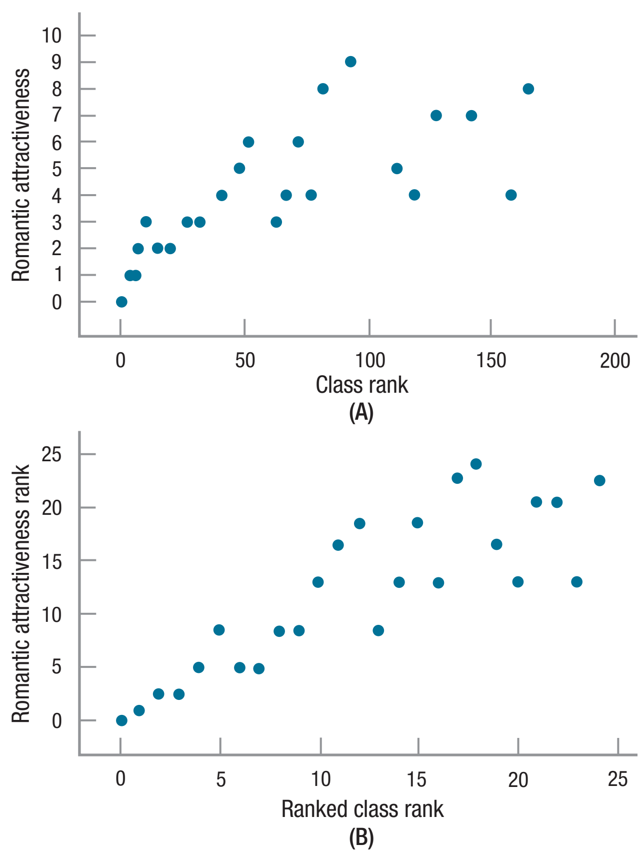

Figure 15.6 Relationship Between Academic Success and Romantic Attractiveness for Unranked Data and Ranked Data Both scatterplots show the relationship between a student's class rank and his or her desirability as a romantic partner. In the top graph (A), the naturally occurring range is used for each variable. In the bottom graph (B), the variables have been converted to ranks. For example, in Panel A, the *X*-axis ranges from 0 to 200, which is sufficient to capture the person with the highest class rank, 165. In Panel B, when the class ranks are ranked from 1 to 24, the person with a rank of 165 now has a score of 24 and the *X*-axis only has to go up to 25.

Class rank is an ordinal-level variable, so a Pearson *r* is not appropriate to use to analyze these data. Instead, a Spearman *r* is the appropriate test. To conduct the Spearman *r*, each variable is assigned a rank from 1 to 24 and the rankings are correlated. Why is each variable assigned a rank from 1 to 24? Because there are 24 cases in the sample. The scatterplot for the ranked data is displayed in Panel B of Figure 15.6, and it shows a very similar pattern to Panel A.

Look at class rank on the *X*-axis in the top panel of Figure 15.6. For these 24 students, it ranges from 1 to 165 and the *X*-axis goes all the way up to 200 to accommodate this. Then look at the *X*-axis in the bottom panel—instead of ranging from 0 to 180, it ranges from 0 to 25. That's because the Spearman *r* takes class rank, which was already an ordinal-level variable, and turns it into pure ranks. The student with the worst class rank in this sample, who had a class rank of 165, now has a score of 24, which indicates the worst class rank in the sample.

Similarly, the *Y*-axis variable, romantic attractiveness, is converted into ranks. In Panel A, the scores ranged from 0 to 10. In Panel B, when the scores are ranked, the highest possible score (for the most attractive person) is now 24, not 10.

The Spearman *r* is just like the Pearson *r*, except that it correlates ranks, not raw scores. The difference between the two can be seen in the axes of the two scatterplots in Figure 15.6. In addition, look closely at the shape of the two scatterplots in Figure 15.6—they are similar, but not exactly the same. Converting raw scores to ranks can change the shape. Sometimes, as here, it makes the relationship appear stronger. Other times it makes the relationship appear weaker. However, making the relationship stronger or weaker isn't the objective. The objective is to use the appropriate statistical test for the data. If correlating two ordinal variables, one ordinal variable with one interval or ratio variable, or two interval and/or ratio variables where it's not possible to proceed with a Pearson *r*, then the Spearman rank-order correlation coefficient is the appropriate test.

The Mann–Whitney *U* Test

The Mann–Whitney *U* test is the nonparametric equivalent of the independent-samples *t* test. It is used to compare two independent populations when the dependent variable is measured at the ordinal level, or as a fallback test when nonrobust assumptions for an independent-samples *t* test have been violated. Because it uses an ordinal outcome variable, the Mann–Whitney *U* test determines if the median of one population is significantly different from the other.

The Mann–Whitney *U* works by combining the two independent samples and then assigning a rank to each case based on its standing in the pooled group. The two samples are then separated, and the sum of the ranks for each of the samples is calculated and used to calculate *U*. If the two populations have similar scores on the dependent variable, then each sample will have a mixture of high ranks and low ranks. However, if the two populations differ, then one will have more high ranks than the other. When this happens, the *U* value will reflect the difference in ranks between the samples, and the null hypothesis of no difference in the medians will be rejected.

Here's an example where the difference between the two samples is very obvious. Table 15.13 shows the speeds, in miles per hour, of the five fastest tennis serves by men and the five fastest serves by women. Clearly, the fastest male tennis serves are faster than the fastest female tennis serves. Let's see how the Mann–Whitney *U* test would show this.

TABLE 15.13 Five Fastest Tennis Serves by Men and by Women

Men	Women
163	129
156	128
155	126
155	126
153	125

These are the speeds, in miles per hour, of the five fastest serves by men and the five fastest serves by women. (From Wikipedia, the source of all contemporary knowledge.)

TABLE 15.14 Fastest Tennis Serves by Men and Women, Combined and Ranked

Speed	Sex	Rank
163	Male	1
156	Male	2
155	Male	3.5
155	Male	3.5
153	Male	5
129	Female	6
128	Female	7
126	Female	8.5
126	Female	8.5
125	Female	10

The data from the two samples in Table 15.13 are combined here in order. It is clear that the five men have faster serves than the five women. And, the sum of the ranks for the two sexes, 15 vs. 40, reflects this.

In Table 15.14, the two samples are combined into one group and the cases are ordered from fastest to slowest. The rank of 1 is given to the fastest serve and the rank of 10 to the slowest serve. There is no overlap in the ranks assigned to men and women. Ranks 1 to 5 belong to the men and ranks 6 through 10 to the women. The sum of the men's ranks, 15, and the sum of the women's ranks, 40, are very different numbers. And, the *U* value that the Mann–Whitney *U* test would calculate for these data would reflect as much.

That's how the Mann–Whitney *U* test works. It combines the two samples into one group, assigns ranks to the whole group, then puts the cases back into their original samples, and compares the ranks. If the ranks are evenly distributed between the two samples, then there is no evidence that the populations differ. But if one sample has a lot of high ranks and the other a lot of low ranks, then the difference between the populations is statistically significant.

Practice Problems 15.4

For these practice questions, select the appropriate statistical test from the following: Pearson r, Spearman r, independent-samples t test, or Mann–Whitney U test.

15.13 A psychologist wanted to examine the relationship, in adolescents, between days of drug use in the past year and IQ score. Days of drug use turned out to be very positively skewed.

15.14 The psychologist then divided the adolescents into two groups, those with no days of drug use over the past year and those with one or more days. She wanted to compare the two groups in terms of IQ.

15.15 A sports economist wanted to know if countries that invested more in their sports programs did better in the Olympics. He divided countries into those that competed in 10 or more sports in the summer Olympics and those that competed in fewer than 10 sports. For each country, he found its medal count rank, and he divided this by the number of sports in which the country competed to obtain his dependent variable, rank per events entered.

Application Demonstration

To see chi-square in action, let's analyze some data from a real-life study. A group of researchers was interested in determining if survival after surgery was influenced by a person's nutritional status before surgery. They took almost 400 patients about to undergo surgery for kidney cancer and classified them as malnourished or not. Surgery was then performed and the researchers followed the patients for three years to see how many lived and how many died (Morgan et al., 2011).

Step 1 Pick a Test. Though the researchers used more complex statistical techniques, we'll apply a chi-square test of independence to the data:

- The grouping variable, nutritional status, is nominal with two categories: (1) malnourished or (2) not malnourished.
- The dependent variable, survival three years later, is nominal with two categories: (1) alive or (2) dead.

Table 15.15 displays the results as a contingency table. The table shows cell frequencies, as well as row percentages and column percentages. Eighty-five of the 369 patients (23%) were classified as malnourished and 76 of the 369 patients (21%) died within three years.

TABLE 15.15 Contingency Table for the Effect of Nutritional Status on Mortality

	Alive	Dead	
Malnourished	A 50 (58.82%)	B 35 (41.18%)	85 (23.04%)
Not Malnourished	C 243 (85.56%)	D 41 (14.44%)	284 (76.96%)
	293 (79.40%)	76 (20.60%)	$N = 369$

The percentages in parentheses are row percentages and column percentages. The percentages in the cells are for that row only. The mortality rate for the malnourished patients (41%) was almost 3 times higher than for the patients who were not malnourished (14%). Data from Morgan et al., 2011.

The percentages in the cells suggest that being malnourished prior to surgery diminishes one's chance of surviving for three years after surgery: 35 of the 85 malnourished patients died—that's 41%—compared to only 41 of the 284 not malnourished patients, or 14%. In other words, there appears to be a relationship between malnourishment status and mortality status. But, is this a statistically significant relationship? To answer that question about these nominal-level variables, a chi-square test of independence is required.

Step 2 Check the Assumptions.

- *Random samples.* The sample was not a random sample, leaving uncertainty about the population to which the results can be generalized. The sample is large; it might be representative of patients with kidney cancer at this hospital.

Whether these patients are like patients at other hospitals or in other states and countries is unknown. The random samples assumption is robust to violation, so it's OK to proceed with the planned chi-square test of independence.

- *Independence of observations.* No evidence exists that the patients influenced each other. There's no reason to believe any patients were in the study twice. The independence of observations assumption does not appear to have been violated.
- *Adequate expected frequencies.* All cells must have expected frequencies of at least 5. This assumption can't be assessed until the expected frequencies are calculated.

Step 3 List the Hypotheses. The null hypothesis says that malnutrition status and survival are independent and the alternative hypothesis states that they have some relationship:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Step 4 Set the Decision Rule. The alpha level is set at .05, as usual. The degrees of freedom are calculated using Equation 15.4:

$$\begin{aligned} df &= (R-1) \times (C-1) \\ &= (2-1) \times (2-1) \\ &= 1 \times 1 \\ &= 1 \end{aligned}$$

Use Appendix Table 9 to find the critical value of chi-square, $\chi^2_{cv} = 3.841$. Here is the decision rule:

- If $\chi^2 \geq 3.841$, reject H_0 and say the results are statistically significant.
- If $\chi^2 < 3.841$, fail to reject H_0 and say the results are not statistically significant.

Step 5 Calculate the Test Statistic. To save time, the expected frequencies for the cells are shown in Table 15.16. Equation 15.5 was used to calculate the expected frequency for Cell A and then logic was utilized to find the expected frequencies for the other three cells. None of the expected frequencies is less than 5, so it's now known that the third assumption was not violated.

TABLE 15.16 Expected Frequencies for the Effect of Nutritional Status on Mortality

	Alive	Dead	
Malnourished	A 67.50	B 17.50	85
Not Malnourished	C 225.50	D 58.50	284
	293	76	$N = 369$

Equation 15.6 was used to find the expected frequency for Cell A. There is 1 degree of freedom in this contingency table, so once one cell frequency is known, the other cell frequencies can be determined from the column frequencies and row frequencies.

Using Equation 15.3, the chi-square value is calculated as 28.63:

$$\chi^2 = \sum \frac{\left(f_{\text{Observed}} - f_{\text{Expected}}\right)^2}{f_{\text{Expected}}}$$

$$= \frac{(50 - 67.50)^2}{67.50} + \frac{(35 - 17.50)^2}{17.50} + \frac{(243 - 225.50)^2}{225.50} + \frac{(41 - 58.50)^2}{58.50}$$

$$\text{Step 1} = \frac{-17.5000^2}{67.50} + \frac{17.5000^2}{17.50} + \frac{17.5000^2}{225.50} + \frac{-17.5000^2}{58.50}$$

$$\text{Step 2} = \frac{306.2500}{67.50} + \frac{306.2500}{17.50} + \frac{306.2500}{225.50} + \frac{306.2500}{58.50}$$

$$\text{Step 3} = 4.5370 + 17.5000 + 1.3581 + 5.2350$$

$$\text{Step 4} = 28.6301$$

$$= 28.63$$

Step 6 Interpret the Results. The calculated value of chi-square (28.63) is greater than the critical value (3.841), so the null hypothesis is rejected and the results are written in APA format as $\chi^2(1, N = 369) = 28.63, p < .05$. A statistically significant difference exists between the mortality rate for malnourished patients and patients who were not malnourished. There are only two groups, so we know the direction of the difference: the mortality rate is higher for malnourished patients than for patients who were not malnourished, 41% vs. 14%.

Using Equation 15.7, Cramer's V is calculated as .28:

$$V = \sqrt{\frac{\chi^2}{N \times c}}$$

$$= \sqrt{\frac{28.63}{369 \times 1}}$$

$$= \sqrt{\frac{28.63}{369}}$$

$$= \sqrt{0.776}$$

$$= 0.2786$$

$$= .28$$

By Cohen's standards (1988), this is a medium effect. However, Cohen's standards are meant as guidelines. A change in the mortality rate from 14% to 41%—almost a threefold increase—seems like a strong effect, especially as death is the outcome.

Putting it all together.

Data were analyzed from a study in which researchers compared the mortality rate for kidney cancer patients who were and were not malnourished prior to surgery. Three years after surgery, 41% of the malnourished patients had died compared to only 14% of the patients who were not malnourished. This almost threefold increase in mortality was a statistically significant difference [$\chi^2(1, N = 369) = 28.63, p < .05$]. The effect seems to be a strong one. For kidney cancer patients undergoing surgery, being malnourished prior to surgery is associated with almost a threefold increase in the risk of dying in the next three years. In future studies, it would be important to assess the reason for the malnourishment. If it resulted from having a more aggressive form of cancer, it might be the cancer, not the malnourishment, that led to the increase in mortality.

DIY

What can grocery shopping tell us about gender equality? Do men do more shopping than women? Are women more likely than men to be accompanied by children?

Make a data collection sheet like the one below. Then go to a grocery store during the day and put a tally mark in Cell A for every woman you see shopping with children. A mark goes in Cell B for every woman shopping by herself. Similarly, use Cell C to keep track of men shopping with children and D for men shopping alone.

Add together Cells A and B, and Cells C and D. Use a chi-square goodness-of-fit test with expected percentages of 50% and 50% to see if men and women share shopping responsibility equally. Do the results say something about contemporary gender roles?

	Shopping with Children	Shopping without Children
Female Shoppers	A	B
Male Shoppers	C	D

Next, let's see if one sex is more likely to shop with children. To do so, use the values in all four cells as the observed frequencies for a chi-square test of independence.

SUMMARY

Differentiate parametric tests from nonparametric tests.

- Parametric tests have interval or ratio outcome variables that must be normally distributed. Nonparametric tests can be done on nominal or ordinal outcome variables and don't require a normal distribution. Less restricted by assumptions, nonparametric tests often have less power than parametric tests.

Calculate and interpret a chi-square goodness-of-fit test.

- The chi-square goodness-of-fit test is a nonparametric, single-sample test for use with a nominal outcome variable. It compares the observed frequencies to the expected frequencies to calculate a chi-square value. Interpretation concerns whether the null hypothesis was rejected and, if so, what the direction of the difference was.

Calculate and interpret a chi-square test of independence.

- A chi-square test of independence determines whether two or more populations differ on a nominal-level outcome variable. Data for it are arrayed in a contingency table, which shows the degree to which the outcome variable is determined by the explanatory variable. The chi-square value is calculated from the discrepancy between the observed frequencies and the expected frequencies. In the interpretation, three questions are addressed: (1) whether the null hypothesis was rejected, (2) what the direction of the difference is, and (3) how big the effect is.

Know when to use a Spearman rank-order correlation coefficient and a Mann–Whitney *U* test.

- The Spearman rank-order correlation coefficient, used with ordinal-level data, is the nonparametric alternative to Pearson *r*, the correlation coefficient used with interval/ratio-level data. The Mann–Whitney *U* test, used with an ordinal outcome variable, is the nonparametric alternative to an independent-samples *t* test.

KEY TERMS

chi-square goodness-of-fit test – a nonparametric, single-sample test used to compare the distribution of a categorical (nominal- or ordinal-level) outcome variable in a sample to a known population value.

chi-square test of independence – a nonparametric test used to determine whether two or more populations of cases differ on a categorical (nominal- or ordinal-level) outcome variable.

contingency table – a table showing the degree to which a case's value on the outcome variable depends on its category on the explanatory variable.

Mann–Whitney U test – a nonparametric test used to compare two independent samples on an ordinal-level outcome variable that utilizes ranking.

nonparametric test – a statistical test for use with nominal- or ordinal-level outcome variables, and for which assumptions about the shape of the population don't have to be met.

parametric test – a statistical test for use with interval- or ratio-level outcome variables, and for which assumptions about the shape of the population must be met.

single-sample test – a statistical test used to compare the results in a sample to a known population value.

Spearman rank-order correlation coefficient – a nonparametric test that examines the relationship between two ordinal-level variables or one ordinal and an interval/ratio variable.

CHAPTER EXERCISES

Review Your Knowledge

15.01 The statistical tests in previous chapters all had ____-level or ____-level outcome variables.

15.02 ____ tests are for nominal-level or ____-level outcome variables.

15.03 The independent-samples *t* test is an example of a ____ test.

15.04 Parametric tests assume that the outcome variable is ____ in the population.

15.05 If a researcher violates a nonrobust assumption for a parametric test, then a nonparametric test can be used as a ____ test.

15.06 Nonparametric tests are less restricted by ____ than are ____ tests.

15.07 Nonparametric tests usually have less ____ than ____ tests.

15.08 The chi-square goodness-of-fit test is a ____-sample test used with a ____ outcome variable.

15.09 The chi-square goodness-of-fit test sees if the difference between what is ____ and what is ____ can be explained by sampling error.

15.10 The abbreviation for chi-square is ____, where the Greek letter is pronounced ____.

15.11 The sample in a chi-square goodness-of-fit test should be a ____ from the population.

15.12 There should be at least ____ cases in each cell in a chi-square goodness-of-fit test.

15.13 The chi-square goodness-of-fit test compares the distribution of the outcome variable in the ____ to the specified distribution in the ____.

15.14 The critical value of chi-square is abbreviated ____.

15.15 If the ____ lands in the ____ of the sampling distribution of chi-square, then the null hypothesis is rejected.

15.16 The degrees of freedom for a chi-square goodness-of-fit test depend on the number of ____ in the study.

15.17 If the differences between the observed frequencies and the expected frequencies in a chi-square goodness-of-fit test are small, the null hypothesis is ____.

15.18 It is *impossible* / *possible* for expected frequencies in a chi-square goodness-of-fit test to be fractional numbers.

15.19 The expected frequencies in a chi-square goodness-of-fit test add up to ____.

15.20 The first step in calculating a chi-square is to ____ the expected frequencies from the ____.

15.21 In APA format for a chi-square goodness-of-fit test, the first number inside the parentheses represents the ____.

15.22 In APA format, if the results are written "$p > .05$," this means the results fell in the ____ zone.

15.23 The direction of the results for a chi-square goodness-of-fit test needs to be determined if the results are ____.

15.24 The chi-square test of independence differs from the chi-square goodness-of-fit test in terms of the number of ____ in the test.

15.25 The chi-square test of independence can be conceptualized as a ____ test or a ____ test.

15.26 To conduct a chi-square test of independence, construct a ____ table that cross-tabulates the values of the ____ with the levels of the ____.

15.27 In a contingency table, each ____ is placed in one, and only one, ____.

15.28 The assumptions for the chi-square test of independence are the same as for the ____.

15.29 The null hypothesis for the chi-square test of independence states that the ____ and the ____ are ____.

15.30 The hypotheses for a chi-square test of independence are stated the same way as they are for a ____.

15.31 The degrees of freedom for a chi-square test of independence depend on the number of ____ and the number of ____.

15.32 The formula for the chi-square test of independence *is similar to* / *not similar to* the formula for the chi-square goodness-of-fit test.

15.33 The number of cases in the rows in a chi-square test of independence add up to ____%.

15.34 The sum of the expected frequencies for the rows in a chi-square test of independence are the same as the sum of the observed frequencies for the ____.

15.35 In a Spearman rank-order correlation coefficient, the ____ of the two variables are correlated.

15.36 The ____ is the nonparametric alternative to the independent-samples t test.

15.37 The Mann–Whitney U test takes two independent ____, combines them into one ____, and then assigns a ____ to each case in the combined group.

15.38 The Mann–Whitney U test compares the ranks of the cases in one ____ to the ____ of the cases in the other sample.

Apply Your Knowledge

Pick the correct test from the following: the chi-square goodness-of-fit test, chi-square test of independence, Spearman rank-order correlation coefficient, and Mann–Whitney U test.

15.39 A high school principal developed a theory that caffeinated sodas cause more burping than decaffeinated sodas. She obtained a large sample of students and randomly assigned them to drink sodas with or without caffeine.

She then waited 15 minutes and classified each student as having burped or not having burped during that time. What test should she use?

15.40 The principal also kept track of how many burps each student produced. For both the caffeinated and de-caffeinated groups, the number of burps was extremely positively skewed. What statistical test should she use to see if the number of burps differs between the two groups?

15.41 A social psychologist assigned ranks to all the second-grade boys in a school based on how much the other boys liked them. He then did the same thing for all the boys again, but based the ranks on how much the girls liked them. What test should the psychologist conduct to see if there's an association between how boys and how girls view boys?

15.42 A statistics teacher looked out at her class of 32 students and noted that 21 of them were female. Assuming there are equal numbers of males and females in the world, what test should this teacher use to see if her class is overpopulated with women?

Checking the assumptions for a chi-square goodness-of-fit test or a chi-square test of independence

15.43 A sociologist planned to study patterns of criminality in small towns. She drew a random sample of 50 small towns from all across America. Before conducting her study, she wanted to make sure that her sample was representative of small towns. From the FBI, she learned that 2% of small towns have experienced a murder over the past 10 years. Can she use a chi-square goodness-of-fit test to compare the percentage of small towns that had experienced a murder to the national percentage?

15.44 An addictions researcher wants to see if male and female alcoholics differ in the type of alcohol they consume. She goes to a large alcohol detox facility, gets a sample of men and a sample of women, and checks each person's chart to find the beverage of choice. She classifies the beverages as (a) wine, (b) beer, or (c) hard liquor. The table that follows shows the expected frequencies. Can the researcher use a chi-square test of independence to analyze her data?

	Beer	Wine	Hard Liquor
Men	9.07	8.53	6.40
Women	7.93	7.47	5.60

Stating the hypotheses for chi-square tests

15.45 State the null and alternative hypotheses for Exercise 15.43.

15.46 State the null and alternative hypotheses for Exercise 15.44.

Finding degrees of freedom for chi-square goodness-of-fit tests

15.47 Given this matrix, how many degrees of freedom are there for this chi-square goodness-of-fit test?

15.48 Given this matrix, how many degrees of freedom are there for this chi-square goodness-of-fit test?

15.49 Given this matrix, how many degrees of freedom are there for this chi-square test of independence?

15.50 Given this matrix, how many degrees of freedom are there for this chi-square test of independence?

Finding χ^2_{cv} and setting the decision rule. Use $\alpha = .05$.

15.51 (a) If $df = 2$, what is χ^2_{cv}? (b) What is the decision rule?

15.52 (a) If $df = 4$, what is χ^2_{cv}? (b) What is the decision rule?

Calculating expected frequencies for a chi-square goodness-of-fit test

15.53 Given $N = 97$ and the expected percentages below, find the expected frequencies:

A 63%	B 37%

15.54 Given $N = 182$ and the expected percentages below, find the expected frequencies:

A 14%	B 38%	C 48%

Calculating χ^2 for a chi-square goodness-of-fit test

15.55 Given the information in the following matrix, calculate χ^2:

f_{Observed}	9	14	18	$\Sigma = 41$
f_{Expected}	6.97	15.58	18.45	$\Sigma = 41.00$

15.56 Given the information in the following matrix, calculate χ^2:

f_{Observed}	28	23	20	$\Sigma = 71$
f_{Expected}	33.37	19.98	17.65	$\Sigma = 71.00$

15.57 Given the information in the following matrix, calculate χ^2:

f_{Observed}	40	50
$f_{\text{\% Expected}}$	35%	65%

15.58 Given the information in the following matrix, calculate χ^2:

f_{Observed}	22	37
$f_{\text{\% Expected}}$	42%	58%

Using APA format with $\alpha = .05$

15.59 If $df = 1$, $N = 43$, $\chi^2 = 4.81$, and $\chi^2_{cv} = 3.841$, write the results in APA format.

15.60 If $df = 3$, $N = 78$, $\chi^2 = 5.99$, and $\chi^2_{cv} = 7.815$, write the results in APA format.

15.61 If $df = 4$, $N = 55$, and $\chi^2 = 9.49$, write the results in APA format.

15.62 If $df = 5$, $N = 234$, and $\chi^2 = 3.85$, write the results in APA format.

Determining the direction of the difference for a statistically significant chi-square goodness-of-fit test

15.63 Given the information below, determine the direction of the results:

	Category I	Category II
f_{Observed}	24	26
f_{Expected}	8.77	41.23

15.64 Given the information below, determine the direction of the results:

	Category I	Category II
f_{Observed}	37	23
f_{Expected}	49.23	10.77

Interpreting chi-square goodness-of-fit test

15.65 Dr. Lowry investigated whether fortunetellers have extrasensory perception. He obtained a random sample of 124 fortunetellers. For each one, a card was randomly selected from a well-shuffled 52-card deck and placed face down. Then the fortuneteller had to determine if the card was red or black. For each fortuneteller, Dr. Lowry recorded if he or she guessed correctly. If that person had no ESP ability, then he or she had a 50% chance of guessing correctly. Dr. Lowry found that 72 of the 124 fortunetellers (58.06%) correctly guessed the color of the card and 52 (41.94%) were wrong. The table below shows the observed and expected frequencies. The chi-square goodness-of-fit test results were $\chi^2(1, N = 124) = 3.23$, $p > .05$. Write a four-point interpretation.

	Guessed Correctly	Category Incorrectly
f_{Observed}	72	52
f_{Expected}	62.00	62.00

15.66 According to the U.S. government, 68% of American adults weigh more than they should. A dietician, Dr. Christiansen, maintained a theory that drinking 8 or more glasses of water a day was associated with not being overweight. She obtained a random sample of 580 American adults who consumed 8 or more glasses of water a day, weighed each adult, and classified each one as overweight or not. The table below shows the observed frequencies and the expected frequencies. The chi-square goodness-of-fit test results were $\chi^2(1, N = 580) = 15.62, p < .05$. Write a four-point interpretation.

	Not Overweight	Overweight
$f_{Observed}$	230	350
$f_{Expected}$	185.60	394.40

Completing all six steps of hypothesis testing for a chi-square goodness-of-fit test

15.67 A consumer psychologist, Dr. Wessells, obtained a random sample of 880 people who had strong preferences as to what cola they preferred. He gave each one, individually, a taste test in which each participant tasted four different colas and then had to pick his or her favorite. Dr. Wessells wondered whether consumers would be able to tell the difference. If they couldn't, they would have a 25% chance of being right and a 75% chance of being wrong. It turned out that 279 participants correctly chose their preferred cola and 601 chose incorrectly. Use hypothesis testing to decide if all colas taste the same, even to people who have strong preferences.

15.68 Dr. Constantinople, a cancer physician, wondered whether wearing a hat protected one against skin cancer. From prior research, she knew that about 1% of adults develop skin cancer every year. She obtained a random sample of 2,040 adults who wore hats and followed them for a year. During that year, 14 developed skin cancer and 2,026 did not. Use hypothesis testing to determine if wearing a hat is associated with a change in the risk of developing skin cancer.

Generating contingency tables

15.69 A researcher compared people in their 20s, 40s, and 60s in terms of whether they supported gay marriage or not. Support was measured as a "yes" or "no." Create a labeled contingency table that could be used to cross-tabulate the results.

15.70 An experimental group and a control group were compared in terms of whether they recovered from an illness in 5 or fewer days, 6 to 10 days, or more than 10 days. Make a labeled contingency table that could be used to cross-tabulate the results.

Finding the number of cases per row and column for a chi-square test of independence

15.71 Given the contingency table below, find the number of cases for each row and column:

	Outcome A	Outcome B
Group I	34	86
Group II	67	113

15.72 Given the contingency table below, find the number of cases for each row and column:

	Outcome A	Outcome B
Group I	28	12
Group II	32	18

Calculating expected frequencies

15.73 Given $N = 72$ and the information in the following contingency table, calculate the expected frequencies for the cells:

	Outcome 1	Outcome 2	
Group I	A	B	34
Group II	C	D	38
	30	42	

15.74 Given $N = 123$ and the information in the following contingency table, calculate the expected frequencies for the cells:

	Outcome 1	Outcome 2	
Group I	A	B	34
Group II	C	D	32
Group III	E	F	57
	27	96	

Calculating χ^2

15.75 Given the information below, calculate χ^2:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	28	12
Group II	12	12

	Expected Frequencies	
	Outcome A	Outcome B
Group I	25.00	15.00
Group II	15.00	9.00

15.76 Given the information below, calculate χ^2:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	35	55
Group II	67	40

	Expected Frequencies	
	Outcome A	Outcome B
Group I	46.60	43.40
Group II	55.40	51.60

15.77 Given the information below, calculate χ^2:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	14	21
Group II	18	18

15.78 Given the information below, calculate χ^2:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	6	18
Group II	30	10

Types of error

15.79 Given $\chi^2(1, N = 45) = 4.92$, $p < .05$, which type of error does the researcher need to worry about?

15.80 Given $\chi^2(1, N = 88) = 3.45$, $p > .05$, which type of error does the researcher need to worry about?

Determining the direction of the difference

15.81 The chi-square yielded statistically significant results. Given the information below, determine the direction of the difference:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	6	18
Group II	30	10

	Expected Frequencies	
	Outcome A	Outcome B
Group I	13.50	10.50
Group II	22.50	17.50

15.82 The chi-square yielded statistically significant results. Given the information below, determine the direction of the difference:

	Observed Frequencies	
	Outcome A	Outcome B
Group I	12	19
Group II	17	9

	Expected Frequencies	
	Outcome A	Outcome B
Group I	15.77	15.23
Group II	13.23	12.77

Calculating and interpreting Cramer's V

15.83 Given $\chi^2(3, N = 78) = 8.92$, $p < .05$ where $R = 4$ and $C = 2$, (a) calculate *V*; (b) classify the effect as small, medium, or large; and (c) determine if the researcher should worry about Type II error.

15.84 Given $\chi^2(1, N = 102) = 3.53$, $p > .05$ where $R = 1$ and $C = 1$, (a) calculate *V*; (b) classify the effect as small, medium, or large; and (c) determine if the researcher should worry about Type II error.

Interpreting a chi-square test of independence

15.85 Tennis players are more likely to win points when they are serving. A tennis journalist wondered whether the server's advantage diminished as a rally lasted longer. She randomly selected 209 points from all the matches played at the major tennis tournaments during a year. For each point, she recorded whether the rally was short (the ball went back and forth no more than twice before the point was decided) or the rally was long (the ball went back and forth more than two times). She also recorded whether the server won the point or not. The table below shows the results. She found $\chi^2(1, N = 209) = 3.93$, $p < .05$ and calculated Cramer's $V = .14$. Write a four-point interpretation.

	Observed Frequencies	
	Server Wins	Server Loses
Short rally	66	41
Long rally	49	53

	Expected Frequencies	
	Server Wins	Server Loses
Short rally	58.88	48.12
Long rally	56.12	45.88

15.86 A college dean noticed that political science seemed to attract more male majors than psychology. She wondered if her observation were true. So, she obtained a random sample from each major and recorded the sex of each student. The table below shows her results. The dean found $\chi^2(1, N = 55) = 3.66$, $p > .05$; $V = .26$. Write a four-point interpretation.

	Observed Frequencies	
	Female	Male
Psychology	23	13
Political science	7	12

	Expected Frequencies	
	Female	Male
Psychology	19.64	16.36
Political science	10.36	8.64

Completing all six steps of hypothesis testing

15.87 A surgeon decided to compare sutures and staples in closing incisions. When he finished operating on a patient, he flipped a coin in order to determine if he would use sutures or staples to close the wound. Six months later, he called each patient and found out whether he or she was satisfied, yes vs. no, with how the scar looked. The results appear below. Use hypothesis testing to determine if there's a difference in satisfaction between the two techniques.

	Observed Frequencies	
	Not Satisfied	Satisfied
Sutures	8	13
Staples	9	12

15.88 A political scientist developed a theory that after an election, supporters of the losing candidate removed the bumper stickers from their cars faster than did supporters of the winning candidate. The day before a presidential election, he randomly selected parking lots, and at each selected parking lot, he randomly selected one car with a bumper sticker and recorded which candidate it supported. The day after the election, he followed the same procedure with a new sample of randomly selected parking lots. For both days, he then classified the bumper stickers as supporting the winning or losing candidate. Below are the results. Use hypothesis testing to see if a difference exists between how winners and losers behave.

	Observed Frequencies	
	Winner	Loser
Before	34	32
After	28	10

Expand Your Knowledge

15.89 Complete the following contingency table. If it can't be done, explain why.

12		20
		30
26	24	

15.90 Complete the following contingency table. If it can't be done, explain why.

9		7	24
			17
3		2	10
18	18	15	

15.91 Complete the following contingency table. If it can't be done, explain why.

9		18
		36
		42
46	50	

15.92 Is it possible, for the contingency table below, where $N = 24$, for f_{Observed} to equal f_{Expected} for each cell?

		$N = 24$

15.93 If $f_{\text{Observed}} = f_{\text{Expected}}$ for every cell, what does that mean for the chi-square value and for the null hypothesis?

15.94 Is there any value for degrees of freedom from 1 to 13 that cannot occur for a chi-square test of independence?

SPSS

Data for a chi-square test of independence in SPSS are entered with each variable in a separate column and each case in a row. Figure 15.7 shows the data for the text-reading study. The second column, text, has a 1 if the student read the text before class and a 2 if the text was read after class. In the third column, grade, a 1 means a high grade (A or B) and a 2 means a low grade (C, D, or F).

	V1	text	grade
1	1	1	2
2	2	2	2
3	3	2	2
4	4	2	2
5	5	2	2
6	6	1	1
7	7	1	1
8	8	1	2
9	9	1	1
10	10	1	1
11	11	1	1
12	12	1	1
13	13	1	1

Figure 15.7 Data Entry for Chi-Square Test of Independence in SPSS Data for a chi-square test of independence are entered in SPSS with each variable in a separate column. (Source: SPSS)

In SPSS, the chi-square test of independence is called "Crosstabs," which is short for a cross-tabulation (contingency) table. Figure 15.8 shows that Crosstabs is located under "Analyze" and "Descriptive Statistics."

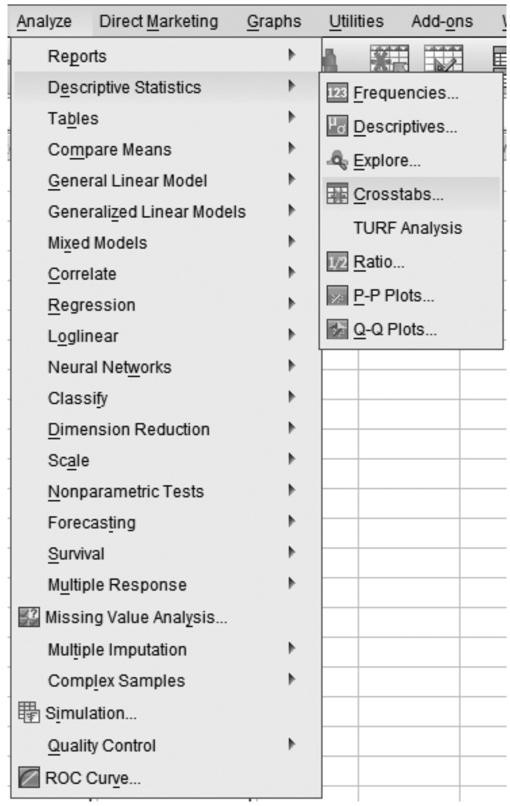

Figure 15.8 The Command for Crosstabs in SPSS "Crosstabs" is the SPSS command for a chi-square test of independence. It is found in the menu under "Analyze" and "Descriptive Statistics." (Source: SPSS)

Clicking on the Crosstabs command opens up the menu seen in Figure 15.9. Note that the variable "text" has already been sent over to be the row variable and the variable "grade" is getting ready to be sent to be the column variable. Be consistent—make the explanatory variable the row variable and the dependent variable the column variable.

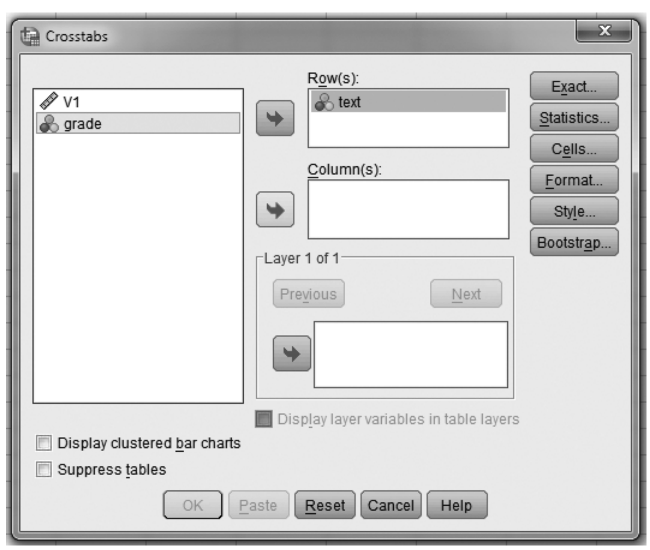

Figure 15.9 Menu for the Crosstabs Command Once this menu is active, one variable is designated as the row variable and a second variable as the column variable. (Source: SPSS)

Clicking on the “Statistics” button in the upper-right-hand corner of Figure 15.9 opens the menu seen in **Figure 15.10**. Note that the boxes for “Chi-Square” and “Phi and Cramer’s V” have been checked. Then click on the “Continue” button in Figure 15.10 and the “OK” button in Figure 15.9.

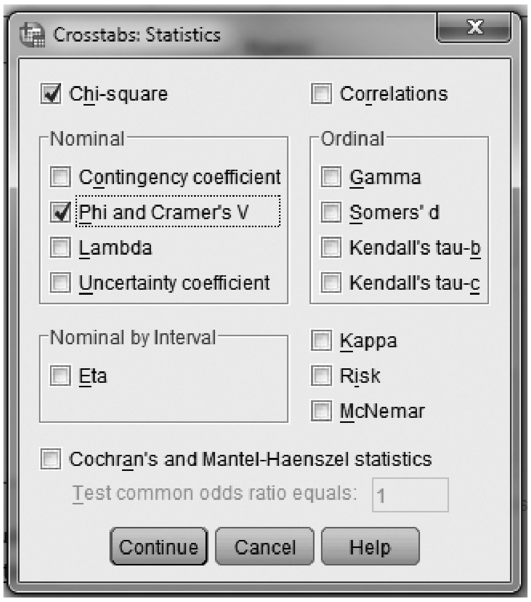

Figure 15.10 Selecting Chi-square and Cramer’s *V* Once the desired statistics have been selected, click on the “Continue” button. (Source: SPSS)

Once the statistics have been selected, clicking on the OK button on the cross-tabs menu generates the output seen in **Figure 15.11**. The first box in the output (A) is the contingency table. The second box (B) gives the chi-square value, the degrees of freedom, and the exact significance level. SPSS calls chi-square “Pearson chi-square” and reports it on the first row as 3.926. Note that the subscript “a” means that all cells had expected frequencies greater than 5. The degrees of freedom are reported as 1 and the significance level as .048. If alpha is set at .05, then the results are statistically significant as long as the significance level is $\leq$.05. Finally, Cramer’s V is reported in the third box of the printout (C) as equal to .280.

text * grade Crosstabulation

Count

		grade		Total
		1	2	
text	1	20	6	26
	2	12	12	24
Total		32	18	50

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.926[a]	1	.048		
Continuity Correction[b]	2.845	1	.092		
Likelihood Ratio	3.980	1	.046		
Fisher's Exact Test				.077	.045
Linear-by-Linear Association	3.848	1	.050		
N of Valid Cases	50				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.64.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.280	.048
	Cramer's V	.280	.048
N of Valid Cases		50	

Figure 15.11 Output for Crosstabs in SPSS The chi-square value, here called "Pearson Chi-Square," can be found on the first line of the second box. Cramer's *V* is found on the second line of the third box. (Source: SPSS)