

Chi-Square: A Test Commonly Used for Nominal-Level Measures

Choosing the chi-square distribution

When is the Chi-Square Distribution Appropriate?

What are Degrees of Freedom?

How Do Degrees of Freedom Affect the Distribution?

Calculating the chi-square statistic

How is the Chi-Square Statistic Calculated?

How Does One Interpret a Chi-Square Statistic?

Substantive examples using the chi-square test

How is the Chi-Square Test Carried Out When There Is Only One Nominal-Scale Measure?

How is the Chi-Square Test Applied in the Case of the Relationship Between Two Nominal-Scale Variables?

Can the Chi-Square Test Be Used to Examine the Relationship Between Ordinal-Level Variables?

THE BINOMIAL TEST provides a good introduction to the problem of statistical inference because it examines relatively simple statistical decisions. Using the binomial test, we illustrated how statisticians build a sampling distribution from probabilities. But the binomial distribution can be applied only to a single binary variable. In this chapter, we look at a more commonly used nonparametric test of statistical significance for nominal-level measures: chi-square. The chi-square test allows the researcher to examine multicategory nominal-level variables as well as the relationship between nominal-level measures.

We begin our discussion of chi-square with an example similar to the one used to introduce the binomial distribution in Chapter 7. In this case, we examine the problem of a fair roll of a die. We then turn to applications of the chi-square test in criminal justice.

Testing Hypotheses Concerning the Roll of a Die

In Chapter 7, we examined how you might make a decision about whether to challenge the fairness of a coin used to decide who would serve first in a weekly volleyball match. But what if you had the same question regarding a die used in a friendly game of chance at a local club? Each week, you and a few friends go down to the club and play a game of chance that involves the toss of a die. Let's say that the house (the club) wins whenever you roll a two or a six. You win whenever you roll a three or a four, and no one wins when you roll a one or a five. Over the month, you have played the game 60 times. Of the 60 rolls of the die, you have lost 24, rolling six 20 times and rolling two 4 times (see [Table 9.1](#)). You have won 10 times in total, rolling three 6 times and rolling four 4 times. The remaining 26 rolls of the die were split, with 16 ones and 10 fives.

Table 9.1

Frequency Distribution for 60 Rolls of a Die

1	No winner	16
2	You lose	4
3	You win	6
4	You win	4
5	No winner	10
6	You lose	20
Total (Σ)		60

As in the case of the coin toss, you and your friends have begun to be suspicious. Does it make sense that there should be such an uneven split in the outcomes of the game if the die is fair? Should you raise this issue with the club and suggest that they change their die? You don't want to appear to be a sore sport. Nonetheless, if the distribution of rolls of the die that you observed is very unlikely given a fair die, you would be willing to make a protest.

The Chi-Square Distribution

You cannot use the binomial distribution to assess the fairness of the die because the binomial distribution assumes that there are only two potential outcomes for each event—for example, a head or tail on each toss of a coin. For the die, there are six potential outcomes: a roll of one, two, three, four, five, or six. In such cases, you can make use of another sampling distribution, called the **chi-square (χ^2) distribution**. Like the binomial distribution, which varies depending on the number of trials conducted, the chi-square distribution varies from problem to problem. However, the chi-square distribution varies not according to the number of trials that are conducted but according to the number of **degrees of freedom** (df) associated with a test. The number of degrees of freedom refers to how much a mathematical operation is free to vary, or take on any value, after an agreed-upon set of limitations has been imposed.

In the chi-square distribution, these limitations are associated with the number of categories, or potential outcomes, examined. To define the degrees of freedom of a chi-square test, we ask how many categories would have to be known for us to predict the remaining categories with certainty. For example, if we know that there are 60 rolls of the die and we also know the precise number of events that fall in five of the six categories, we will be able to predict the sixth category simply by subtracting from the total number of events (60) the number in the five known categories (see [Table 9.2](#)). If two categories are blank, we can predict the total of both, but not the exact split between them. Accordingly, the number of degrees of freedom for this example is 5. Once we know the number of events or observations in five categories,

Table 9.2

Frequency Distribution for 60 Rolls of a Die with Information Missing

1	No winner	16
2	You lose	4
3	You win	6
4	You win	4
5	No winner	10
6	You lose	?
Total (Σ)		60

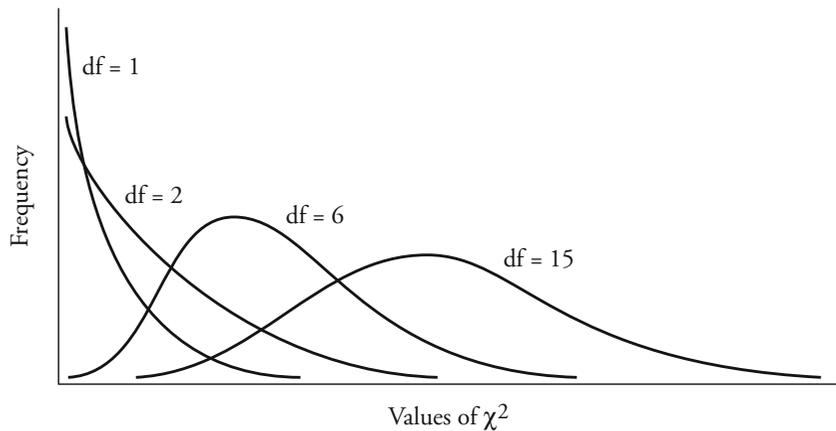
Frequency of category 6 = (total frequency) – (sum of categories 1 to 5)
 $20 = 60 - 40$

we can predict the number in the sixth with certainty. More generally, you can identify the degrees of freedom for a one-variable chi-square distribution using the equation $df = k - 1$, where k equals the number of categories in your measure (for our example, $6 - 1 = 5$).

Figure 9.1 shows how chi-square distributions vary according to the number of degrees of freedom. The height of the distribution represents the proportion of cases found at any specific value of the **chi-square statistic**. As the number of degrees of freedom grows, the height of the chi-square distribution decreases, with a longer and longer tail to the right. This means that the proportion of cases found above higher values of the chi-square statistic grows as the number of degrees of freedom increases. To understand what this means substantively, as well as how the chi-square distribution is used in making decisions about hypotheses, it is important to see how the chi-square statistic is calculated.

Figure 9.1

Chi-Square Distributions for Various Degrees of Freedom



Calculating the Chi-Square Statistic

The formula for the chi-square statistic is presented in Equation 9.1.

$$\chi^2 = \sum_{i=1}^k \frac{(f_o - f_e)^2}{f_e} \tag{Equation 9.1}$$

The summation symbol in the body of the equation has $i = 1$ below it and k above it. This means that we sum the quantity that follows the summation symbol for each category from the first to the k th, or last, category. Since there are six categories in our example, we will have to carry out the same calculation six times, once for each of the six potential outcomes of the roll of a die.

The quantity that follows the summation symbol includes two symbols, f_o and f_e . The symbol f_o represents the frequency of the events observed in a category, or the **observed frequencies**. For example, in 20 of the 60 trials, a six was rolled (see Table 9.1). The observed frequency for a roll of six is 20. The symbol f_e represents the **expected frequency** of a category. The expected frequencies are ordinarily defined by the null hypothesis. In our example, they represent the number of events that would be expected in each category in the long run if the die were fair. Because a fair die is one for which there is an equal chance of obtaining any of the six potential outcomes, we divide the 60 observations evenly across the six categories. This leads to an expected frequency of 10 for each potential outcome. Table 9.3 shows the expected and observed frequencies for our example.

To calculate the chi-square statistic, Equation 9.1 tells us first to subtract the expected frequency from the observed frequency in each category. We then square the result and divide that quantity by the expected frequency of the category. For example, for a roll of six, we subtract 10 (the expected frequency) from 20 (the observed frequency). We then square that quantity (to get 100) and divide the result by 10. This gives us 10 for a roll of six. After carrying out this computation for each category, as is done in Table 9.4, we add up the results for all six categories

Table 9.3

Expected and Observed Frequencies for 60 Rolls of a Fair Die

	f_e	f_o
1	10	16
2	10	4
3	10	6
4	10	4
5	10	10
6	10	20
Total (Σ)	60	60

Table 9.4

Computation of Chi-Square for 60 Rolls of a Die

OUTCOME A	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1	16	10	6	36	3.6
2	4	10	-6	36	3.6
3	6	10	-4	16	1.6
4	4	10	-6	36	3.6
5	10	10	0	0	0
6	20	10	10	100	10.0
					$\Sigma = 22.4$

to obtain the total chi-square statistic. The chi-square statistic for this example is 22.4.

The chi-square statistic measures how much the observed distribution differs from that expected under the null hypothesis. If the observed frequencies are similar to the expected frequencies, the chi-square statistic is small. If the observed frequencies are the same as the expected frequencies, the chi-square statistic equals 0. The more the observed frequencies differ from the expected frequencies, the larger the chi-square statistic will be. What does this mean in terms of making a decision about the fairness of the die? To find out, we have to turn to a table of probabilities associated with the chi-square distribution.

Linking the Chi-Square Statistic to Probabilities: The Chi-Square Table

In Chapters 7 and 8, we used the binomial formula to calculate the probability associated with each of the possible outcomes in our sample. For other tests of statistical significance, including chi-square, we can take advantage of already calculated probability distributions. Appendix 2 presents a table of probabilities associated with chi-square distributions with degrees of freedom from 1 to 30. The chi-square table does not give us the probability associated with every possible outcome, but rather provides probabilities and then lists the chi-square statistics associated with them.

As illustrated in the chi-square table in Appendix 2, a larger chi-square statistic is associated with a smaller significance level, or α value. For example, under one degree of freedom, a statistic of 2.706 is associated with a significance level of 0.10, a statistic of 3.841 with an α value of 0.05, and a statistic of 10.827 with an α value of 0.001. This also means that the larger the chi-square statistic obtained in a test, the less likely it is that the observed distribution is drawn from the expected distribution. This logic makes good common sense. For our example of the roll of a die, it is reasonable to become more suspicious about the fairness of the die as the number of events in the different categories

becomes more uneven. If we expect 10 events in each category and actually get one with 20, one with 16, and two others with only 4, this should begin to make us suspicious. If one or two categories have 25 cases and two or three have none, it seems even more likely that the die is not a fair one. But if each category has about 10 cases, which is to be expected in the long run with 60 rolls of a fair die, both common sense and chi-square give us little reason to suspect a biased die.

Notice as well in Appendix 2 that as the number of degrees of freedom gets larger, a larger chi-square statistic is needed to arrive at the same probability value. For example, with one degree of freedom, a chi-square statistic of 3.841 is associated with an α value of 0.05. With 30 degrees of freedom, a statistic of 43.773 is needed to achieve the same threshold. This reflects the difference in the shape of chi-square distributions with different degrees of freedom and makes good sense if you consider how the chi-square statistic is calculated. A separate addition is made to the chi-square statistic for each possible category. Accordingly, it makes sense to demand a larger statistic as the number of categories in the test increases.

What about our decision regarding the roll of the die? Looking at Appendix 2, we can see that with five degrees of freedom a chi-square statistic of 11.070 is associated with a significance level of 0.05. This means that in the long run we would expect to obtain a chi-square statistic of 11.070 in only 5 in 100 samples if the die is fair. In fact, we obtained a chi-square statistic of 22.4. This number is even larger than that needed for a significance level of 0.001. Accordingly, the observed significance level for this test is less than 0.001 ($p < 0.001$). If the die were fair, the probability of getting a distribution like the one observed in our 60 rolls of a die would be less than 1 in 1,000. Given this result, we would likely come to the conclusion that the die was not a fair one and call for the club to use a new one.

A Substantive Example: The Relationship Between Assault Victims and Offenders

We can illustrate the chi-square test for a single variable by considering the responses from a random sample survey of Illinois residents.¹ One of the primary purposes of the survey was to examine the effect of victimization on the physical and mental health of adults. Each respondent was asked about a variety of possible victimization experiences. When the person claimed to have experienced a crime, a series of follow-up questions were asked about the circumstances of the event. Table 9.5 presents the frequency distribution of the responses to a question about the

¹See Chester L. Britt, "Health Consequences of Criminal Victimization," *International Review of Victimology* 8 (2001): 63–73 for a description of the study.

Table 9.5

Relationship Between Assault Victim and Offender

CATEGORY	FREQUENCY (<i>N</i>)
Stranger	166
Acquaintance	61
Friend	35
Boyfriend/girlfriend	38
Spouse	66
Other relative	44
Total (Σ)	410

relationship between the victim and the offender for those persons who claimed to have been assaulted.

A simple research question using these data might focus on whether the victim-offender relationship was unevenly distributed among the population of assault victims. To answer our research question, we follow the form of a statistical test introduced in Chapter 8.

We begin by stating the assumptions of our chi-square test. The level of measurement required for chi-square is nominal. We make no specific assumptions regarding the shape of the population distribution, as the chi-square test is a nonparametric test of statistical significance. Although the chi-square test ordinarily requires a fully independent random sample, this sample was selected without replacement.² This is not a serious violation of our assumptions because the sample is very small relative to the population of interest. Note that our null hypothesis is that the victim-offender relationship in the population is evenly or randomly distributed across the categories examined. The research hypothesis is that the victim-offender relationship is not randomly or evenly distributed in the population.

Assumptions:

Level of Measurement: Nominal scale.

Population Distribution: No assumption made.

Sampling Method: Independent random sampling (no replacement; sample is small relative to population).

Sampling Frame: Persons aged 18 and over in the state of Illinois.

Hypotheses:

H_0 : The type of victim-offender relationship for assault victims is randomly distributed.

H_1 : The type of victim-offender relationship for assault victims is not randomly distributed.

²There are certain specific situations in which the chi-square test does not require sampling with replacement; see B. S. Everitt, *The Analysis of Contingency Tables* (London: Chapman and Hall, 1997).

The Sampling Distribution Since we are analyzing the distribution of cases for a nominal variable, the chi-square distribution provides an appropriate means of assessing whether the observations are randomly distributed across the six categories of victim-offender relationships. For a single nominal variable, the number of degrees of freedom for the chi-square test is $df = k - 1 = 6 - 1 = 5$.

Significance Level and Rejection Region Since we have no reason to impose a stricter or more lenient level of statistical significance on our analysis, we will use a significance level (α) of 0.05. Given that the number of degrees of freedom associated with this chi-square test is 5 and the significance level is 0.05, we see from Appendix 2 that the corresponding critical value of the chi-square distribution is 11.070. Accordingly, if the calculated value of the chi-square statistic in our example is greater than 11.070, we will reject the null hypothesis and conclude that type of victim-offender relationship among assault victims is not randomly distributed.

The Test Statistic Equation 9.1 provides the formula for calculating the chi-square statistic to test for random assignment of cases to each category or value. We begin by calculating the expected frequency (f_e) for each cell in the table. Again, as in the example of the die, we would expect under the null hypothesis that there would be an equal number of cases in each of the categories examined. To calculate the expected frequency mathematically, we divide the total N of cases by the number of categories. This is done below, where we get an expected value for each category of 68.333:

$$f_e = \frac{N}{k} = \frac{410}{6} = 68.333$$

After calculating the expected frequency, we can proceed to calculate the chi-square statistic. Table 9.6 presents the observed and expected frequencies for each cell and the appropriate calculations for determining the value of the chi-square statistic. We find the value of the test statistic to be 178.85.

Table 9.6

Computation of Chi-Square for Type of Victim-Offender Relationship

CATEGORY	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Stranger	166	68.333	97.667	9,538.843	139.593
Acquaintance	61	68.333	-7.333	53.773	0.787
Friend	35	68.333	-33.333	1,111.089	16.260
Boyfriend/girlfriend	38	68.333	-30.333	920.091	13.465
Spouse	66	68.333	-2.333	5.443	0.080
Other relative	44	68.333	-24.333	592.095	8.665
					$\Sigma = 178.849$

The Decision The critical value for our test of statistical significance was 11.070, meaning that a calculated chi-square statistic greater than this critical value would lead to rejection of the null hypothesis. The value of our test statistic is 178.85, which is much larger than our critical chi-square value. Accordingly, the observed significance level of our test is less than the significance criterion we set at the outset ($p < 0.05$). On the basis of this outcome, we reject the null hypothesis and conclude that type of victim-offender relationship among assault victims is not randomly distributed. Of course, we cannot be certain that the null hypothesis is false for the population we are examining. We make our decision with a set risk of a Type I error defined at the outset of our test.

Relating Two Nominal-Scale Measures in a Chi-Square Test

In criminal justice and criminology, we seldom examine research issues like the fairness of a die or the randomness of type of victim-offender relationship, which are concerned with outcomes on only one measure. More often, we are interested in describing the relationships among two or more variables. For example, we may want to assess whether men and women are likely to be placed in different types of treatment facilities or whether different ethnic groups receive different types of sanctions. For each of these examples, two measures must be assessed at the same time. In the former, we examine both gender and type of treatment facility. In the latter, we examine type of sentence and ethnicity. Below, we use the example of a study of white-collar criminals to illustrate the use of chi-square in making inferences about the relationship between two variables: recidivism and sanction type.

A Substantive Example: Type of Sanction and Recidivism Among Convicted White-Collar Criminals

In a study of white-collar offenders, data on reoffending from FBI records over a ten-year period were examined. The sample included offenders from seven U.S. district courts, convicted of eight different white-collar crimes (antitrust violations, securities fraud, mail and wire fraud, false claims and statements, credit and lending institution fraud, bank embezzlement, income tax fraud, and bribery). The sample was chosen randomly without replacement.³ The research question concerned whether imprisonment of white-collar offenders impacted upon reoffending.

³In this case, a stratified random sample was selected in order to ensure a broad sampling of white-collar offenders. For our example here, we treat the sample as a simple random sample. See David Weisburd, Elin Waring, and Ellen Chayet, "Specific Deterrence in a Sample of Offenders Convicted of White Collar Crimes," *Criminology* 33 (1995): 587–607.

The likelihood of rearrest for a group of offenders who received a prison sanction was compared with that of a matched group who did not receive a prison sanction. The researchers found that 33.0% of the prison group ($N = 100$) was rearrested during the follow-up period, in contrast to 28.4% of the no-prison group ($N = 67$). What conclusions can we come to concerning white-collar criminals generally?

To answer our research question, we follow the standard format of a test of statistical significance. We begin by stating our assumptions. Remember that to state the assumptions you must choose the type of test you will use. In this case, we have chosen a chi-square test for relating two nominal-level measures.

Assumptions:

Level of Measurement: Nominal scales.

Population Distribution: No assumption made.

Sampling Method: Independent random sampling (no replacement; sample is small relative to population).

Sampling Frame: Offenders from seven federal judicial districts convicted of eight different white-collar crimes.

Hypotheses:

H_0 : There is no difference in the likelihood of rearrest among similar white-collar offenders sentenced or not sentenced to prison. (Or, likelihood of rearrest and imprisonment are independent.)

H_1 : There is a difference in the likelihood of rearrest among similar white-collar offenders sentenced or not sentenced to prison. (Or, likelihood of rearrest and imprisonment are not independent.)

The level of measurement required for a chi-square test is nominal. Our example includes two variables: rearrest and type of sanction. Each is measured as a binary nominal variable. For rearrest, we examine those rearrested versus those not rearrested in the follow-up period. For type of sanction, we differentiate between those who were sentenced to prison and those who did not receive a prison sanction. In regard to the population distribution, chi-square is a nonparametric test and therefore requires no specific assumption.

The sample was selected randomly, but as is the case with most criminal justice studies, the investigators did not sample with replacement. At the same time, the population from which the sample was drawn is very large relative to the sample examined, and thus we have no reason to suspect that this violation of the assumptions will affect our test result.

The sampling frame includes offenders from seven federal judicial districts convicted of eight different white-collar crimes. Accordingly, our inferences relate directly to the population of those offenses and those districts. As discussed in Chapter 8, it is necessary to explain why your

sample is representative of a broader population of cases if you want to make inferences beyond your sampling frame. In this study, the seven districts examined were seen as providing a sample with geographic spread throughout the United States, and the selected white-collar offenses were defined as offering a “broad sampling of white-collar offenders.”

In most applications of the chi-square test, one cannot assign a directional research hypothesis. This is because chi-square requires a nominal level of measurement, which does not assign order or value to the categories examined. Nonetheless, in the special case of two binary variables, as examined here, the researcher can choose between a directional and a nondirectional research hypothesis. In our example, a directional hypothesis would be that the prison group is more likely than the no-prison group to be rearrested or that the no-prison group is more likely than the prison group to be rearrested. The research hypothesis stated by the investigators in this study was nondirectional. It stated simply that the two groups (prison and no-prison) differ in terms of likelihood of rearrest during the follow-up period. The null hypothesis was that there is no difference between the prison and no-prison groups.

Researchers often state the hypotheses of a chi-square test in terms of the independence of the variables that are examined. Stated this way, the null hypothesis would be that prison group (prison or no-prison) is independent, or unrelated to rearrest, in the follow-up period. The research hypothesis is that prison group is not independent. While this method of stating the hypotheses for your test sounds different, it leads to the same conclusions. If the two groups differ in terms of likelihood of rearrest in the follow-up period, then prison group and recidivism are related and thus not independent. If there is no difference, then prison group is unrelated to, or independent of, recidivism in the follow-up period.

The Sampling Distribution Because we are examining the relationship between two nominal-scale variables, the chi-square distribution provides an appropriate sampling distribution for our test. However, our decision about degrees of freedom is not as straightforward as that in the example of a roll of a die. In this case, we must take into account the

Table 9.7

Recidivism Among 167 White-Collar Criminals According to Whether They Did or Did Not Receive Prison Sentences

	Imprisoned	Not imprisoned	Row total
Subsequently arrested	Cell A 33	Cell B 19	52
Not subsequently arrested	Cell C 67	Cell D 48	115
Column total	100	67	167

joint distribution of our measures. This is illustrated in Table 9.7, which shows two potential outcomes for the prison variable and two potential outcomes for the arrest variable. We have four **cells**, or four possible combined outcomes. Cell A is for offenders who received a prison sanction and were arrested in the follow-up period. Cell B is for offenders who did not receive a prison sanction and were arrested in the follow-up period. Cell C is for offenders who received a prison sanction and were not arrested in the follow-up period. Cell D is for offenders who did not receive a prison sanction and were not arrested in the follow-up period.

If we sum across and down the cells, we gain two row **marginals** and two column marginals. The row marginals represent the totals for the rows: 52 for those arrested and 115 for those not arrested. The column marginals represent the totals for the columns: $N = 100$ for the prison group, and $N = 67$ for the no-prison group. If you know the row and column marginals, as is assumed in computing the degrees of freedom for chi-square, you can predict with certainty the remaining cells, once the value of any one cell is known (see Table 9.8). Degrees of freedom for a two-variable chi-square can be gained more simply through the formula $df = (r - 1)(c - 1)$, where r represents the number of rows and c the number of columns. For our example, there are two rows and two columns, so $df = (2 - 1)(2 - 1) = 1$.

Significance Level and Rejection Region We stated no reason at the outset for choosing for our example a stricter or more lenient significance threshold than is used by convention. Accordingly, we use a significance level of 0.05 for our test. Our rejection region is defined by the chi-square table (see Appendix 2). Importantly, the chi-square distribution is not concerned with the direction of outcomes in a test. It tells us to what extent the observed frequencies in our example differ from those that would be expected under the null hypothesis of no difference. Whether they differ in one direction or another, the chi-square statistic will always be positive.

Table 9.8

Predicting the Missing Cells in a Two-Variable Chi-Square Test

	Imprisoned	Not imprisoned	Row total
Subsequently arrested	Cell A 33	Cell B ?	52
Not subsequently arrested	Cell C ?	Cell D ?	115
Column total	100	67	167
Given that cell A = 33: Cell B = $(52 - 33) = 19$ Cell C = $(100 - 33) = 67$ Cell D = $(115 - 67) = 48$			

The terms “directional” and “nondirectional” are very tenuous ones in a chi-square test. Chi-square assumes nominal-scale variables, which by definition do not provide information about the order of values in a measure. If we cannot specify the order of two measures, we cannot speak of the direction of their relationship. As noted earlier, in most situations a directional hypothesis is not appropriate for a chi-square test. In the special case of two binary variables, however, researchers do sometimes use chi-square to examine directional research hypotheses. For example, we might have stated our research hypothesis as “The likelihood of arrest in the follow-up period for white-collar offenders sentenced to prison is lower than that of similar white-collar offenders not sentenced to prison.”

However, our research hypothesis is nondirectional, as is the table of chi-square values. To define our rejection region, we turn to the row of the table associated with one degree of freedom. Under a significance level (α) of 0.05, we see a score of 3.841. For us to reject the null hypothesis, our test statistic will have to be greater than this value.⁴

The Test Statistic To apply chi-square to the two-variable case, we need to adapt our original equation. The formula for the chi-square statistic for relating two measures is presented in Equation 9.2.⁵

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_o - f_e)^2}{f_e} \quad \text{Equation 9.2}$$

The only difference between Equation 9.2 and Equation 9.1 is that we have an additional summation symbol. In this case, we do not sum simply across the categories of one measure; rather, we sum across each row (r) and column (c) of the joint distribution of two measures. Accordingly, Equation 9.2 reminds us that we must examine the expected and observed frequencies for every potential outcome we can observe—or, in terms of the chi-square equation, for every cell in our table.

⁴What if we had defined a directional research hypothesis? In this case, we look to the column of the table for twice the value of the desired significance level, since we now have placed all risk of falsely rejecting the null hypothesis in only one direction. For example, for a 0.05 significance level, we turn to the test statistic for a 0.10 level.

⁵When a chi-square test has only one degree of freedom, it is recommended that a correction factor be added if the expected frequencies of any cell fall below 20. The correction provides a more conservative, or smaller, chi-square statistic:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

For our example, this means we must sum across cells A, B, C, and D. As before, we want to compare the observed frequency to the expected frequency in each cell. The observed frequencies are those gained in our research. The expected frequencies are defined through the null hypothesis. The null hypothesis states that there is no difference in arrest rates between the prison and no-prison groups. If this is true, then we should expect the same proportion of arrests in both groups. To calculate the expected frequencies, accordingly, we first need to define the overall proportion of offenders arrested in the follow-up period.

The proportion of offenders arrested overall in the sample in the follow-up period is obtained by dividing the total number of offenders arrested ($N_{\text{cat}} = 52$) by the total number of offenders in the sample ($N_{\text{total}} = 167$):

$$\text{Proportion} = \frac{N_{\text{cat}}}{N_{\text{total}}} = \frac{52}{167} = 0.3114$$

To get the expected frequency for cell A, we multiply this proportion by the marginal total of 100 ($f_e = 31.14$). For the no-prison group, we have a total of 67 offenders. Applying the proportion of the total sample to this group, we multiply 0.3114 by 67 and get an expected frequency of 20.86 for cell B. In practice, we do not need to compute the expected frequencies for the remaining two cells, C and D. Indeed, we could have assigned all of the cells expected frequencies based on knowledge of only one cell. This is what the number of degrees of freedom for this example tells us. If you know the number of cases in one cell, you can predict with certainty the rest. The expected and observed frequencies for our example are shown in [Table 9.9](#).

Now that we have calculated the observed and expected frequencies for each potential outcome, or cell, we can calculate the chi-square statistic. To do this, we first square the difference of the observed and

Table 9.9

Expected and Observed Frequencies of Recidivism and Nonrecidivism for White-Collar Offenders According to Whether They Received Prison Sentences

	Imprisoned	Not imprisoned	Row total
Subsequently arrested	Cell A $f_o = 33$ $f_e = 31.14$	Cell B $f_o = 19$ $f_e = 20.86$	52
Not subsequently arrested	Cell C $f_o = 67$ $f_e = 68.86$	Cell D $f_o = 48$ $f_e = 46.14$	115
Column total	100	67	167

Table 9.10

Computation of Chi-Square for 167 White-Collar Criminals

CELL	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
A	33	31.14	1.86	3.4596	0.1111
B	19	20.86	-1.86	3.4596	0.1658
C	67	68.86	-1.86	3.4596	0.0502
D	48	46.14	1.86	3.4596	0.0750
					$\Sigma = 0.4021$

expected frequencies for each cell, and then we divide this quantity by the expected frequency of the cell:

$$\frac{(f_o - f_e)^2}{f_e}$$

This is done in [Table 9.10](#) for each of the four cells in our problem. Using cell A as an example, we first subtract the expected frequency of 31.14 from the observed frequency of 33. We then square this quantity (1.86), obtaining a result of 3.4596. Dividing this result by the expected frequency in the cell (31.14) gives us 0.1111. The sum of all four cells, 0.4021, is our test statistic.

The Decision Our rejection region was defined as including any chi-square statistic greater than 3.841. The test statistic for our example is only 0.402. Accordingly, we choose not to reject the null hypothesis. The observed significance level for our test is greater than the significance level, or threshold, we set at the outset ($p > 0.05$). We conclude that there is no significant difference in the likelihood of recidivism between white-collar offenders who have and have not been sentenced to prison. Our inferences are made directly to the specific offenses and judicial districts defined in the sampling frame.

Extending the Chi-Square Test to Multicategory Variables: The Example of Cell Allocations in Prison

The previous example illustrates the use of chi-square in the case of two binary variables. We now turn to an extension of the chi-square test to an example including a multicategory nominal-level variable. Our example is drawn from a study of the relationship between prisoners' race and their cell assignments in a large state prison in the northeastern

Table 9.11

Proportions of Non-Hispanic White Prisoners in Seven Cell Blocks

	Non-Hispanic whites	Nonwhites	Row total
Cell block C	48 18.7%	208 81.3%	256 100%
Cell block D	17 31.5%	37 68.5%	54 100%
Cell block E	28 25.0%	84 75.0%	112 100%
Cell block F	32 28.8%	79 71.2%	111 100%
Cell block G	37 12.2%	266 87.8%	303 100%
Cell block H	34 60.7%	22 39.3%	56 100%
Cell block I	44 14.1%	268 85.9%	312 100%
Column total	240 19.9%	964 80.1%	1,204 100%

United States.⁶ We examine the placement of non-Hispanic white and “nonwhite” inmates (including Hispanics) into seven cell blocks. The sample includes all prisoners in the general prison population for a single day. The distribution of cases is presented in [Table 9.11](#).

If cell assignments were made on considerations unrelated to race, we would expect to find the proportion of non-Hispanic whites in each cell block roughly equivalent to the proportion of non-Hispanic whites in the general prison population (19.9%; see the marginal for non-Hispanic whites in [Table 9.11](#)). Such equivalence is not evident. In block G, for example, non-Hispanic whites constituted 12.2% of the inmates. In block H, they comprised 60.7%. Do results for this sample allow us to conclude that there is disparity in cell-block assignments throughout the year in the prison.

Assumptions:

Level of Measurement: Nominal scales.

Population Distribution: No assumption made.

Sampling Method: Independent random sampling (the entire sampling frame is examined).

Sampling Frame: All prisoners in the general prison population on a specific day.

⁶See Douglas McDonald and David Weisburd, “Segregation and Hidden Discrimination in Prisons: Reflections on a Small Study of Cell Assignments,” in C. Hartchen (ed.), *Correctional Theory and Practice* (Chicago: Nelson Hall, 1991).

Hypotheses:

H_0 : Cell-block assignment and race are independent.

H_1 : Cell-block assignment and race are not independent.

As in our previous example, we assume a nominal level of measurement for our test and do not make assumptions regarding the form of the population distribution. Prisoner race is measured at the binary nominal level, and cell block is a multicategory nominal scale.

The sample includes all cases in the sampling frame. Accordingly, we do not need to use statistical inference to make statements about that population. However, the study was designed not only to describe prison-cell allocations on that day, but also to make more general statements about cell allocations in the prison studied throughout the year. This is not an uncommon scenario in criminal justice research, in good part because the realities of the criminal justice system often preclude sampling beyond specific institutions or outside of specific time frames. This means, however, that the researchers seek to make inferences beyond their sampling frame.

If cell allocations on the day examined in this study are representative of cell allocations more generally throughout the year, then the inferences made on the basis of the test will be reliable. If not, then the test will not provide for valid inferences. In our example, the investigators argue:

There was no reason to suspect that the cell assignments of prisoners on that day differed substantially from assignments on other days. Moreover, these cell assignments represented the results of decisions made over the course of months and perhaps years prior to the date of drawing the sample. There was every reason to believe, consequently, that cell assignments on that date constituted a valid representation of cell assignment decisions made during the several months prior to that day.

Our research question asks whether we would be likely to obtain the distribution we observe in our sample if assignment to cell blocks were colorblind in the population. Stated in the form of hypotheses, we ask whether race and cell-block assignment are independent. If they are independent, as proposed in our null hypothesis, then we would expect about the same proportion of nonwhite and non-Hispanic white prisoners in each cell block. Our research hypothesis is nondirectional. It states that race and cell-block assignment are not independent. In this example, as in most chi-square tests, use of nominal-scale measures, which do not assign order or value to categories, means that one cannot define a directional research hypothesis.

The Sampling Distribution Because we are examining the relationship between two nominal variables, one binary and one multcategory, we use the chi-square sampling distribution. The number of degrees of freedom for our problem is defined as in the previous example:

Working It Out

$$\begin{aligned} df &= (r - 1)(c - 1) \\ &= (7 - 1)(2 - 1) \\ &= 6 \end{aligned}$$

In this case, we have seven categories for our row variable (cell block) and two categories for our column variable (ethnicity). The number of degrees of freedom for our sampling distribution is six.

Significance Level and Rejection Region As we have no reason to propose more lenient or stricter significance criteria than are used by convention, we will set a 0.05 significance level. To define our rejection region, we turn to the row of the chi-square table associated with six degrees of freedom. Under the 0.05 column, a chi-square statistic of 12.592 is listed. If the test statistic is greater than this critical value, then it falls within the rejection region of the test.

The Test Statistic To calculate the test statistic in this multcategory example, we follow the same procedure used for the two-by-two table in the previous section. Our first task is to define the expected frequency for each cell of the table. We do this, as before, by dividing a marginal of the table by the total proportion of cases. Taking the overall number of non-Hispanic whites in the sample, we obtain a proportion of 0.1993:

Working It Out

$$\begin{aligned} \text{Proportion} &= \frac{N_{\text{cat}}}{N_{\text{total}}} \\ &= \frac{240}{1,204} \\ &= 0.199335 \end{aligned}$$

Table 9.12

Observed Frequencies and Expected Frequencies for Non-Hispanic White and Nonwhite Prisoners in Seven Cell Blocks

	Non-Hispanic whites	Nonwhites	Row total
Cell block C	$f_o = 48$ $f_e = 51.030$	$f_o = 208$ $f_e = 204.970$	256
Cell block D	$f_o = 17$ $f_e = 10.764$	$f_o = 37$ $f_e = 43.236$	54
Cell block E	$f_o = 28$ $f_e = 22.326$	$f_o = 84$ $f_e = 89.674$	112
Cell block F	$f_o = 32$ $f_e = 22.126$	$f_o = 79$ $f_e = 88.874$	111
Cell block G	$f_o = 37$ $f_e = 60.399$	$f_o = 266$ $f_e = 242.601$	303
Cell block H	$f_o = 34$ $f_e = 11.163$	$f_o = 22$ $f_e = 44.837$	56
Cell block I	$f_o = 44$ $f_e = 62.193$	$f_o = 268$ $f_e = 249.807$	312
Colum total	240	964	1,204

To calculate the expected frequency in each cell in the non-Hispanic whites column, we multiply this proportion by the marginal total for each row. So, for example, for cell block C, we multiply 256 by 0.199335, leading to an expected frequency for non-Hispanic whites of 51.030. We then replicate this procedure for each of the six other cells in the non-Hispanic whites column. To calculate the expected frequencies for the nonwhites column, we simply subtract the expected frequency for the non-Hispanic whites column from the row marginal. So, for example, for nonwhites in cell block C, the expected frequency is 256 (the marginal total) minus 51.030 (the expected frequency for non-Hispanic whites for that cell block), or 204.970. Table 9.12 includes the expected and observed frequencies for the 14 cells in our example.

To obtain the test statistic, we use Equation 9.2, which may be applied to any two-variable chi-square problem:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_o - f_e)^2}{f_e}$$

Again we begin by subtracting the expected frequency from the observed frequency in each cell and squaring the result. This quantity is then divided by the expected frequency of the cell. The chi-square statistic is found by summing the result across all 14 cells. The full set of calculations for the test statistic is presented in Table 9.13. The chi-square score for our example is 88.3610.

Table 9.13

Computation of Chi-Square for Non-Hispanic White (W) and Nonwhite (NW) Prisoners in Seven Cell Blocks

CELL BLOCK	RACE	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
C	W	48	51.030	-3.030	9.1809	0.1799
C	NW	208	204.970	3.030	9.1809	0.0448
D	W	17	10.764	6.236	38.8877	3.6128
D	NW	37	43.236	-6.236	38.8877	0.8994
E	W	28	22.326	5.674	32.1943	1.4420
E	NW	84	89.674	-5.674	32.1943	0.3590
F	W	32	22.126	9.874	97.4959	4.4064
F	NW	79	88.874	-9.874	97.4959	1.0970
G	W	37	60.399	-23.399	547.5132	9.0649
G	NW	266	242.601	23.399	547.5132	2.2568
H	W	34	11.163	22.837	521.5286	46.7194
H	NW	22	44.837	-22.837	521.5286	11.6317
I	W	44	62.193	-18.193	330.9852	5.3219
I	NW	268	249.807	18.193	330.9852	1.3250
						$\Sigma = 88.3610$

The Decision The outcome of 88.3610 is much greater than the critical value for our test of 12.592. Accordingly, we reject the null hypothesis that race and cell-block allocation are independent (using a 5% significance level). We conclude that there is a statistically significant relationship between the distribution of prisoners across cell blocks and their race.

Extending the Chi-Square Test to a Relationship Between Two Ordinal Variables: Identification with Fathers and Delinquent Acts

The examples of the application of the chi-square test presented so far have used only nominal-scale variables. This is consistent with the assumptions of the chi-square test. But in practice researchers sometimes use chi-square to test for independence when one or both of the variables have been measured at the ordinal level of measurement. This test for independence can provide important information to the researcher. However, because the chi-square test assumes a nominal scale of measurement, it does not pay attention to the order of the categories in an ordinal scale. This means that a statistically significant finding can tell the researcher only that the distribution of scores observed is different from that expected had there been no relationship. It cannot test for whether the values of one variable increase as the values of the other increase or, conversely, whether the scores on one measure increase as those on the

Table 9.14

Affectional Identification with Father by Number of Delinquent Acts

AFFECTIONAL IDENTIFICATION WITH FATHER	DELINQUENT ACTS			Row total
	None	One	Two or More	
In every way	77 63.636%	25 20.661%	19 15.702%	121 100%
In most ways	263 65.099%	97 24.010%	44 10.891%	404 100%
In some ways	224 57.881%	97 25.065%	66 17.054%	387 100%
In just a few ways	82 47.674%	52 30.233%	38 22.093%	172 100%
Not at all	56 40.580%	30 21.739%	52 37.681%	138 100%
Column total	702 57.447%	301 24.632%	219 17.921%	1,222 100%

other decrease. When you use the chi-square test for ordinal-scale variables, the test itself treats the variables as if they were simply composed of a group of nominal categories.

Table 9.14 presents data from the Richmond Youth Survey report on the relationship between number of delinquent acts and affectional identification with one's father. The distribution of cases presented refers only to the white males who responded to the survey.⁷ The sample was a random sample (drawn without replacement) for all high school-age white males in Richmond, California, in 1965. The size of the sample is small relative to the sampling frame.

If delinquency were unrelated to attachment to one's family—here indicated by the level of affectional identification with one's father—we would expect to find the distribution of cases for each level of delinquency to be roughly equal across levels of identification. The distribution of cases provides some indication that these variables are not, in fact, independent. For example, among the youths who wanted to be like their father in every way, 63% reported that they had not committed a delinquent act. This was true for only 41% of those who did not want to be at all like their fathers. Our question is whether the differences we observe in our sample are large enough for us to conclude that identification with one's father and delinquency are related in the population from which our sample has been drawn.

⁷David F. Greenberg, "The Weak Strength of Social Control Theory," *Crime and Delinquency* 45:1 (1999): 66–81.

Assumptions:

Level of Measurement: Nominal scales (our study examines two ordinal-scale measures).

Population Distribution: No assumption made.

Sampling Method: Independent random sampling (no replacement; sample is small relative to population).

Sampling Frame: High school–age white males in Richmond, California, in 1965.

Hypotheses:

H_0 : Affectional identification with father and delinquency are independent.

H_1 : Affectional identification with father and delinquency are not independent.

The Sampling Distribution Although we are using two ordinal-scale measures rather than two nominal-scale measures, we have chosen to use the chi-square sampling distribution to test for a relationship. This violation of the nominal-scale assumption for the chi-square test is acceptable. However, by placing the violation of the assumption in parentheses next to the test requirement of a nominal level of measurement, we remind ourselves that chi-square is not concerned with the order of the categories in the measures examined. It treats the two ordinal-scale measures as if they were nominal-scale measures and simply tests for whether the distributions among the categories depart from what we would expect under an assumption of independence. As we did in the two previous examples, we calculate degrees of freedom as follows:

Working It Out

$$\begin{aligned}df &= (r - 1)(c - 1) \\ &= (5 - 1)(3 - 1) \\ &= 8\end{aligned}$$

Significance Level and Rejection Region There is no reason to propose a more lenient or stricter significance level for this analysis, so we will stick with a 0.05 significance level. Given that we have eight degrees of freedom and a significance level of 0.05, we can consult the chi-square table and determine that our critical value of the chi-square statistic is 15.507. If the test statistic is greater than this value, it falls in the rejection

region of the test, and we can conclude that delinquency is significantly related to affectional identification.

The Test Statistic To determine the expected frequency for each cell in the table, we follow the same format we have used in the previous two examples. As before, we start with the calculation of the marginal for no delinquent acts and divide by the total number of cases, which gives us a value of 0.574468:

Working It Out

$$\begin{aligned} \text{Proportion} &= \frac{N_{\text{cat}}}{N_{\text{total}}} \\ &= \frac{702}{1,222} \\ &= 0.574468 \end{aligned}$$

To calculate the expected frequency for each cell in the no delinquent acts column, we take this proportion and multiply it by the marginal total for the row. For the first row, we multiply 0.574468 by 121, which gives us an expected frequency of 69.511. Similarly, for the second row, we multiply 0.574468 by 404, giving us an expected frequency of 232.085. We continue this procedure for the remaining three rows in the no delinquent acts column.

For the second column, we need to determine the marginal proportion for those cases with one delinquent act. Since there are 301 cases in the marginal for one delinquent act, the corresponding proportion is $(301/1,222) = 0.246318$. To obtain the expected frequencies for this second column, we multiply 0.246318 by the corresponding row marginal. So, for the first row of the second column, the expected frequency is obtained by multiplying 0.246318 by 121, which gives us 29.804. This procedure is repeated to complete the remaining cells in the second column.

Finally, to determine the expected frequencies for the cells in the third column, we simply add the expected frequencies for the first two columns and subtract that sum from the row marginal. For example, in the first row, the two expected frequencies obtained thus far are 69.511 and 29.804. If we add these two values ($69.511 + 29.804 = 99.315$) and subtract this sum from the row marginal (121), we find that the expected frequency for the cell in the third column of the first row is equal to $(121 - 99.315) = 21.685$. To complete the table of expected frequencies,

Table 9.15

Observed and Expected Frequencies for Affectional Identification with Father and Number of Delinquent Acts

AFFECTIONAL IDENTIFICATION WITH FATHER	DELINQUENT ACTS			Row total
	None	One	Two or More	
In every way	$f_o = 77$ $f_e = 69.511$	$f_o = 25$ $f_e = 29.804$	$f_o = 19$ $f_e = 21.685$	121
In most ways	$f_o = 263$ $f_e = 232.085$	$f_o = 97$ $f_e = 99.512$	$f_o = 44$ $f_e = 72.403$	404
In some ways	$f_o = 224$ $f_e = 222.319$	$f_o = 97$ $f_e = 95.325$	$f_o = 66$ $f_e = 69.355$	387
In just a few ways	$f_o = 82$ $f_e = 98.809$	$f_o = 52$ $f_e = 42.367$	$f_o = 38$ $f_e = 30.824$	172
Not at all	$f_o = 56$ $f_e = 79.277$	$f_o = 30$ $f_e = 33.992$	$f_o = 52$ $f_e = 24.731$	138
Column total	702	301	219	1,222

we repeat this operation for the remaining cells in the third column. Table 9.15 contains all the observed and expected frequencies.

To obtain the test statistic, we again use Equation 9.2, which may be applied to any two-variable chi-square problem:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_o - f_e)^2}{f_e}$$

Again we begin by subtracting the expected from the observed frequency in each cell and squaring the result. This quantity is then divided by the expected frequency of the cell. The chi-square statistic is found by summing the result across all cells. The full set of calculations necessary for obtaining the value of the chi-square test statistic appears in Table 9.16. The chi-square statistic for our test has a value of 61.532.

The Decision

The calculated chi-square statistic of 61.532 is much larger than the critical value of 15.507 for the chi-square distribution with eight degrees of freedom. This means that the observed significance level for our test is less than the criterion significance level we set at the outset ($p < 0.05$). Thus, we reject the null hypothesis that affectional identification with father is not related to number of delinquent acts (at a 5% significance level). In turn, we conclude that for adolescent males there is a statistically significant relationship between delinquency and affectional identification with father. Importantly, this statistical inference refers directly to our sampling frame: high school-age white males in Richmond, California, in 1965.

Table 9.16

Computation of Chi-Square for Affectional Identification
with Father and Delinquency

IDENTIFICATION	DELINQUENCY	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Every way	None	77	69.511	7.489	56.085	0.807
Every way	One	25	29.804	-4.804	23.078	0.774
Every way	Two or more	19	21.685	-2.685	7.209	0.332
Most ways	None	263	232.085	30.915	955.737	4.118
Most ways	One	97	99.512	-2.512	6.310	0.063
Most ways	Two or more	44	72.403	-28.403	806.730	11.142
Some ways	None	224	222.319	1.681	2.826	0.013
Some ways	One	97	95.325	1.675	2.806	0.029
Some ways	Two or more	66	69.355	-3.355	11.256	0.162
Few ways	None	82	98.809	-16.809	282.543	2.859
Few ways	One	52	42.367	9.633	92.795	2.190
Few ways	Two or more	38	30.824	7.176	51.495	1.671
Not at all	None	56	79.277	-23.277	541.819	6.835
Not at all	One	30	33.992	-3.992	15.936	0.469
Not at all	Two or more	52	24.731	27.269	743.598	30.067
						$\Sigma = 61.532$

The Use of Chi-Square When Samples Are Small: A Final Note

The chi-square test is often used by criminal justice researchers. However, it has a very important limitation in its application to studies with small or highly skewed samples. When more than one in five (20%) of the cells in your table has an expected frequency of five or less, it is generally considered inappropriate to use a chi-square test. In such situations, it is recommended that you combine categories of your variables until you meet the minimum expected-frequencies requirement.⁸

Chapter Summary

Whereas the binomial distribution is relevant only for binary variables, the **chi-square distribution** can be used to examine a variable with more than two categories.

The shape of the chi-square distribution chosen depends on the **degrees of freedom** associated with the test. The formula for degrees of freedom defines how many categories would have to be known for us to be able to predict the remaining categories with certainty. The greater the number of degrees of freedom, the flatter the distribution. In practical terms, as the number of degrees of freedom increases, a larger chi-square statistic is required to reject the null hypothesis.

⁸Another alternative solution is to use another group of non-parametric tests, defined as 'exact' tests, to estimate the observed significance level (see Alan Agresti, *Categorical Data Analysis*, New York, John Wiley, 1990). Such tests (e.g., Fisher's Exact Test) which develop

The chi-square test of statistical significance is a nonparametric test. To calculate the test statistic, the researcher must first identify the **observed frequency** and the **expected frequency** of each category. The expected frequencies are those one would expect under the assumption of the null hypothesis. They are distributed in the same proportions as the **marginal frequencies**. The chi-square formula is then applied to each category, or **cell**, in the table. If the observed frequencies differ substantially from the expected frequencies, then the **chi-square statistic** will be large. If the observed frequencies are similar to the expected frequencies, then the chi-square statistic will be small. If the two frequencies are the same, the statistic will be 0. The larger the statistic (and the smaller the number of degrees of freedom), the easier it will be to reject the null hypothesis. The chi-square statistic is always positive. Because the chi-square test relies on nominal nonordered data, it is not concerned with the direction of outcomes.

Key Terms

cells The various entries in a table, each of which is identified by a particular row and column. When we use a table to compare two variables, it is convenient to refer to each combination of categories as a cell.

chi-square distribution A sampling distribution that is used to conduct tests of statistical significance with binary or multicategory nominal variables. The distribution is nonsymmetrical and varies according to degrees of freedom. All the values in the distribution are positive.

chi-square statistic The test statistic resulting from applying the chi-square formula to the observed and expected frequencies for each cell. This statistic tells us how much the observed distribution

differs from that expected under the null hypothesis.

degrees of freedom A mathematical index that places a value on the extent to which a particular operation is free to vary after certain limitations have been imposed. Calculating the degrees of freedom for a chi-square test determines which chi-square probability distribution we use.

expected frequency The number of observations one would predict for a cell if the null hypothesis were true.

marginal The value in the margin of a table that totals the scores in the appropriate column or row.

observed frequency The observed result of the study, recorded in a cell.

Symbols and Formulas

χ^2 Chi-square

df Degrees of freedom

← a sampling distribution for each problem examined, have been made more practical with the advent of powerful Computers. SPSS provides exact tests for two by two tables as Computational options for cross tabulations. A few statistical programs (e.g., SAS) have begun to provide exact test options for larger tables.

f_o Observed frequency

f_e Expected frequency

c Number of columns

r Number of rows

k Number of categories

To determine the degrees of freedom for a chi-square test including only one variable:

$$df = k - 1$$

To determine the degrees of freedom for a chi-square test including two variables:

$$df = (r - 1)(c - 1)$$

To determine the chi-square statistic for one variable:

$$\chi^2 = \sum_{i=1}^k \frac{(f_o - f_e)^2}{f_e}$$

To determine the chi-square statistic for two variables:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_o - f_e)^2}{f_e}$$

Exercises

- 9.1 Local community leaders are concerned about the distribution of homicides in their small town. The local police department broke the city into six recognizable neighborhoods of the same size and discovered the following distribution of homicides:

Neighborhood	Number of Homicides
A	14
B	9
C	17
D	3
E	7
F	10

Community leaders would like to know whether the homicides are randomly distributed across these six neighborhoods.

- a. Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. What can you conclude about the distribution of homicides across these six neighborhoods?
- 9.2 Sergeant Bob is in charge of the duty roster at Gatley police station. Every week, it is his responsibility to randomly assign the five beat officers, including his son Bob Jr., to patrol in each of the five zones that make up the city of Gatley. Zones A and D are favored by all the officers because they are usually quiet. Of the others, Zone C is notoriously dangerous. The officers have recently begun to suspect Sergeant Bob of favoritism toward his son. In the last 30 weeks, Bob Jr. has been assigned to Zone A 12 times, Zone B and Zone C 2 times each, Zone D 9 times, and Zone E 5 times.
- a. Do the other officers have reason to believe that Sergeant Bob is not assigning zones in a random manner? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. Would your answer be any different if a 1% level of significance were used?
- 9.3 In the past 100 years, there have been more than 250 successful breakouts from Didsbury Prison. Mike is a researcher who has been hired by the prison governor to investigate the phenomenon. Details are available only for those breakouts that took place in the past ten years—a total of 30. Using the records of these 30 breakouts as a sample, Mike decides to break the figures down to see whether breakouts were more common in certain wings of the prison than in others. It transpires that of the 30 breakouts, 4 have been from A-Wing, 8 from B-Wing, 15 from C-Wing, and 3 from D-Wing.
- a. Does Mike have enough evidence to conclude that, over the 100-year period, breakouts were more (or less) likely to occur from certain wings than from others? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. Would your answer be any different if a 1% level of significance were used?
 - c. Are there any problems with Mike's choice of a sample? Explain your answer.
- 9.4 A study of death penalty cases (all first-degree murder charges with aggravating circumstances) revealed the following relationship between the victim's race and the chances the offender was sentenced to death: In 100 cases involving white victims, 20 offenders were sen-

- tenced to death. In 100 cases involving black victims, 10 offenders were sentenced to death.
- a. Is there a relationship between the race of the victim and the likelihood an offender was sentenced to death? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. Would your answer be different if a 1% level of significance were used?
- 9.5 At a local school, 46 children were accused of cheating on exams over the course of a semester. In an innovation, the principal decided that every second child accused of cheating would be brought before a “peer jury” to decide guilt or innocence. In all other cases, the decision would be made by the examiners as usual. Of the 30 children who were adjudged guilty over the course of the semester, 18 were convicted by the peer jury, and the rest were convicted by the examiners. Of the children who were adjudged not guilty, 5 were acquitted by their peers.
- a. The principal is mainly interested in the educational value of the experiment, but he will discontinue it if it becomes clear that the peer jury and the examiners make different decisions to a statistically significant degree. He is willing to take a 5% risk of error. Should the scheme be continued? Outline each of the steps of a test of statistical significance.
 - b. Could the principal base the test on a directional hypothesis? If so, what would that hypothesis be, and would it make a difference in his final decision?
- 9.6 In the course of a year, Jeremy, a law student, observed a total of 55 cases in which an accused male pleaded guilty to a serious traffic offense. He observed that of the 15 who were sentenced to prison, 6 wore a shirt and tie in court. Of the 40 who were not sentenced to prison, 8 wore a shirt and tie in court.
- a. Can Jeremy conclude that there is a link between the physical appearance of the accused and whether he is imprisoned? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. What level of significance would be required for his decision to be reversed?
- 9.7 Sasha was interested in the extent to which people are prepared to intervene to help a stranger and whether the race of the stranger is relevant to the likelihood of intervention. She hired four male actors: one of African ancestry, one of Asian ancestry, one of European ancestry, and one of Indian ancestry. The actors were each told to fake a fall in a busy shopping street and pretend to be in some pain. Sasha

observed from nearby and recorded whether, within five minutes of the actor's fall, anyone had stopped to see if he was okay. Each actor repeated the experiment 40 times.

The results were as follows:

Ancestry	√	×
African	4	36
Asian	0	40
European	20	20
Indian	8	32

(√ = Intervention within 5 mins; × = no intervention)

- Can Sasha conclude that there is a link between race of victim and readiness to intervene? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - Would your answer be any different if a 1% level of significance were used?
- 9.8 Dave takes a random sample of the speeches, interviews, and official statements given by the prime minister and the interior minister of a particular country over the course of a year in which reference is made to "prison policy." He analyzes the content of the statements in his sample and discovers five different types of justification for the government's prison policy. Dave then records each time the prime minister or interior minister refers to any of the five justification types. The results are as follows:

Justification Type	Prime Minister	Interior Minister
Incapacitation or protecting society	6	16
Specific deterrence	2	14
General deterrence	4	20
Rehabilitation	0	15
Retribution	13	10

- Is there a statistically significant difference between the policy statements of the prime minister and those of the interior minister? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - Would your answer be any different if a 1% level of significance were used?
- 9.9 The Television Complaints Board monitors the standards of morality for a nation's TV channels. It has recently set up a telephone hotline for viewers who wish to complain about sex, violence, or foul language on any of the nation's three TV channels. In its first month of operation, the board received the following complaints:

	Channel 1	Channel 2	Channel 3
Sex	2	8	10
Violence	10	12	10
Foul language	3	10	15

- a. Which of the following questions would a chi-square test of these results seek to answer?
 - i. Is there a statistically significant difference between the number of complaints made against each channel?
 - ii. Is there a statistically significant difference between the number of each type of complaint made?
 - iii. Is there a statistically significant difference between the types of different complaints received about the three different stations?
 - b. Answer the question you chose in part a by running a chi-square test at a 5% level of significance. Should the null hypothesis be overturned?
- 9.10 A survey of public opinion about the criminal justice system asked respondents to complete the following statement: “The criminal justice system treats offenders. . . .” The researchers found the following distribution of responses by gender of the respondent:

Gender of Respondent	Too Lenient	About Right	Too Harsh
Female	15	50	35
Male	40	35	25

- a. Is there a relationship between the gender of the respondent and perceptions of punishment severity? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
 - b. Would your answer be different if the significance level were 0.01?
 - c. What can you conclude about the relationship between gender and perceptions of punishment severity?
- 9.11 A researcher is interested in the link between the type of offense a defendant is charged with and the manner in which a conviction is obtained. An examination of court records of a random sample of convicted offenders reveals the following distribution of cases:

Type of Charge Offense	How Conviction Was Obtained		
	Jury Trial	Bench Trial	Guilty Plea
Violent	19	13	67
Property	5	8	92
Drug	8	11	83
Other	10	6	74

- a. Is there a relationship between type of charge offense and method of conviction? Use a 5% level of significance and outline each of the steps required in a test of statistical significance.
- b. Would your answer be any different if a 1% level of significance were used?
- c. What can you conclude about the relationship between type of charge offense and method of conviction?

Computer Exercises

Up to this point, all of our computer examples have relied on individual-level data, where each line of data in the spreadsheet view, whether SPSS or Stata, referred to one case. However, there are many other times where we have been given data in tabular form, say from a government publication, and we wish to perform some analysis on that data. In these instances, we may have data on several hundred, or even thousand, cases, but the data appear in aggregated form. All of the examples in this chapter take that form, too.

Prior to discussing how to obtain chi-square values to test for a relationship between two categorical variables, we illustrate how to enter tabular data into the software program.

SPSS

Entering Tabular Data

The most direct way of entering tabular data is to think of the rows and columns in a table as two separate variables and the number of cases in each cell of the table as a third variable (representing a count of the number of cases with the combination of the values for the two variables). For example, consider the data presented in [Table 9.7](#):

	Imprisoned	Not Imprisoned	Row Margin
Subsequently Arrested	33	19	52
Not Subsequently Arrested	67	48	115
Column Margin	100	67	167

We can enter the information from this table as three variables: the row, the column, and the count. Note that as you enter data into SPSS, the default variable names will be VAR00001, VAR00002, and VAR00003. There are two equivalent ways of renaming these variables in SPSS. We will rename the row variable from VAR00001 to *arrested* (to represent whether or not the person was subsequently arrested), the column variable from VAR00002 to *prison* (to represent whether or not the person had been imprisoned), and the number of cases in each cell from VAR00003 to *count*. The equivalent syntax for renaming these variables is:

```
RENAME VARIABLES VAR00001 = arrest VAR00002 = prison
VAR00003 = count.
```

Or,

```
RENAME VARIABLES (VAR00001 VAR00002 VAR00003 = arrest
prison count).
```

After entering these data into SPSS and changing the variable names, you would have three variables and four lines of data like the following (with the variable names appearing at the top of each column of data):

arrest	prison	count
1.0000	1.0000	33.0000
1.0000	2.0000	19.0000
2.0000	1.0000	67.0000
2.0000	2.0000	48.0000

If you were to begin working with SPSS at this point, you would not obtain the correct results, since SPSS will treat these data as representing only *four* observations (You may want to confirm this for yourself—run the FREQUENCIES command on either arrest or prison and see how many observations SPSS thinks it is working with.).

To have SPSS recognize the 167 observations represented by these four lines, you need to tell SPSS to weight the cases by the “count” variable. This is done with the WEIGHT command:

```
WEIGHT BY variable_name.
```

For our example, this would be:

```
WEIGHT BY count.
```

The WEIGHT command tells SPSS that you have entered data in tabular form.

Obtain the Chi-Square Statistic

The chi-square statistic is obtained by using the CROSSTABS command. The basic format of the CROSSTABS command is

```
CROSSTABS
  /TABLES = row_variable BY column_variable
  /STATISTICS = CHISQ.
```

To continue our example using the data from [Table 9.7](#), we would run the following command:

```
CROSSTABS
  /TABLES = arrest BY prison
  /STATISTICS = CHISQ.
```

The output produced by executing this command will contain a cross-tabulation of the data that should be identical to [Table 9.7](#). Immediately below this table will be another table labeled “Chi-square tests.” The “Pearson Chi-Square” is the name of the chi-square statistic that you have learned to calculate in this chapter. The value reported by SPSS is 0.403, which differs from the value reported above (0.402) by 0.001, which can be attributed to rounding error above.

All of the commands discussed above are illustrated in the SPSS syntax file `Chapter_9.sps`.

Stata

A similar process is used in Stata to enter the data from [Table 9.7](#). Recall that to enter data, you will need to click on the “Data Editor” button at the top center of the Stata window. Note that as you enter the data, the default variable names are `var1`, `var2`, and `var3`. To change these to `arrest`, `prison`, and `count`, respectively, enter the following commands:

```
rename var1 arrest
```

```
rename var2 prison
```

```
rename var3 count
```

Alternatively, you could edit the variable name in the far right box in the Stata window (be sure to click on the lock symbol to “unlock” the variable properties and allow you to make changes to the variable name).

In contrast to SPSS, there is no direct weighting of the data directly. Rather, Stata includes an option in nearly all of its statistical procedures that will allow the user to include a variable that represents frequency counts, as we have in our example.

To obtain the chi-square statistic and a two-way cross-tabulation, we enter the following command:

```
tabulate row_variable column_variable, chi2
```

Note that we have assumed we are not working with tabular data in this example, but individual-level data as we have in all previous chapters. The addition of the **chi2** option after the comma requests the chi-square statistic.

For our example, we would enter the following command:

```
tabulate arrest prison [fweight=count], chi2
```

We have added the **fweight**= option [in required brackets] to indicate that we have a frequency weight variable named `count` that indicates the number of cases with each characteristic of the two variables.

The output from running this command includes a cross-tabulation that should be identical to [Table 9.7](#). Just below the table is the chi-square statistic,

reported as “Pearson $\chi^2(1) = 0.4031$.” The value “1” in the parentheses refers to the number of degrees of freedom for the chi-square statistic. This is a common reporting format in Stata output for a chi-square statistics—the degrees of freedom associated with the test will be listed in the adjacent parentheses.

All of these commands are illustrated in the Stata do file `Chapter_9.do`.

Problems

1. Input the data on race and cell-block assignment from [Table 9.11](#). Compute the value of the chi-square statistic for these data. How does it compare to the value reported in the text?
2. Input the data on affectional identification with father and delinquency from [Table 9.14](#). Compute the value of the chi-square statistic for these data. How does it compare to the value reported in the text?
3. Enter the data from Exercise 9.7. Compute the value of the chi-square statistic for these data. How does it compare with the value that you have calculated for this exercise?
4. Enter the data from Exercise 9.11. Compute the value of the chi-square statistic for these data. How does it compare with the value that you have calculated for this exercise?
5. Open the NYS data file (`nys_1.sav`, `nys_1_student.sav`, or `nys_1.dta`). Use a 5% level of significance and outline each of the steps required in a test of statistical significance for each of the following possible relationships:
 - a. Is ethnicity related to grade point average?
 - b. Is marijuana use among friends related to the youth’s attitudes about marijuana use?
 - c. Is the importance of going to college related to the importance of having a job?
 - d. Is grade point average related to the importance of having a job?
 - e. Is the sex of the youth related to the importance of having friends?
 - f. Is the importance of having a job related to the youth’s attitudes about marijuana use?
 - g. SPSS notes at the bottom of each cross-tabulation the number and percentage of all cells that had expected frequencies less than 5. For parts a through f, are there any cross-tabulations that produce expected frequencies of less than 5 for 20 % or more of all cells in the table? If so, what are the consequences for interpreting the chi-square statistic? Explain how the categories of one or more variables could be combined to produce a table that has fewer cells with expected frequencies of less than 5.

6. Open the data file `pcs_98.sav`. These data represent a random sample of 1,300 offenders convicted of drug, property, or violent offenses and sentenced in Pennsylvania in 1998. The full data file contains information on all offenders (more than 40,000) who were convicted of misdemeanor and felony offenses and sentenced to some form of punishment in Pennsylvania in 1998. The full data file is available through the National Archive of Criminal Justice Data at <http://www.icpsr.umich.edu/NACJD> (Prior to answering the following questions, you may find it helpful to review the list of variables and how each is coded.). Use a 5% level of significance and outline each of the steps required in a test of statistical significance for each of the following relationships:
 - a. Is the sex of the offender related to the method of conviction (i.e., plea, bench trial, or jury trial)?
 - b. Is the race-ethnicity of the offender related to whether the offender was incarcerated or not?
 - c. Is the method of conviction related to the type of punishment received (i.e., probation, jail, or prison)?
 - d. Is the type of conviction offense (i.e., drug, property, or violent) related to the method of conviction?