Contents

Chapter 10

Chi Square Tests

10.1 Introduction

The statistical inference of the last three chapters has concentrated on statistics such as the mean and the proportion. These summary statistics have been used to obtain interval estimates and test hypotheses concerning population parameters. This chapter changes the approach to inferential statistics somewhat by examining whole distributions, and the relationship between two distributions. In doing this, the data is not summarized into a single measure such as the mean, standard deviation or proportion. The whole distribution of the variable is examined, and inferences concerning the nature of the distribution are obtained.

In this chapter, these inferences are drawn using the chi square distribution and the chi square test. The first type of chi square test is the goodness of fit test. This is a test which makes a statement or claim concerning the nature of the distribution for the whole population. The data in the sample is examined in order to see whether this distribution is consistent with the hypothesized distribution of the population or not. One way in which the chi square goodness of fit test can be used is to examine how closely a sample matches a population. In Chapter 7, the representativeness of a sample was discussed in Examples **??** through **??**. At that point, hypothesis testing had not yet been discussed, and there was no test for how well the characteristics of a sample matched the characteristics of a population. In this chapter, the chi square goodness of fit test can be used to provide a test for the representativeness of a sample.

The second type of chi square test which will be examined is the chi square test for independence of two variables. This test begins with a cross classification table of the type examined in Section 6.2 of Chapter 6. There these tables were used to illustrate conditional probabilities, and the independence or dependence of particular events. In Chapter 6, the issue of the independence or dependence of the variables as a whole could not be examined except by considering all possible combinations of events, and testing for the independence of each pair of these events.

In this chapter, the concept of independence and dependence will be extended from events to variables. The chi square test of independence allows the researcher to determine whether variables are independent of each other or whether there is a pattern of dependence between them. If there is a dependence, the researcher can claim that the two variables have a statistical relationship with each other. For example, a researcher might wish to know how the opinions of supporters of different political parties vary with respect to issues such as taxation, immigration, or social welfare. A table of the distribution of the political preferences of respondents cross classified by the opinions of respondents, obtained from a sample, can be used to test whether there is some relationship between political preferences and opinions more generally.

The chi square tests in this chapter are among the most useful and most widely used tests in statistics. The assumptions on which these tests are based are minimal, although a certain minimum sample size is usually required. The variables which are being examined can be measured at any level, nominal, ordinal, interval, or ratio. The tests can thus be used in most circumstances. While these tests may not provide as much information as some of the tests examined so far, their ease of use and their wide applicability makes them extremely worthwhile tests.

In order to lay a basis for these tests, a short discussion of the chi square distribution and table is required. This is contained in the following section. Section 10.3 examines the chi square goodness of fit test, and Section 10.4 presents a chi square test for independence of two variables.

10.2 The Chi Square Distribution

The chi square distribution is a theoretical or mathematical distribution which has wide applicability in statistical work. The term 'chi square' (pronounced with a hard 'ch') is used because the Greek letter χ is used to define this distribution. It will be seen that the elements on which this distribution is based are squared, so that the symbol χ^2 is used to denote the distribution.

An example of the chi squared distribution is given in Figure 10.1. Along the horizontal axis is the χ^2 value. The minimum possible value for a χ^2 variable is 0, but there is no maximum value. The vertical axis is the probability, or probability density, associated with each value of χ^2. The curve reaches a peak not far above 0, and then declines slowly as the χ^2 value increases, so that the curve is asymmetric. As with the distributions introduced earlier, as larger χ^2 values are obtained, the curve is asymptotic to the horizontal axis, always approaching it, but never quite touching the axis.

Each χ^2 distribution has a degree of freedom associated with it, so that there are many different chi squared distributions. The chi squared distributions for each of 1 through 30 degrees of freedom, along with the distributions for 40, 50 , . . . , 100 degrees of freedom, are given in Appendix **??**.

The χ^2 distribution for 5 degrees of freedom is given in Figure 10.1. The total area under the whole χ^2 curve is equal to 1. The shaded area in this figure shows the right 0.05 of the area under the distribution, beginning at $\chi^2 = 11.070$. You will find this value in the table of Appendix **??** in the fifth row (5 df) and the column headed 0.05. The significance levels are given across the top of the χ^2 table and the degrees of freedom are given by the various rows of the table.

The chi square table is thus quite easy to read. All you need is the degree of freedom and the significance level of the test. Then the critical χ^2 value can be read directly from the table. The only limitation is that you are restricted to using the significance levels and degrees of freedom shown in the table. If you need a different level of significance, you could try interpolating between the values in the table.

The Chi Square Statistic. The χ^2 statistic appears quite different from the other statistics which have been used in the previous hypotheses tests. It also appears to bear little resemblance to the theoretical chi square distribution just described.

For both the goodness of fit test and the test of independence, the chi square statistic is the same. For both of these tests, all the categories into which the data have been divided are used. The data obtained from the sample are referred to as the **observed** numbers of cases. These are the frequencies of occurrence for each category into which the data have been

Figure 10.1: χ^2 Distribution with 5 Degrees of Freedom

grouped. In the chi square tests, the null hypothesis makes a statement concerning how many cases are to be **expected** in each category if this hypothesis is correct. The chi square test is based on the difference between the observed and the expected values for each category.

The chi square statistic is defined as

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed number of cases in category i, and E_i is the expected number of cases in category i. This chi square statistic is obtained by calculating the difference between the observed number of cases and the expected number of cases in each category. This difference is squared and divided by the expected number of cases in that category. These values are then added for all the categories, and the total is referred to as the chi squared value.

> **Chi Square Calculation**
>
> Each entry in the summation can be referred to as "The observed minus the expected, squared, divided by the expected." The chi square value for the test as a whole is "The sum of the observed minus the expected, squared, divided by the expected."

The null hypothesis is a particular claim concerning how the data is distributed. More will be said about the construction of the null hypothesis later. The null and alternative hypotheses for each chi square test can be stated as

$$H_0 : O_i = E_i$$
$$H_1 : O_i \neq E_i$$

If the claim made in the null hypothesis is true, the observed and the expected values are close to each other and $O_i - E_i$ is small for each category. When the observed data does not conform to what has been expected on the basis of the null hypothesis, the difference between the observed and expected values, $O_i - E_i$, is large. The chi square statistic is thus small when the null hypothesis is true, and large when the null hypothesis is not true. Exactly how large the χ^2 value must be in order to be considered large enough to reject the null hypothesis, can be determined from the level of significance and the chi sqaure table in Appendix **??**. A general formula for determining the degrees of freedom is not given at this stage, because this differs for the two types of chi square tests. In each type of test though, the degrees of freedom is based on the number of categories which are used in the calculation of the statistic.

The chi square statistic, along with the chi square distribution, allow the researcher to determine whether the data is distributed as claimed. If the chi square statistic is large enough to reject H_0, then the sample provides evidence that the distribution is not as claimed in H_0. If the chi square statistic is not so large, then the researcher may have insufficient evidence to reject the claim made in the null hypothesis.

The following paragraphs briefly describe the process by which the chi square distribution is obtained. You need not read the following page, and can go directly to the goodness of fit test in Section 10.3. But if you do read the following paragraphs, you should be able to develop a better understanding of the χ^2 distribution.

Derivation of the Distribution. In mathematical terms, the χ^2 variable is the sum of the squares of a set of normally distributed variables. Imagine a standardized normal distribution for a variable Z with mean 0 and standard deviation 1. Suppose that a particular value Z_1 is randomly selected from this distribution. Then suppose another value Z_2 is selected from the same standardized normal distribution. If there are d degrees of freedom, then let this process continue until d different Z values are selected from this distribution. The χ^2 variable is defined as the sum of the squares of these Z values. That is,

$$\chi^2 = Z_1^2 + Z_2^2 + Z_3^2 + \cdots + Z_d^2$$

Suppose this process of independent selection of d different values is repeated many times. The variable χ^2 will vary, because each random selection from the normal distribution will be different. Note that this variable is a **continuous** variable since each of the Z values is continuous.

This sum of squares of d normally distributed variables has a distribution which is called the χ^2 distribution with d degrees of freedom. It can be shown mathematically that the mean of this χ^2 distribution is d and the standard deviation is $2d$.

An intuitive idea of the general shape of the distribution can also be obtained by considering this sum of squares. Since χ^2 is the sum of a set of squared values, it can never be negative. The minimum chi squared value would be obtained if each $Z = 0$ so that χ^2 would also be 0. There is no upper limit to the χ^2 value. If all the Z values were quite large, then χ^2 would also be large. But note that this is not too likely to happen. Since large Z values are distant from the mean of a normal distribution, these large Z values have relatively low probabilities of occurrence, also implying that the probability of obtaining a large χ^2 value is low.

Also note that as there are more degrees of freedom, there are more squared Z values, and this means larger χ^2 values. Further, since most values under the normal curve are quite close to the center of the distribution, within 1 or 2 standard deviations of center, the values of χ^2 also tend to be concentrated around the mean of the χ^2 distribution. The chi square distribution thus has the bulk of the cases near the center of the distribution, but it is skewed to the right. It has a lower limit at 0, and declines as χ^2 increases to the right, in the general shape of Figure 10.1.

While the above provides a general description of the theoretical χ^2 distribution of Appendix **??**, this may seem to be quite different than the chi square statistic:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

The theoretical distribution is a continuous distribution, but the χ^2 statistic is obtained in a discrete manner, on the basis of discrete differences between the observed and expected values. As will be seen on page 717, the probabilities of obtaining the various possible values of the chi square statistic can be approximated with the theoretical χ^2 distribution of Appendix **??**.

The condition in which the chi square statistic is approximated by the theoretical chi square distribution, is that the sample size is reasonably large. The rule concerning the meaning of large sample size is different for the χ^2 test than it is for the approximations which have been used so far in this textbook. The rough rule of thumb concerning sample size is that there should be at least 5 expected cases per category. Some modifications of this condition will be discussed in the following sections.

10.3 Goodness of Fit Test

This section begins with a short sketch of the method used to conduct a chi square goodness of fit test. An example of such a test is then provided, followed by a more detailed discussion of the methods used. Finally, there is a variety of examples used to illustrate some of the common types of goodness of fit tests.

The chi square goodness of fit test begins by hypothesizing that the distribution of a variable behaves in a particular manner. For example, in order to determine daily staffing needs of a retail store, the manager may wish to know whether there are an equal number of customers each day of the week. To begin, an hypothesis of equal numbers of customers on each day could be assumed, and this would be the null hupothesis. A student may observe the set of grades for a class, and suspect that the professor allocated the grades on the basis of a normal distribution. Another possibility is that a researcher claims that the sample selected has a distribution which is very close to distribution of the population. While no general statement can be provided to cover all these possibilities, what is common to all is that a claim has been made concerning the nature of the whole distribution.

Suppose that a variable has a frequency distribution with k categories into which the data has been grouped. The frequencies of occurrence of the variable, for each category of the variable, are called the **observed** values. The manner in which the chi square goodness of fit test works is to determine how many cases there would be in each category if the sample data were distributed exactly according to the claim. These are termed the **expected** number of cases for each category. The total of the expected number of cases is always made equal to the total of the observed number of cases. The null hypothesis is that the observed number of cases in each category is exactly equal to the expected number of cases in each category. The alternative hypothesis is that the observed and expected number of cases differ sufficiently to reject the null hypothesis.

Let O_i is the observed number of cases in category i and E_i is the expected number of cases in each category, for each of the k categories $i = 1, 2, 3, \cdots, k$, into which the data has been grouped. The hypotheses are

$$H_0 : O_i = E_i$$

$$H_1 : O_i \neq E_i$$

and the test statistic is

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where the summation proceeds across all k categories, and there are $k - 1$ degrees of freedom. Large values of this statistic lead the researcher to reject the null hypothesis, smaller values mean that the null hypothesis cannot be rejected. The chi square distribution with $k - 1$ degrees of freedom in Appendix **??** is used to determine the critical χ^2 value for the test.

This sketch of the test should allow you to follow the first example of a chi square goodness of fit test. After Example 10.3.1, the complete notation for a chi square goodness of fit test is provided. This notation will become more meaningful after you have seen how a goodness of fit test is conducted.

Example 10.3.1 Test of Representativeness of a Sample of Toronto Women

In Chapter 7, a sample of Toronto women from a research study was examined in Example **??**. *Some of this data is reproduced here in Table 10.1.*

This data originally came from the article by Michael D. Smith, "Sociodemographic risk factors in wife abuse: Results from a survey of Toronto women," **Canadian Journal of Sociology** *15 (1), 1990, page 47. Smith claimed that the sample showed*

> a close match between the age distributions of women in the sample and all women in Toronto between the ages of 20 and 44. This is especially true in the youngest and oldest age brackets.

Age	Number in Sample	Per Cent in Census
20-24	103	18
25-34	216	50
35-44	171	32
Total	490	100

Table 10.1: Sample and Census Age Distribution of Toronto Women

Using the data in Table 10.1, conduct a chi square goodness of fit test to determine whether the sample does provide a good match to the known age distribution of Toronto women. Use the 0.05 level of significance.

Solution. *Smith's claim is that the age distribution of women in the sample closely matches the age distribution of all women in Toronto. The alternative hypothesis is that the age distribution of the women in the sample do not match the age distribution of all Toronto women. The hypotheses for this goodness of fit test are:*

H_0: The age distribution of respondents in the sample is the same as the age distribution of Toronto women, based on the Census.
H_1: The age distribution of respondents in the sample differs from the age distribution of women in the Census.

The frequencies of occurrence of women in each age group are given in the middle column of Table 10.1. These are the observed values O_i, and there are $k = 3$ categories into which the data has been grouped so that $i = 1, 2, 3$.

Suppose the age distribution of women in the sample conforms exactly with the age distribution of all Toronto women as determined from the Census. Then the expected values for each category, E_i, could be determined. With these observed and expected numbers of cases, the hypotheses can be written

$$H_0 : O_i = E_i$$

$$H_1 : O_i \neq E_i$$

For these hypotheses, the test statistic is

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where the summation proceeds across all k categories and there are $d = k - 1$ degrees of freedom.

The chi square test should always be conducted using the actual number of cases, rather than the percentages. These actual numbers for the observed cases, O_i, are given as the frequencies in the middle column of Table 10.1. The sum of this column is the sample size of $n = 490$.

The expected number of cases for each category is obtained by taking the total of 490 cases, and determining how many of these cases there would be in each category if the null hypothesis were to be exactly true. The percentages in the last column of Table 10.1 are used to obtain these. Let the 20-24 age group be category $i = 1$. The 20-24 age group contains 18% of all the women, and if the null hypothesis were to be exactly correct, there would be 18% of the 490 cases in category 1. This is

$$E_1 = 490 \times \frac{18}{100} = 490 \times 0.18 = 88.2$$

For the second category 25-34, there would be 50% of the total number of cases, so that

$$E_2 = 490 \times \frac{50}{100} = 490 \times 0.50 = 245.0$$

cases. Finally, there would be

$$E_3 = 490 \times \frac{32}{100} = 490 \times 0.32 = 156.8$$

cases in the third category, ages 35-44.

Both the observed and the expected values, along with all the calculations for the χ^2 statistic are given as in Table 10.2. Note that the sum of *each of the observed and expected numbers of cases totals 490. The difference between the observed and expected numbers of cases is given in the column labelled $O_i - E_i$, with the squares of these in the column to the right. Finally, these squares of the observed minus expected numbers of cases divided by the expected number of cases is given in the last column. For example, for the first row, these calculations are*

$$\frac{(O_i - E_i)^2}{E_i} = \frac{(103 - 88.2)^2}{88.2} = \frac{14.8^2}{88.2} = \frac{219.04}{88.2} = 2.483.$$

The other rows are determined in the same manner, and this table provides all the material for the chi square goodness of fit test.

Category	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	103	88.2	14.8	219.04	2.483
2	216	245.0	-29.0	841.00	3.433
3	171	156.8	14.2	201.64	1.286
Total	490	490.0	0.0		7.202

Table 10.2: Goodness of Fit Calculations for Sample of Toronto Women

The sum of the last column of the table gives the required χ^2 statistic. This is

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} \\ &= 2.483 + 3.433 + 1.286 \\ &= 7.202 \end{aligned}$$

The next step is to decide whether this is a large or a small χ^2 value. The level of significance requested is the $\alpha = 0.05$ level. The number of degrees of freedom is the number of categories minus one. There are $k = 3$ categories into which the ages of women have been grouped, so that there are $d = k - 1 = 3 - 1 = 2$ degrees of freedom. From the 0.05 column and the second row of the χ^2 table in Appendix **??**, *the critical value is $\chi^2 = 5.991$. This means that there is exactly 0.05 of the area under the curve to the right of $\chi^2 = 5.991$. A chi square value larger than this leads to rejection of* *the null hypothesis, and a chi square value from the data which is smaller than 5.991 means that the null hypothesis cannot be rejected.*

The data yields a value for the chi squared statistic of 7.202 and this exceeds 5.991. Since $7.202 > 5.991$ *the null hypothesis can be rejected, and the research hypothesis accepted at the 0.05 level of significance.*

The statement of the author that the sample is a close match of the age distribution of all women in Toronto is not exactly justified by the data presented in Table 10.1. Of course, there is a chance of Type I error, the error of rejecting a correct null hypothesis. In this case, the sampling method may generally be a good method, and it just happens that the set of women selected in this sample ends up being a bit unrepresentative of the population of all women in Toronto. If this is so, then there is Type I error. However, the chance of this is at most 0.05. Since this is quite small, the test shows that the sample is representative of all Toronto women.

Additional Comments. *There are a few parts of the calculations which should be noted. First, note the differences between the observed and expected values in the middle of Table 10.2. The largest difference is -29.0, for group 2, the 25-34 age group. This group appears to be considerably underrepresented in the sample, with there being 29 less observed than expected cases. As can be noted in the last column of the table, this group contributes 3.433 to the* χ^2 *total of 7.202, the largest single contribution.*

For the other two age groups, the discrepancy between the observed and expected is not so large. Each of the first and third groups is somewhat overrepresented in the sample, compared with the Census distribution. But the smaller chi squared values for these first and third groups, 2.483 and 1.286 (in the last column), means that these groups are not overrepresented as much as the middle aged group is underrepresented. If the sample were to be adjusted to make it more representative, the sample of 25-34 year old women should be boosted.

Note also that the sum of the $O_i - E_i$ *column is 0. This is because the differences between the observed and expected numbers of cases are sometimes positive and sometimes negative, but these positives and negatives cancel if the calculations are correct.*

Finally, an appreciation of the meaning of the degrees of freedom can be obtained by considering how the expected values are constructed. Note that the expected values are calculated by using the percentage of cases that are in each category, in order to determine the expected numbers of cases. After the expected values for each of the first two categories have *been determined, the number of expected cases in the third category can be determined by subtraction. That is, if there are 88.2 and 245.0 expected cases in the first two categories, this produces a total of* $88.2 + 245.0 = 333.2$ *expected cases. In order to have the sum of the expected cases equal 490, this means that there must be* $490.0 - 333.2 = 156.8$ *cases in the third category. In terms of the degrees of freedom, this means that there are only two degrees of freedom. Once values for two of the categories have been specified, the third category is constrained by the fact that the sum of the observed and expected values must be equal.*

Notation for the Chi Square Goodness of Fit Test. As noted earlier, there is no general formula which can be given to cover all possible claims which could be made in a null hypothesis. In the examples which follow, you will see various types of claims, each of which required a slightly different method of determining the expected numbers of cases in each category. However, notation can be provided to give a more complete explanation of the test.

Begin with k categories into which a frequency distribution has been organized. The frequencies of occurrence for the variable are $f_1, f_2, f_3, \cdots, f_k$, but for purposes of carrying out the χ^2 goodness of fit test, these are called the observed values, O_i and given the symbols $O_1, O_2, O_3, \cdots, O_k$. Let the sum of these frequencies or observed values be n, that is,

$$\sum f_i = \sum O_i = n.$$

The expected values for the test must also total n, so that

$$\sum E_i = \sum O_i = n.$$

The expected numbers of cases E_i is determined on the basis of the null hypothesis. That is, some hypothesis is made concerning how the cases should be distributed. Let this hypothesis be that there is a proportion p_i cases in each of the categories i, where $i = 1, 2, 3, \cdots, k$. Since all the cases must be accounted for, the sum of all the proportions must be 1. That is

$$\sum_{i=1}^{k} p_i = 1.$$

The null hypothesis could then be stated as

H_0: The proportion of cases in category 1 is p_1, in category 2 is p_2, ... , in category k is p_k.

Alternatively, this hypothesis could be stated to be:

$$H_0 : \text{The proportion of cases in category } i \text{ is } p_i, \ i = 1, 2, 3, \cdots, k.$$

The alternative hypothesis is that the proportions are not equal to p_i as hypothesized.

In Example 10.3.1 the proportions hypothesized in the null hypothesis are the percentages divided by 100. The hypotheses are

$$H_0 : p_1 = 0.18, p_2 = 0.50, p_3 = 0.32$$

$$H_1 : \text{Proportions are not as given in } H_0.$$

The alternative hypothesis is that at least one of the proportions in the null hypothesis is not as stated there. Note that

$$\sum_{i=1}^{3} p_i = 0.18 + 0.50 + 0.32 = 1.$$

The expected cases are obtained by multiplying these p_i by n. That is, for each category i,

$$E_i = p_i \times n.$$

Since the sum of the p_i is 1, the sum of the E_i is n.

These values of O_i and E_i can now be used to determine the values of the chi squared statistic

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

where there are $d = k - 1$ degrees of freedom.

The only assumption required for conducting the test is that each of the expected numbers of cases is reasonably large. The usual rule of thumb adopted here is that each of the expected cases should be greater than or equal to 5, that is,

$$E_i \geq 5, \text{ for each } i = 1, 2, 3, \cdots, k.$$

This rule may be relaxed slightly if most of these expected values exceed 5. For example, suppose there are 6 categories, and all of the categories have more than 5 expected cases, but one category has only 4.3 expected cases. Unless the chi square value comes out quite close to the borderline of acceptance or not acceptance of the null hypothesis, this should cause little inaccuracy. The larger the expected number of cases, the more accurate the chi squared calculation will be. Also note that this condition is placed on the expected values, not the observed values. There may be no observed values in one or more categories, and this causes no difficulty with the test. But there must be around 5 or more expected cases per category. If there are considerably less than this, adjacent categories may be grouped together to boost the number of expected cases. An example of this is given in Example 10.3.4.

The reason why the number of expected cases must be 5 or more is as follows. As noted earlier, there are really two chi squares which have been given here. The **chi square distribution** is a theoretically derived mathematical distribution, based on the sums of squares of normally distributed variables. This is a continuous distribution. In contrast, the **chi square statistic** is a discrete statistic, based on a finite number of possible values $O_i - E_i$. These differences are used to compute the chi square statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

This statistic can be shown to have what mathematicians refer to as a **multinomial distribution**. This distribution is a generalization of the binomial probability distribution, obtained by examining the distribution of the possible values where there are more than 2 categories. It can also be shown mathematically that the multinomial distribution can be approximated by the theoretical chi square distribution under certain conditions. The condition ordinarily specified by statisticians is that the expected cases in each category be 5 or more. This is large enough to ensure that the discrete multinomial distribution is reasonably closely approximated by the continuous chi square distribution.

Several examples of different types of goodness of fit tests now follow. If you study these examples, this should provide you with the background required to carry out most goodness of fit tests.

Example 10.3.2 Suicides in Saskatchewan

The media often seizes on yearly changes in crime or other statistics. A large jump in the number of murders from one year to the next, or a large *decline in the the support for a particular political party may become the subject of many news reports and analysis. These statistics may be expected to show some shift from year to year, just because there is variation in any phenomenon. This question addresses this issue by looking at changes in the annual number of suicides in the province of Saskatchewan.*

'Suicides in Saskatchewan declined by more than 13% from 1982 to 1983," was the headline in the Saskburg **Times** *in early 1984. The article went on to interview several noted experts on suicide who gave possible reasons for the large decline. As a student who has just completed a statistics course, you come across the article and decide to check out the data. By consulting Table 25 of Saskatchewan Health, Vital Statistics,* **Annual Report***, for various years, you find the figures for 1978-1989. The data is given in Table 10.3.*

Year	Number of Suicides in Saskatchewan
1978	164
1979	142
1980	153
1981	171
1982	171
1983	148
1984	136
1985	133
1986	138
1987	132
1988	145
1989	124

Table 10.3: Suicides in Saskatchewan, 1978 – 1989

Use the χ^2 test for goodness of fit to test the hypothesis that the number of suicides reported for each year from 1978 to 1989 does not differ significantly from an equal number of suicides in each year. (0.05 significance).

Based on this result, what might you conclude about the newspaper report, especially in light of the extra information for 1984 – 1989 that is now available to you? Also comment on any possible errors in the conclusion.

Solution. *The null hypothesis is that there is the same number of suicides each year in the province. While this is obviously not exactly true, the test of this hypothesis will allow you to decide whether the actual number of suicides is enough different from an equal number to reject the null hypothesis.*

The hypotheses can be stated as follows:

H_0: The distribution of the observed number of suicides for each year does not differ significantly from an equal number of suicides in each year.

H_1: The distribution of the observed number of suicides for each year differs significantly from an equal number of suicides in each year.

For purposes of conducting the test, these hypotheses translate into

$$H_0 : O_i = E_i$$

$$H_1 : O_i \neq E_i$$

where the E_i are equal for each year, and sum to the same total number as the total of the observed number of suicides over all these years.

Since we observe a total of 1757 suicides over the 12 years shown here, this means an average of $1757/12 = 146.4$ *suicides each year. If there were an equal number of suicides for each of the 12 years, then this would mean 146.4 each year, and this is the value of the expected number of cases for each year. The* χ^2 *test is then conducted as follows, with the calculations as shown in Table 10.4.*

For the χ^2 *test, the number of degrees of freedom is the number of categories minus one. Here this is* $d = 12 - 1 = 11$ *degrees of freedom. For* $\alpha = 0.05$ *and 11 degrees of freedom, the* χ^2 *value from the table in Appendix* **??** *is 19.675. The region of rejection is all* χ^2 *values of 19.675 or more. In Figure 10.2, the chi square distribution with 11 degrees of freedom and* $\alpha = 0.05$ *is shown. The critical* χ^2 *value is* $\chi^2 = 19.675$*, and 0.05 of the area under the curve lies to the right of this.*

From the set of the observed yearly number of suicides, $\chi^2 = 18.135 < 19.675$*. As can be seen in Figure 10.2,* $\chi^2 = 18.135$ *is not in the critical region for this test. and this means that there is insufficient evidence to reject* H_0 *at the 0.05 level of significance. As a result the null hypothesis that the distribution of the actual number of suicides is not any different than an equal number of suicides each year cannot be rejected at the 0.05 level of significance.*

There is a considerable chance of Type II error here, the error of failing to reject the null hypothesis when it is not correct. It is obvious that the

Figure 10.2: Chi Square Distribution for Test of Equality in the Annual Number of Suicides

number of suicides is not equal for each of these 12 years and as a result, the null hypothesis is not exactly correct. What the chi square goodness of fit test shows is that the null hypothesis is also not all that incorrect. The test shows that there is little difference between the fluctuation in the number of suicides from year to year and an assumption that there was no change in the number of suicides each year. What might be concluded then is that the number of suicides has stayed much the same over these years and that whatever factors determine the suicides on a yearly basis have remained much the same over these years. A certain degree of random fluctuation in the yearly total of suicides is bound to take place even if the basic factors determining the number of suicides does not change. Also note that the population of Saskatchewan changed little over these years, so that fluctuations in the total population cannot be used to explain the yearly variation in suicides.

As a result, not too much should be made of the drop in the number of suicides from 171 to 148, or $(23/171) \times 100\% = 13.5\%$ over one year. The

Year	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1978	164	146.4	17.6	309.76	2.116
1979	142	146.4	-4.4	19.36	0.132
1980	153	146.4	6.6	43.56	0.298
1981	171	146.4	24.6	605.16	4.134
1982	171	146.4	24.6	605.16	4.134
1983	148	146.4	1.6	2.56	0.017
1984	136	146.4	-10.4	108.16	0.739
1985	133	146.4	-13.4	179.56	1.227
1986	138	146.4	-8.4	70.56	0.482
1987	132	146.4	-14.4	207.36	1.416
1988	145	146.4	-1.4	1.96	0.013
1989	124	146.4	-22.4	501.76	3.427
Total	1,757	1,756.8	0.2		$\chi^2 = 18.135$

Table 10.4: Calculations for χ^2 Test

totals often change from year to year even though the underlying causes do not really change. In this case then, no real explanation of changes in the number of suicides need be sought. This change is more likely just the result of random yearly fluctuation.

The data from 1983 to 1989 show that for these seven years for which mortality data is available in the province, the number of suicides is lower than in the early 1980s. Since the recent high point of 171 suicides of 1981 and 1982, the number of suicides has declined considerably. This evidence points in a different direction than indicated by the hypothesis test and may indicate that suicides peaked in the early 1980s, then declined with a new, lower number of suicides being a more normal level in the mid 1980s.

Example 10.3.3 Birth Dates of Hockey Players

The Toronto **Globe and Mail** *of November 27, 1987 contained an article, written by Neil Campbell, entitled "NHL career can be preconceived." In the article, Campbell claimed that the organization of hockey*

has turned half the boys in hockey-playing countries into second class citizens.

The disadvantaged are those unlucky enough to have been born in the second half of the calendar year.

Campbell calls this the **Calendar Effect**, *arguing that it results from the practice of age grouping of very young boys by calendar year of birth. For example, all boys 7 years old in 1991, and born in 1984, would be in the same grouping. By 1991, those boys born in the first few months of 1984 are likely to be somewhat larger and better coordinated than those boys born in the later months of 1984. Yet all these players compete against each other. Campbell argues that this initial advantage stays with these players, and may become a permanent advantage by the time the boys are separated into elite leagues at age 9.*

In order to test whether this Calendar Effect exists among hockey players who are somewhat older, a statistics student collected data from the Western Hockey League (WHL) Yearbook for 1987-88. The birth dates for the players are as shown in Table 10.5.

Quarter	Number of Players
January to March	84
April to June	77
July to September	35
October to December	34

Table 10.5: Distribution of Birth Dates, WHL Players, 1987-88

The results from Table 10.5 seem quite clear, and one may take this as evidence for the Calendar Effect. The proportion of players born in the first half of the year is $(84+77)/230 = 161/230 = 0.70$, *well over half. However, a test to determine whether the distribution in Table 10.5 differs significantly from what would be expected if the birth dates of hockey players match those of the birth dates in the population as a whole should be conducted.*

In order to do this, note first that there is a seasonal pattern to births, and this seasonal pattern should be considered to see whether it accounts for this difference. In order to determine this, the birth dates of births in the four Western provinces of Canada was obtained for the births in the *years 1967-70. This would approximate the dates of the hockey players in the WHL in 1987-88. This data is presented in Table 10.6.*

	Births in 1967-70	
Quarter	Number	Proportion
January to March	97,487	0.242
April to June	104,731	0.260
July to September	103,974	0.258
October to December	97,186	0.241

Table 10.6: Distribution of Births by Quarter, 4 Western Provinces, 1967-70

As can be seen in Table 10.6, while there is a seasonal pattern to births, it does not appear to be very different from an equal number of, or proportion of, births in each quarter. Also, it appears to be very different from the pattern of births by quarter for the hockey players.

In order to conduct a test of whether the distribution of birth dates of hockey players differs from the distribution of the birth dates shown in Table 10.6, conduct a chi square test of goodness of fit. Use 0.001 significance.

Solution.

The hypotheses can be stated as follows:

H_0: The distribution of the observed number of births by quarter does not differ significantly from that of the distribution of births in the Western provinces in 1967-70.

H_1: The distribution of the observed number of births by quarter differs significantly from that of the distribution of births in the Western provinces in 1967-70.

For the χ^2 *test, it is necessary to determine the expected number of births in each quarter, based on* H_0*. In order to do this, use the proportion of births for each quarter in Table 10.6. Then multiply these proportions by the total number of players. This is the total of 230 players in Table 10.5. For example, for January to March, there are 0.242 of total births expected. This amounts to* $0.242 \times 230 = 55.6$ *expected births between January and March. This is done for each of the four quarters, and the results shown* *as the expected E_i column in Table 10.7. The rest of the calculations in Table 10.7 are the calculations required for the χ^2 test.*

i	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	84	55.6	28.4	806.56	14.506
2	77	59.7	17.3	299.29	5.383
3	35	59.3	- 24.3	590.49	9.958
4	34	55.4	- 21.4	457.96	8.266
Total	230	230.0			$\chi^2 = 38.113$

Table 10.7: Calculations for χ^2 Test

For the χ^2 test, the number of degrees of freedom is the number of categories minus one. This is $d = k - 1 = 4 - 1 = 3$ degrees of freedom. For $\alpha = 0.001$ and 3 degrees of freedom, the χ^2 value from the table is 16.266. The region of rejection is all χ^2 values of 16.266 or more. The value from the WHL data is 38.113 and this is well into the region of rejection of the null hypothesis. As a result the null hypothesis can be rejected at the 0.001 level of significance. There appears to be very strong evidence of the Calendar Effect, based on the birth dates of the Western Hockey League for 1987-88. The chance of Type I error, rejecting the null hypothesis when it is really true, is extremely small. There is a probability of less than 0.001 of Type I error.

Goodness of Fit to a Theoretical Distribution. Suppose that a researcher wishes to test the suitability of a model such as the normal distribution or the binomial probability distribution, as an explanation for an observed phenomenon. For example, a grade distribution for grades in a class may be observed, and someone claims that the instructor has allocated the grades on the basis of a normal distribution. The test of the goodness of fit then looks at the observed number of grades in each category, determines the number that would be expected if the grades had been exactly normally distributed, and compares these with the chi square goodness of fit test.

In general this test is the same as the earlier goodness of fit tests except for one alteration. The degrees of freedom for this test must take into account the number of parameters contained in the distribution being fitted to the data. The new formula for the degrees of freedom becomes $d = k-1-r$ where k is the number of categories into which the data has been grouped, and r is the number of parameters contained in the distribution which is being used as the model. In the case of the normal distribution, there are two parameters, μ and σ, so that there are $d = k-1-2 = k-3$ degrees of freedom when an acutal distribution is being compared with a normal distribution with the goodness of fit test. If the suitability of the binomial as an explanation for the data is being requested, there are $d = k-1-1 = k-2$ degrees of freedom. For the binomial, there is only one parameter, the proportion of successes.

Using this modification, the suitability of the normal distribution in explaining a grade distribution is examined in the following example.

Example 10.3.4 Grade Distribution in Social Studies 201

The grade distribution for Social Studies 201 in the Winter 1990 semester is contained in Table 10.8. Along with the grade distribution for Social Studies 201 is the grade distribution for all the classes in the Faculty of Arts in the same semester.

Grade	All Arts (Per Cent)	Social Studies 201 1990 Winter (Number)
Less than 50	8.3	2
50s	15.4	7
60s	24.7	10
70s	30.8	15
80s	17.8	8
90s	3.0	1
Total	100.0	43
Mean	68.2	68.8
Standard Deviation		12.6

Table 10.8: Grade Distributions

Use the data in Table 10.8 to conduct two chi square goodness of fit tests. First test whether the model of a normal distribution of grades adequately explains the grade distribution of Social Studies 201. Then test whether the grade distribution for Social Studies 201 differs from the grade distribution for the Faculty of Arts as a whole. For each test, use the 0.20 level of significance.

Solution. *For the first test, it is necessary to determine the grade distribution that would exist if the grades had been distributed exactly as the normal distribution. The method outlined in Example 6.4.6 on pages 382-386 of Chapter 6 will be used here. That is, the normal curve with mean and standard deviation the same as the actual Social Studies 201 distribution will be used. This means that the grade distribution for the normal curve with mean* $\mu = 68.8$ *and standard deviation* $\sigma = 12.7$ *is used to determine the grade distribution if the grades were normally distributed. The* Z *values were determined as -1.48, -0.69, 0.09, 0.88 and 1.67 for the* X *values 50, 60, 70, 80 and 90, respectively. By using the normal table in Appendix* **??**, *the areas within each range of grades were determined. These are given as the proportions of Table 10.9. These proportions were then multiplied by 43, the total of the observed values, to obtain the expected number of grades in each of the categories into which the grades have been grouped. These are given in the last column of Table 10.9.*

	Normal Distribution of Grades	
Grade	Proportion (p_i)	Number (E_i)
Less than 50	0.0694	3.0
50s	0.1757	7.6
60s	0.2908	12.5
70s	0.2747	11.8
80s	0.1419	6.1
90s	0.0475	2.0
Total	1.000	43.0

Table 10.9: Normal Distribution of Grades

The hypotheses can be stated as follows:

H_0: The distribution of the observed number of students in each grade category does not differ significantly from what would be expected if the grade distribution were exactly normally distributed.

H_1: The distribution of the observed number of students in each category differs significantly from a normal distribution of grades.

In order to apply the χ^2 test properly, each of the expected values should exceed 5. The less than 50 category and the 90s category both have less than 5 expected cases. In this test, the 90s have only 2 expected cases, so this category is merged with the grades in the 80s. For the grades less than 50, even though there are only 3 expected cases, these have been left in a category of their own. While this may bias the results a little, the effect should not be too great. The calculation of the χ^2 statistic is as given in Table 10.10.

Category	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	2	3.0	-1.0	1.00	0.333
2	7	7.6	-0.6	0.36	0.047
3	10	12.5	-2.5	6.25	0.500
4	15	11.8	3.2	10.24	0.868
5	9	8.1	0.9	0.81	0.100
Total	43	43.0	0.0		$\chi^2 = 1.848$

Table 10.10: Test for a Normal Distribution of Grades

For this χ^2 test, the number of degrees of freedom is the number of categories minus one, minus the number of parameters. Here this is $5 - 1 - 2 = 2$ degrees of freedom. For $\alpha = 0.20$ and 2 degrees of freedom, the χ^2 value from the table is 3.219. The region of rejection is all χ^2 values of 3.219 or more. The value of the chi square statistic for this test is 1.848 and this is not in the region of rejection of the null hypothesis. While the grades are not exactly normally distributed, the hypothesis that they are distributed normally cannot be rejected at the 0.20 level of significance.

For the test concerning a difference in the grade distributions of Social

Category	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	2	3.6	-1.6	2.56	0.711
2	7	6.6	0.4	0.16	0.024
3	10	10.6	-0.6	0.36	0.034
4	15	13.2	1.8	3.24	0.245
5	9	9.0	0.0	0.00	0.000
Total	43	43.0	0.0		$\chi^2 = 1.014$

Table 10.11: Test for Difference Between SS201 and Arts Grades

Studies 201 and the distribution for the Faculty of Arts, the hypotheses can be stated as follows:

H_0: The distribution of the observed number of students in each grade category, in Social Studies 201, does not differ significantly from what would be expected if the grade distribution exactly matched that for all Arts grades.

H_1: The distribution of the observed number of students in each category differs significantly from that of all Arts grades.

To obtain the expected number of grades in each category under this null hypothesis, multiply 43 by the percentages from the distribution of grades in the Faculty of Arts. This provides the expected number of cases in each category. For example, in the less than 50 category, there would be expected to be 8.3% of 43, or 3.6 expected cases.

Using the same method, and moving from the lowest to the highest grades, the expected numbers of cases for each category are 3.6, 6.6, 10.6, 13.2, 7.7 and 1.3. For the 90s, there are only 1.3 expected cases, violating the assumption of more than 5 expected cases per cell. This category is merged with the 80s so that the assumption of 5 or more cases per category is met. For the less than 50 category, the same problem emerges. However, there are 3.6 expected cases in this category, and this is only a little less than 5, so this is kept as a separate category, even though this violates the assumptions a little. The calculation of the χ^2 *statistic is given in Table 10.11.*

For the χ^2 *test, the number of degrees of freedom is the number of categories minus one. Here this is* $5 - 1 = 4$ *degrees of freedom. For* $\alpha = 0.05$

and 4 degrees of freedom, the χ^2 value from the table is 9.4877. The region of rejection is all χ^2 values of 9.4877 or more. The value from this data is $\chi^2 = 1.014$, and this is not in the region of rejection of the null hypothesis. As a result, the null hypothesis that the distribution of the grades in Social Studies 201 does not differ from the distribution of all grades in the Faculty of Arts cannot be rejected. This conclusion can be made at the 0.20 level of significance.

Additional Comments. *1. Note that* **two different null hypotheses** *have been tested here, and neither can be rejected. One hypothesis claimed that the grades are normally distributed, and the other null hypothesis was that the grade distribution is the same as that of grades in the Faculty of Arts as a whole. Since neither of these could be rejected, there is some question concerning which of these is the correct null hypothesis.*

Within the method of hypothesis testing itself, there is no clear way of deciding which of these two null hypotheses is correct. Since the χ^2 value is smaller in the the case of the second null hypothesis, that the grades are the same as the Faculty of Arts as a whole, it could be claimed that this is the proper hypothesis, and is the best explanation of the Social Studies 201 grade distribution. However, the normal distribution also provides a reasonable explanation.

The problem here is that of Type II error. In each case, the null hypothesis could not be rejected, but the null hypothesis may not be exactly correct either. Where there are two null hypotheses, both of which seem quite reasonable and neither of which can be rejected, then the uncertainty is increased.

The only real solution to this problem would be to obtain more data. If the grades for Social Studies 201 over several semesters were to be combined, then a clearer picture might emerge. If a sample of grades over the course of 4 or 5 semesters were to be examined, it is likely that one or other of these two null hypotheses could be rejected. Since this data is not provided here, there is little that can be done except to regard each of these two null hypotheses as possible explanations for the distribution of Social Studies 201 grades.

2. **Accepting the Null Hypothesis.** *In the discussion of hypothesis tests so far in the textbook, the null hypothesis has never been accepted. If the data is such that the null hypothesis cannot be rejected, then the conclusion is left at this point, without accepting the null hypothesis. The reason for this is that the level of Type II error is usually fairly considerable,* *and without more evidence, most researchers feel that more proof would be required before the null hypothesis could be accepted.*

For the second of the two tests, $\chi^2 = 1.014$ *is so low that the null hypothesis might actually be accepted. On the second page of the chi square table of Appendix* **??**, *there are entries for the chi square values such that most of the distribution lies to the right of these values. For 4 degrees of freedom, and* $\chi^2 = 1.014$, *there is approximately 0.900 of the area under the curve to the right of this. That is, for 4 degrees of freedom, the significance* $\alpha = 0.900$ *is associated with* $\chi^2 = 1.064$. *This means that just over 90% of the area under the chi square distribution lies to the right of a chi square value of 1.014. Since this* χ^2 *value is so close to the left end of the distribution, this might be taken as proof that the distribution is really the same as that of the Faculty of Arts as a whole.*

If this is to be done, then it makes sense to reduce Type II error to a low level, and increase Type I error. Note that if a significance level of 0.900 is selected, and the null hypothesis is rejected, there is a 0.900 chance of making Type I error. Since this is usually regarded as the more serious of the two types of error, the researcher wishes to reduce this error. But if the common hypothesis testing procedure is reversed, so that the aim is to reject the alternative hypothesis, and prove that a particular null hypothesis is correct, then the usual procedures are reversed. In order to do this, the significance level should be a large value, so that the null hypothesis can be rejected quite easily. Only in those circumstances where the data conforms very closely with the claim of the null hypothesis, should the null hypothesis be accepted.

From this discussion, it would be reasonable to accept the claim made in the second null hypothesis, that the Social Studies 201 distribution is the same as the distribution of grades for all classes in the Faculty of Arts. While this conclusion may be in error, this error is very minimal, because the actual distribution conforms so closely to the hypothesized distribution.

Conclusion. The examples presented here show the wide range of possible types of goodness of fit test. Any claim that is made concerning the whole distribution of a variable can be tested using the chi square goodness of fit test. The only restriction on the test is that there should be approximately 5 or more expected cases per cell. Other than that, there are really no restrictions on the use of the test. The frequency distribution could be measured on a nominal, ordinal, interval or ratio scale, and could be either discrete or continuous. All that is required is a grouping of the values of the variable into categories, and you need to know the number of observed cases which fall into eachi category. Once this is available, any claim concerning the nature of the distribution can be tested.

At the same time, there are some weaknesses to the test. This test is often a first test to check whether the frequency distribution more or less conforms to some hypothesized distribution. If it does not conform so closely, the question which emerges is how or where it does not match the distribution claimed. The chi square test does not answer this question, so that further analysis is required. For example, in the case of the sample of Toronto women in Example 10.3.1, the chi square goodness of fit test showed that the sample was not exactly representative of Toronto women in age. By looking back at the observed and expected numbers of cases, it was seen that the middle aged group was underrepresented. But this latter conclusion really falls outside the limits of the chi square goodness of fit test itself.

The chi square test for goodness of fit is a very useful and widely used test. The following section shows another, and quite different way, of using the chi square test and distribution.

10.4 Chi Square Test for Independence

The chi square test for independence of two variables is a test which uses a cross classification table to examine the nature of the relationship between these variables. These tables are sometimes referred to as **contingency tables**, and they have been discussed in this textbook as **cross classification tables** in connection with probability in Chapter 6. These tables show the manner in which two variables are either related or are not related to each other. The test for independence examines whether the observed pattern between the variables in the table is strong enough to show that the two variables are dependent on each other or not. While the chi square statistic and distribution are used in this test, the test is quite distinct from the test of goodness of fit. The goodness of fit test examines only one variable, while the test of independence is concerned with the relationship between two variables.

Like the goodness of fit test, the chi square test of independence is very general, and can be used with variables measured on any type of scale, nominal, ordinal, interval or ratio. The only limitation on the use of this test is that the sample sizes must be sufficiently large to ensure that the expected number of cases in each category is five or more. This rule can be modified somewhat, but as with all approximations, larger sample sizes are preferable to smaller sample sizes. There are no other limitations on the use of the test, and the chi square statistic can be used to test any contingency or cross classification table for independence of the two variables.

The chi square test for independence is conducted by assuming that there is no relationship between the two variables being examined. The alternative hypothesis is that there is some relationship between the variables. The nature of statistical relationships between variables has not been systematically examined in this textbook so far. The following section contains a few introductory comments concerning the nature of statistical relationships among variables.

10.4.1 Statistical Relationships and Association

There are various types of statistical relationships which can exist among variables. Each of these types of relationship involves some form of connection or association between the variables. The connection may be a causal one, so that when one variable changes, this causes changes in another variable or variables. Other associations among variables are no less real, but the causal nature of the connection may be obscure or unknown. Other variables may be related statistically, even though there is no causal or real connection among them. A few examples of the types of statistical relationship that can exist, and how these might be interpreted follow. In this Chapter, the chi square test provides a means of testing whether or not a relationship between two variables exists. In Chapter **??**, various summary measures of the extent of the association between two variables are developed.

One way to consider a relationship between two variables is to imagine that one variable affects or influences another variable. For example, in agriculture, the influence of different amounts of rainfall, sunshine, temperature, fertilizer and cultivation can all affect the yield of the crop. These effects can be measured, and the manner in which each of these factors affect crop yields can be determined. This is a clear cut example of a relationship between variables.

In the social sciences, experimentation of the agriculture type is usually not possible, but some quite direct relationships can be observed. Education is observed to have a positive influence on incomes of those who are employed in the paid labour force. Those individuals who have more years of education, on average have higher income levels. While the higher incomes are not always caused by more education, and are certainly not assured to any inidividual who obtains more years of education, this relationship does hold true in the aggregate. In terms of conditional probabilities, the probability of being in a higher income category is greater for those with more years of education than for those with fewer years of education. Studies of education and the labour force generally give these results, so that there is a relationship between the variables income and education.

Both of the previous examples involve a causal, and a one directional statistical relationship. Another type of relationship that is common in the social sciences is the association between two variables. An example of this could be opinions and political party supported. In Example **??** it was observed that PC supporters in Edmonton were more likely to think that trade unions were at least partially responsible for unemployment than were Liberal party supporters. It is difficult to know exactly what the nature of the connection between these variables is, but it is clear that the two variables, political preference and opinion, are related. Perhaps those who tend to have a negative attitude toward trade unions are attracted to the program of the Conservative Party. Or it may be that Conservative Party supporters together tend to develop views that trade unions are partially responsible for unemployment. It is also possible that opinions and political preferences are developed simultaneously, so that those who come to support the PCs, at the same time develop negative views toward trade unions. Regardless of the manner in which opinions and political preferences are formed, there are often associations of the type outlined here. This association can be observed statistically, and the nature of the pattern or connection between the variables can be examined with many statistical procedures. The chi square test for independence provides a first approach to examining this association.

The chi square test of independence begins with the hypothesis of no association, or no relationship, between the two variables. In intuitive terms this means that the two variables do not influence each other and are not connected in any way. If one variable changes in value, and this is not associated with changes in the other variable in a predictable manner, then there is no relationship between the two variables. For example, if the cross section of opinions is the same regardless of political preference, then opinions and political preference have no relationship with each other.

When one variable does change in value, and this is associated with some systematic change in values for another variable, then there is statistical dependence of the two variables. For example, shen the distribution of opinions does differ across different political preferences, then opinion and political preference have an association or relationship with each other. This is the alternative hypothesis which will be used in the chi square test for independence.

The conditional probabilities of Chapter 6 can be extended to discuss statistical relationships. In Chapter 6, two events A and B were said to be independent of each other when the conditional probability of A, given event B, equals the probability of A. That is,

$$P(A/B) = P(A) \text{ or } P(B/A) = P(B).$$

If independence of events is extended so that all possible pairs of events across two variables are independent, then the two variables can also be considered to be independent of each other.

The two variables X and Y are defined as being **independent** variables if the probability of the occurrence of one category for variable X does not depend on which category of variable Y occurs. If the probability of occurrence of the different possible values of variable X depend on which category of variable Y occurs, then the two variables X and Y are **dependent** on each other.

Based on these notions of independence and dependence, the chi square test for independence is now discussed.

10.4.2 A Test for Independence

The test for independence of X and Y begins by assuming that there is no relationship between the two variables. The alternative hypothesis states that there is some relationship between the two variables. If the two variables in the cross classification are X and Y, the hypotheses are

$$H_0 : \text{No relationship between } X \text{ and } Y$$

$$H_1 : \text{Some relationship between } X \text{ and } Y$$

In terms of independence and dependence these hypotheses could be stated

$$H_0 : X \text{ and } Y \text{ are independent}$$

$$H_1 : X \text{ and } Y \text{ are dependent}$$

The chi square statistic used to conduct this test is the same as in the goodness of fit test:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$

The observed numbers of cases, O_i, are the numbers of cases in each cell of the cross classification table, representing the numbers of respondents that take on each of the various combinations of values for the two variables. The expected numbers of cases E_i are computed on a different basis than in the goodness of fit test. Under the null hypothesis of no relationship between X and Y, the expected cases for each of the cell can be obtained from the multiplication rule of probability for independent events. The manner in which these expected cases are computed will be shown in the following example, with a general formula given later in this section.

The level of significance and the critical region under the χ^2 curve are obtained in the same manner as in the goodness of fit test. The formula for the degrees of freedom associated with a cross classification table is also given a little later in this section. The chi square statistic computed from the observed and expected values is calculated, and if this statistic is in the region of rejection of the null hypothesis, then the assumption of no relationship between X and Y is rejected. If the chi square statistic is not in the critical region, then the null hypothesis of no relationship is not rejected.

An example of a chi square test for independence is now given. The general format and formulas for the chi square test of independence are provided following Example 10.4.1.

Example 10.4.1 Test for a Relationship between Sex and Class

In Section 6.2.9 of Chapter 6, a survey sampling example showing a cross classification of sex by class was given. The cross classification table is presented again in Table 10.12. If variable X is the sex of the respondent, and variable Y is the social class of the respondent, use the chi square test of independence to determine if variables X and Y are independent of each other. Use the 0.05 level of significance.

Solution. *The solution provided here gives quite a complete description of the procedures for conducting this test. In later tests, some of these procedures can be streamlined a little.*

Y (Social Class)	X (Sex) Male (M)	Female (F)	Total
Upper Middle (A)	33	29	62
Middle (B)	153	181	334
Working (C)	103	81	184
Lower (D)	16	14	30
Total	305	305	610

Table 10.12: Social Class Cross Classified by Sex of Respondents

The test begins, as usual, with the statement of the null and research hypotheses. The null hypothesis states that there is no relationship between the variables X and Y, so that sex and social class are independent of each other. This means that the distribution of social class for males should be the same as the distribution of social class for females. By examining the two columns of Table 10.12, it can be seen that the distributions are not identical. The question is whether the differences in the distributions of class between males and females are large enough to reject the hypothesis of independence. The alternative hypothesis is that sex and class are related, so that the variables X and Y are dependent. These hypotheses could be stated in any of the following forms.

$$H_0 : \textit{No relationship between sex and class}$$

$$H_1 : \textit{Some relationship between sex and class}$$

or

$$H_0 : \textit{Sex and Class are independent}$$

$$H_1 : \textit{Sex and Class are dependent}$$

or

$$H_0 : O_i = E_i$$

$$H_1 : O_i \neq E_i$$

The last format is the least desirable form in which to write the hypotheses, in that it does not really say exactly what is being tested. However, each of *the null hypotheses listed above are equivalent to each other, and the last form is the one in which the test is actually carried out. That is, if there is no relationship between the two variables, or if X and Y are independent of each other, then $O_i = E_i$ for each of the cells in the table. This is because the E_is are computed assuming independence of the two variables.*

The chi square statistic used to conduct this test is the same as in the goodness of fit test:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

and the main problem becomes one of computing the expected number of cases for each cell of the table. The observed number of cases are the 8 entries $33, 29, 153, \cdots, 16, 14$ *in Table 10.12. These are repeated in Table 10.13 with the expected numbers of cases also given.*

	X (Sex)		
Y (Social Class)	Male (M)	Female (F)	Total
Upper Middle (A)	33	29	62
	(31.0)	(31.0)	
Middle (B)	153	181	334
	(167.0)	(167.0)	
Working (C)	103	81	184
	(92.0)	(92.0)	
Lower (D)	16	14	30
	(15.0)	(15.0)	
Total	305	305	610

Table 10.13: Observed and Expected Number of Cases for Social Class Cross Classified by Sex of Respondents

The expected numbers of cases are obtained under the null hypothesis that the two variables are independent of each other. If the two variables are independent, then each pair of possible events is also independent. Take the top left cell, the males who are upper middle class. This is the combination of events A and M. In order to determine the expected number of cases for this cell, assuming independence of A and M, the probability of obtaining *this combination is determined first. If these two events are independent of each other, the probability of the two events occuring together is*

$$P(A \text{ and } M) = P(A)P(M).$$

As was shown in Section 6.2.11, when two events are independent of each other, their joint probability is the product of their individual probabilities. This probability is

$$P(A \text{ and } M) = \frac{62}{610} \times \frac{305}{610}.$$

Note that this probability does not take into account the observed number of cases in this cell of the table, it uses only the information about the proportion of males and the proportion of upper middle class respondents.

If being male and being upper middle class are independent of each other, the expected number of respondents in this category is the probability of finding this combination of characteristics, times the number of respondents in the whole sample. That is

$$E(A \text{ and } M) = P(A \text{ and } M) \times n.$$

where $n = 610$ *is the sample size. This means that the above probability, multiplied by* n*, gives the expected number of upper middle class males. This is*

$$\begin{aligned} E(A \text{ and } M) &= \frac{62}{610} \times \frac{305}{610} \times 610 \\ &= \frac{62 \times 305}{610} \\ &= 31.0 \end{aligned}$$

This same procedure can be followed for each of the cells.

For the number of middle class males, the probability of finding this combination, assuming no relationship between male and being middle class, is $(334/610) \times (305/610)$. *Since there are again* $n = 610$ *cases in the sample, the expected number of middle class males is*

$$\begin{aligned} E(B \text{ and } M) &= \frac{334}{610} \times \frac{305}{610} \times 610 \\ &= \frac{334 \times 305}{610} \\ &= 167.0 \end{aligned}$$

For the number of working class females, the probability of finding this combination, assuming no relationship between female and being working class, is (184/610) × (305/610). *Since there are again* $n = 610$ *cases in the sample, the expected number of working class females is*

$$\begin{aligned} E(B \text{ and } M) &= \frac{184}{610} \times \frac{305}{610} \times 610 \\ &= \frac{184 \times 305}{610} \\ &= 92.0 \end{aligned}$$

That is, 184 out of the 610 respondents are working class, and if sex and class are independent, there should be 184/610 *of each sex who are working class. Of this proportion,* 305/610 *of the sample are females, so the proportion of female, working class members in the sample should be* (184/610) × (305/610). *Since this is a proportion, this is multiplied by the number of cases,* $n = 610$ *to obtain the expected number of working class females.*

Note that each of these expected numbers of cases follows a pattern. For each row, the row total is multiplied by the column total, and this product is divided by the sample size. This produces a general format as follows. For each cell in the cross classification table, the expected number of cases is

$$E = \frac{\textit{row total} \ \times \ \textit{column total}}{\textit{grand total}}.$$

For example, the expected number of lower class females is

$$E = \frac{30 \times 305}{610} = 15.0.$$

Each of the expected numbers in the cells can be obtained in the same manner. All of the expected values are in Table 10.13. There are some short cuts which could be taken though, and these short cuts also give an idea of the degrees of freedom associated with the test. Take the first row. Once it has been determined that there are 31.0 expected upper middle class males, this means there must be $62 - 31.0 = 31.0$ *upper middle class females, in order to preserve the total of 62 upper middle class respondents overall. Similarly, when there are 92.0 working class females to be expected, there must be* $184 - 92.0 - 92.0$ *working class males. From this, it can be seen that once one column of expected values has been determined, in this table the second column can be determined by subtraction.*

Also note that the last row of the table has a similar property. If there are expected to be 31.0, 167.0 and 92.0 males in the first three rows, then the fourth column can be obtained by subtracting the sum of these three expected numbers from the total of 305 males. That is, the expected number of lower class males is

$$305 - (31.0 + 167.0 + 92.0) = 305 - 290.0 = 15.0$$

The expected number of lower class females can be determined in s similar manner.

This set of calculations also illustrates the number of degrees of freedom. In this table, there are only 3 degrees of freedom. Once the expected number of males in the first three rows is determined, each of the other entries in the table is constrained because of the given row and column totals. This means that only 3 cells of the table can be freely assigned values, and once these 3 values have been assigned, all the other expected values are determined by the row and column totals.

The general formula for the degrees of freedom is the number of rows minus one, times the number of columns minus 1. Here there are 4 rows and 2 columns, so that the degrees of freedom is

$$(4 - 1) \times (2 - 1) = 3 \times 1 = 3.$$

The chi square values must still be calculated, and these are given in Table 10.14. Each entry in this table is labelled according to the cell it is in. For example, category AM is the upper left cell, that is events A and M.

For $\alpha = 0.05$ *significance and 3 degrees of freedom, the critical chi squared value is 7.815. Since the chi square statistic from this data is* $\chi^2 = 5.369 < 7.815$*, the null hypothesis of independence of sex and class cannot be rejected. While the male and female social class distributions differ somewhat, they do not differ enough to conclude that there is a relationship between sex and class.*

10.4.3 Notation for the Test of Independence

The steps involved in carrying out the χ^2 test for independence of two variables are as follows.

1. State the null and research hypotheses as

$$H_0 : \text{No relation between the two variables}$$

$$H_1 : \text{Some relation exists between the two variables}$$

Category	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
AM	33	31.0	2	4	0.129
BM	153	167.0	-14	196	1.174
CM	103	92.0	11	121	1.315
DM	16	15.0	1	1	0.067
AF	29	31.0	2	4	0.129
BF	181	167.0	14	196	1.174
CF	81	92.0	11	121	1.315
DF	14	15.0	1	1	0.067
Total	610	610.0	0.0		$\chi^2 = 5.370$

Table 10.14: Chi Square Test for Independence between Sex and Class

2. The test statistic is
$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$
3. The degrees of freedom for the test is the number of row minus one times the number of columns minus 1.
4. The level of significance for the test is selected.
5. From the χ^2 table, the critical value is obtained. This is based on the level of significance and the degrees of freedom. The critical region is all values of the χ^2 statistic that exceed the critical value.
6. The expected number of cases for each cell of the table is obtained for each cell in the table. For each cell of the table, this is accomplished by multiplying the row total times the column total, and divided by the grand total.
7. The chi square statistic is computed by using the observed and the expected numbers of cases for each cell. For each category, subtract the observed from the expected number of cases, square this difference, and divide this by the expected number of cases. Then sum these values for each of the cells of the table. The resulting sum is the chi square statistic.

8. If the chi square statistic exceeds the critical chi square value, reject the null hypothesis and accept the alternative hypothesis that there is a relationship between the two variables. If the chi square statistic does not exceed the critical value, then do not reject the null hypothesis that there is no relationship between the two variables.

The only restriction on the use of the test is that the expected number of cases should exceed 5 in most cells of the cross classification table. One widely used rule is as follows.

Assumptions about E_i

No expected case E_i should be less than 1. No more than 20% of the cells should have less than 5 expected cases.

This is the rule which SPSS prints out as part of its output when the chi square test for independence is requested.

The following paragraphs discuss this test in a little more detail, and provide a general notation for the test. If you have difficulty following this example, you can skip the next part and proceed to the examples. If you follow the above steps, and can understand the examples, then you should have little difficulty in carrying out chi square tests of independence.

In order to develop a general notation for the chi square test of independence, some new notation is necessary. Let the two variables being examined be X and Y. Suppose that the distribution for variable X has been grouped into c categories, and variable Y into r categories. Let the variable X represent the columns of the table, so that the columns are X_j, where $j = 1, 2, 3, \cdots, c$. The variable Y represents the rows of the table so that the rows are $Y_1, Y_2, Y_3, \cdots, Y_r$. Next let the oberserved values be O with two subscripts, so that O_{ij} represents the observed number of cases in row i and column j. Let the row totals be $R_1, R_2, R_3, \cdots, R_r$, the column totals $C_1, C_2, C_3, \cdots, C_c$, and the grand total is n, the sample size. All of this can be summarized in Table 10.15.

The null hypothesis is that variables X and Y are independent of each other. The alternative hypothesis is that X and Y are dependent variables.

	Variable X					
Variable Y	X_1	X_2	X_3	$\cdots$	X_c	Total
Y_1	O_{11}	O_{12}	O_{13}	$\cdots$	O_{1c}	R_1
Y_2	O_{21}	O_{22}	O_{23}	$\cdots$	O_{2c}	R_2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Y_r	O_{r1}	O_{r2}	O_{r3}	$\cdots$	O_{rc}	R_r
Total	C_1	C_2	C_3	$\cdots$	C_c	n

Table 10.15: General Notation for the χ^2 Test

In order to carry out the χ^2 test, it is necessary to compute the expected values. The expected values are computed on the basis of the null hypothesis. If the two variables X and Y are independent of each other, let the probability that $X = X_j$ and $Y = Y_i$ be P_{ij}.

$$P_{ij} = P(Y_i \text{ and } X_j) = P(Y_i)P(X_j) = \frac{R_i}{n} \times \frac{C_j}{n}.$$

This result holds based on the principles of probability and of independence in Chapter 6. The expected number of cases in each category under the assumption of independence of X and Y is the above probability multiplied by the number of cases. That is, P_{ij} is the probability of selecting a case which is in row i and column j, under the independence assumption. If n cases are selected, then the number of cases which would be expected to be in row i and column j would be the probability P_{ij} multiplied by the number of cases n. Thus the expected number of cases in row i and column j is

$$E_{ij} = P_{ij} \times n = \frac{R_i}{n} \times \frac{C_j}{n} \times n = \frac{R_i \times C_i}{n}.$$

These E_{ij} values are then used along with the observed numbers of cases O_{ij} to compute

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$

The degrees of freedom are the number of rows minus one times the number of columns minus 1. Since there are r rows and c columns, the degrees of freedom are

$$d = (r - 1) \times (c - 1).$$

This can be seen as follows. Begin with the first row and move across. Each of the first $c-1$ columns can be assigned any values the researcher likes, but once these are assigned, the last entry is determined by the requirement that the sum of the first row of the table be R_1. For the second row, the procedure is the same, with there being $c - 1$ entries which could be hypothesized to equal any value, but the last entry is constrained by the requirement that the sum of this row be R_2. The same argument can be made for each of the first $r - 1$ rows. This means that there are $(r - 1) \times (c - 1)$ degrees of freedom to this point. The last row adds no extra degrees of freedom, because each of the entries in the last row is constrained by the requirement that the column totals be $C_1, C_2, C_3, \cdots, C_c$.

What the above discussion means is that there are $(r-1)\times(c-1)$ cells of the table which could have values assigned to them in practically any manner the researcher wishes. But the remainder of the cells are constrained by the requirement that the proper row and column totals be produced. In terms of the construction of the χ^2 statistic, this means that the multinomial distribution determines the probabilities of the various possible combinations of observed and expected numbers of cases for these $(r - 1) \times (c - 1)$ cells. This multinomial distribution is approximated by the χ^2 distribution with $(r - 1) \times (c - 1)$ degrees of freedom. This distribution is constructed by summing the squares of $(r - 1) \times (c - 1)$ normally distributed values. While all of this would have to be proved mathematically, hopefully this discussion provides an intuitive idea of the manner in which this test and distribution works.

Several examples of chi square tests for independence follow. Some of the difficulties of interpreting the results will be discussed in these examples.

Example 10.4.2 Political Preferences and Opinions in Edmonton Study

The relationship between PC and Liberal party supporters in Edmonton was examined in several examples in Chapter 9. The opinion question examines there concerned views on whether or not trade unions were responsible for unemployment. The attitude scale which was used was a 7 point ordinal scale, and this did not satisfy the assumptions that the scale *be interval or ratio level in order that the means be interpreted meaningfully. In spite of this, tests for the difference of two means, and of two variances were conducted. However, the chi square test for independence might be a more appropriate first test to conduct when examining this data. Since the sample sizes of the earlier tests were small, all the PCs and Liberals in the Alberta study are used for this example.*

Table 10.16 gives the distribution of the opinions of 214 PC supporters and 53 Liberal supporters in the Edmonton study. The opinion question asked is the same as that used earlier. That is, repondents were asked their view concerning the opinion "Unemployment is high because trade unions have priced their members out of a job." Respondents gave their answers on a 7 point scale, with 1 being strongly disagree and 7 being strongly agree. Use the data in this table to test whether political preference and opinion are independent of each other or not. Use the 0.01 level of significance.

	Political Preference	
Opinion	PC	Liberal
1	9	3
2	7	5
3	7	11
4	28	3
5	51	12
6	54	7
7	58	12

Table 10.16: Distribution of Opinions of Edmonton PCs and Liberals

Solution. *The null hypothesis for the test is that there is no relationship between political preference and opinion concerning whether trades unions are partly responsible for unemployment. The alternative hypothesis is that there is a relationship between these two variables. While it is a little difficult to determine whether supporters of the two different parties have different views by examing Table 10.16, it appears that the PCs are more concentrated in the 5 through 7 categories of opinion than are the Liberals. Based on this, there does appear to be a relationship between opinion and political* *preference, but this must be tested.*

The null and alternative hypotheses are

$$H_0 : \textit{No relation between opinion and political preference.}$$

$$H_1 : \textit{Some relation between opinion and political preference.}$$

There are The statistic used to test these hypotheses are

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$

The observed values are in Table 10.16 and both the observed and the expected values are given in Table 10.17. The first step in computing these is to obtain each of the row and column totals. For the first row, there are 9 PCs and 3 Liberals for a total of 12. The first column has 214 PCs, and the second column has 53 Liberals.

Using these row and column totals, the expected numbers of cases for each category are computed. For the first row, the expected numbers are obtained by multiplying the first row total by the column total for the first column and dividing by the grand total. For the first row and first column let the expected number of cases be E_{11}.

$$E_{11} = \frac{12 \times 214}{267} = \frac{2568}{267} = 9.62$$

For the first row and the second column,

$$E_{12} = \frac{12 \times 53}{267} = \frac{636}{267} = 2.38$$

This could be obtained by subtracting E_{11} *from the row total of 12, so that* $E_{12} = 12 - 9.62 = 2.38$.

For each of the next rows, the procedure is the same. For example, for the fourth row and the second column,

$$E_{42} = \frac{31 \times 53}{267} = \frac{1643}{267} = 6.15$$

Once all the expected numbers of cases have been obtained, the next step is to compute the chi square value from all the observed and expected numbers of cases. These values are given in Table 10.18. Each of these is obtained by subtracting the observed from the expected number of cases in

Opinion	Political Preference PC	Liberal	
1	9	3	12
	9.62	2.38	
2	7	5	12
	9.62	2.38	
3	7	11	18
	14.43	3.57	
4	28	3	31
	24.85	6.15	
5	51	12	63
	50.49	12.51	
6	54	7	61
	48.89	12.11	
7	58	12	70
	56.10	13.90	
Total	214	53	267

Table 10.17: Observed and Expected Values of Opinions of Edmonton PCs and Liberals

Table 10.17, squaring this difference and dividing by the expected number of cases. The first entry, for example, is

$$\frac{(9-9.62)^2}{9.62} = \frac{-0.62^2}{9.62} = 0.040.$$

Each of the entries in Table 10.18 is computed in the same manner, with the entries in this table set up to correspond to the entries in the earlier two tables.

The sum of the chi square values in Table 10.18 is $\chi^2 = 28.105$*. There are 7 rows and 2 columns, so that there are* $(7-1) \times (2-1) = 6 \times 1 = 6$ *degrees of freedom. At the* $\alpha = 0.01$ *level of significance, the critical chi square value is* $\chi^2 = 16.812$*, and the region of rejection of the null hypothesis is all chi square values of 16.812 or more. Since the chi square value from the data is*

0.040	0.160
0.713	2.877
3.823	15.438
0.400	1.616
0.005	0.020
0.534	2.155
0.064	0.258

Table 10.18: Chi square Calculations

in the critical region, the hypothesis of no relationship between attitude and political preference can be rejected. At the 0.01 level of significance, there is evidence of a relationship between political preference and opinion. That is, assuming that this is a random sample of Edmonton adults, and given the data here, the probability that there is no relationship between political preference and opinion is less than 0.01.

Additional Comments. *1. One problem with the chi square test for this sample is that some of the cells have too few cases. The top three categories for the Liberals all have less than 5 expected cases. None of these is less than one, but 3 of the 34 cells, just over 20% of the cells, have less than 5 expected cases. Strictly speaking, this violates the rules for use of the chi square. However, the result is very significant statistically, with the null hypothesis being rejected at the 0.01 level of significance. The small violation of the rules in this circumstance would not affect this result much.*

If the assumptions are not to be violated, then the first two groups could be grouped together. That is, if all those of opinions 1 and 2 are grouped together, there will be only 6 rows for the table, with 16 PCs observed having opinions 1 or 2, and 8 Liberals with opinions 1 or 2. If the expected cases are then recalculated, the result would be to have 12 cells in the table, and only 2 of these would have less than 5 expected cases. This would then satisfy the assumptions, with less than 20% of the cells having fewer than 5 expected cases. There would be only 5 degrees of freedom in this case, so that with $\alpha = 0.01$*, the critical value is* $\chi^2 = 15.086$ *from the chi square table in Appendix* **??**. *If the* χ^2 *statistic is recalculated, it is 27.057, greater than the critical value. Thus the conclusion does not change when the first* *two categories are grouped together. At the 0.01 level of signficance, there is still strong evidence that there is a relationship between political preference and opinion for all Edmonton adults.*

2. The chi square test shows that there is a relationship between the two variables. One weakness of the test though, is that the test itself says nothing concerning the nature of the relationship, other than that the relationship exists. With the chi square test for independence, there is no one directional test. However, by returning to the table, you are often able to determine the nature of the relationship. One way to do this is to compare the observed and expected cases, and examine the manner in which they differ.

In Table 10.17, first look at the differences between observed and expected for the PCs in the first data column. For each of the first 3 rows of the table, there are fewer observed than expected PCs. In contrast, for the last 4 rows, and especially for rows 5-7, there are considerably more PCs observed than expected. This means that at the low values of opinion (the disagree) end of the scale, there are fewer PCs than what would be expected if there is no relationship between the two variables. In contrast, at the agree to strongly agree end of the scale (5-7), there are many more PCs than expected. This means that PCs are in agreement more than what would be expected if there was no relationship between opinion and political preference.

Now examine the Liberal column in Table 10.17. There the opposite pattern prevails, with there being considerably more Liberals than expected at the disagree end of the scale. In the first three rows, there are 2.38, 2.38 and 3.57 expected cases, but there are 3, 5 and 11 observed cases, respectively. At the disagree end of the scale, there are less observed than expected cases among the Liberals.

Together these two patterns show the nature of the relationship. Liberals are more likely to disagree than expected, and Conservatives more likely to agree than expected when there is no relationship between the two variables. This means that PCs are inclined to agree or even strongly agree than trade unions have sometimes priced their members out of a job. Among Liberal party supporters, this view is much less commonly held.

Some of the other measures of association examined later in this textbook are able to test for the direction of a relationship. But the chi square test by itself tests only for the existence or nonexistence of a relationship.

3. Comparing the results of the earlier tests with this test shows that the conclusions are much the same. The test for a difference of two means, for the smaller sample in Chapter 9, was technically illegitimate because the *opinion scale was only an ordinal level scale. But the result from that test was very similar to this, showing that the PCs were considerably more in agreement on this issue, than were the Liberals. The results was statistically significant there, as it is here.*

While tests for the difference between two means in the case of an ordinal scale do not satisfy all the assumptions, they are commonly used. One rationale for using them is that they often give much the same results as does the more legitimate chi square test. The test for the difference of two means also can be used as a one directional test, and it even gives some idea of the size of the difference between the two means. In that sense, it may be a preferable test to conduct, even though the assumptions for conducting it are not really satisfied.

One way to look on the chi square test for independence is to regard it as a first test to be conducted. If you know little about the relationship between two variables, then conduct a chi square test for independence to see whether there is any relationship or not. No assumptions are require for this test, other than that the expected number of cases i each cell of the table be 5 or more, or close to this. If there is no relationship between the two variables, based on the chi square test, then other tests are unlikely to demonstrate any relationships either. If there is a relationship between the two variables, using the chi square test, then further investigation of the nature of the relationship is called for. The researcher may then spend some time closely examining the pattern of differences between observed and expected cases in the table. These may allow the researcher to determine the nature the relationships that exist between the variables. On this basis, other statistical test, such as the test for difference of means or proportions may be conducted.

10.4.4 Reporting Chi Square Tests for Independence

When reporting the results of chi square tests of independence, the table of expected values and the value of the chi square statistic are usually given. The expected values are not ordinarily reported, and these can easily be computed by the reader. In addition to the chi square statistic obtained from the table, it is common to report the exact significance level, or the probability, of for the statistic. The degrees of freedom may also be reported, although this is not really necessary, since the reader can easily compute the degrees of freedom. Using these values, the nature of the relationship can be determined. The following example shows how results might be reported and interpreted.

Example 10.4.3 Attitudes to Social Spending in Newfoundland

A sample of adults in Eastern and Central Newfoundland was conducted early in 1988 to examine public attitudes toward government cuts in social spending. Some of the results from this study are described in Morris Saldov, "Public Attitudes to Social Spending in Newfoundland," **Canadian Review of Social Policy**, *26, November 1990, pages 10-14. The data in Table 10.19 comes from Table 2 of this article. Concerning this data, the author comments,*

> Respondents who knew someone on social assistance, were more likely to feel that welfare rates were too low, ...

	Knows Someone on Social Assistance		Row
Welfare Spending	Yes	No	Totals
Too Little	40	6	46
About Right	16	13	29
Too Much	9	7	16
Column Totals	65	26	91

$$\chi^2 = 11 \quad \text{df} = 2 \quad p < .005$$

Table 10.19: Knowing Someone on Social Assistance by Adequacy of Welfare Rates

Implied in the test and the comments are the null and research hypothesis

H_0 : *No relation between attitude and knowing someone on assistance*

H_1 : *Some relation between attitude and knowing someone on assistance*

If the null hypothesis of no relationship between attitudes and whether the respondent knows someone who receives social assistance is assumed, then the expected values and the chi square statistic can be determined in the *same manner as the previous examples. As an exercise, you should be able to confirm that the* χ^2 *statistic for this table is 11.00, to two decimal places. There are 3 rows and 2 columns so there are* $(3-1) \times (2-1) = 2 \times 1 = 2$ *degrees of freedom for this table. At* $\alpha = 0.005$ *and 2 degrees of freedom, the critical value from the chi square table in Appendix* **??** *is 10.597. Since* $\chi^2 = 11.00 > 10.597$*, the null hypothesis of no relationship between the two variables can be rejected. For the 0.001 level of significance,* $\chi^2_{0.001,2} = 13.816$*. The chi square statistic does not exceed this value, so the null hypothesis cannot be rejected at the 0.001 level of significance. Thus the significance level is reported as being less than 0.005.*

From the chi square test, it is known that there is a relationship between the variables. By examining the table, the nature of the relationship can be determined. Of those respondents who know someone receiving social assistance, 40 out of 65 or $(40/65) \times 100\% = 61.5\%$ *said that there is too little welfare spending. In contrast, of those who did not know someone who received social assistance, only* $(6/26)\% = 23.1\%$ *considered welfare spending inadequate. Some of the other cells can be examined in a similar manner. The result is that larger percentages of those who know someone on social assistance support more welfare spending than for those who do not know anyone receiving such assistance. Saldov's result shows that whether or not an individual knows someone on social assistance appears to be associated with that individual's views concerning welfare spending.*

10.4.5 Test of Independence from an SPSS Program

When survey data has been entered into a computer program, it is relatively easy to obtain cross classification tables and chi square tests of independence from these tables. This section gives several examples of such tables, along with some guidelines concerning how such tables can be organized and interpreted. Since the SPSS program calculates the expected values, the χ^2 statistic, and the exact significance level of the statistic, you need not carry out any calculations by hand. Instead, with these tests you can spend more time organizing and interpreting the results. Rather than providing a general description of methods to be used, some of the guidelines concerning analysis of cross classifications and the chi square test of independence are introduced in the examples which follow. The SPSS program commands for obtaining the tables in this section are not given here. For those, consult an SPSS manual.

Example 10.4.4 Relationship Between Age and Attitudes toward Welfare Assistance

In the Regina Labour Force Survey, respondents were asked their views concerning whether it is too easy to get welfare assistance. Table 10.20 gives a cross classification of respondent's age by respondent's opinion concerning welfare assistance. Notice that a four point attitude scale has been used with the categories being strongly agree to strongly disagree, but with no neutral category. The question asked was

> *Do you strongly agree, somewhat agree, somewhat disagree, or stronlgy disagree that it's too easy to get welfare assistance?*

In this table, age has been grouped into the categories 'less than 30,' '30-44,' '45-59,' and '60 and over.' Other age groupings could have been used, and this just happens to be the particular age grouping chosen for this table.

Each cell in Table 10.20 has two entries. The first entry in each cell is the observed number of cases for each combination of age and attitude. The second entry in each cell is the expected number of cases under the null hypothesis that there is no relationship between age and attitude. For example, the top left cell shows that there were 84 respondents less than age 30 who strongly agreed that it is too easy to get welfare assistance. If there had been no relationship between age and attitude, 78.9 respondents would have been expected to have this combination of characteristics.

The row totals are given on the right, the column totals at the bottom, and the grand total of 753 respondents on the lower right. There were 232 respondents in the Survey under age 30, and this constitutes $(232/753) \times 100\% = 30.8\%$ *of all respondents. For the 'Strongly Disagree' category, there were 85 respondents in total, or* $(85/753) \times 100\% = 11.3\%$ *of all respondents.*

For the test of independence, the assumptions are the usual hypotheses, H_0 *that there is no relationship, and* H_1 *that there is some relationship between age and attitude. The chi square statistic is given at the bottom of the table, to the right of 'Pearson.' The chi square test of this section was originally developed early in the 1900s by the statistician Karl Pearson, and this statistic is sometimes named after him. For the cross classification of Table 10.20,* $\chi^2 = 17.463$ *has been computed by the SPSS program, using the number of observed and expected cases, and the chi square formula. As an exercise, you could use these values to check that this is the proper value of the chi square statistic for this table. For this table, there are*

```
AGE   RESPONDENT'S AGE  by
WELF  TOO EASY TO GET WELFARE ASSISTANCE?

             WELF
     Count  |
    Exp Val |STRONGLY SOMEWHAT SOMEWHAT STRONGLY
            | AGREE    AGREE    DISAGRE  DISAGRE   Row
            |     1  |     2  |     3  |     4  | Total
AGE         +--------+--------+--------+--------+
         1  |    84  |    76  |    52  |    20  |   232
< 30        |  78.9  |  72.4  |  54.5  |  26.2  | 30.8%
            +--------+--------+--------+--------+
         2  |    88  |    82  |    80  |    34  |   284
30-44       |  96.6  |  88.6  |  66.8  |  32.1  | 37.7%
            +--------+--------+--------+--------+
         3  |    39  |    44  |    33  |    15  |   131
45-59       |  44.5  |  40.9  |  30.8  |  14.8  | 17.4%
            +--------+--------+--------+--------+
         4  |    45  |    33  |    12  |    16  |   106
60+         |  36.0  |  33.1  |  24.9  |  12.0  | 14.1%
            +--------+--------+--------+--------+
     Column     256      235      177       85      753
      Total   34.0%    31.2%    23.5%    11.3%   100.0%

 Chi-Square      Value        DF     Significance
 ----------    -----------   ----    ------------

 Pearson        17.46299       9        .04194
```

Table 10.20: Relation between Age and Attitude toward Welfare Assistance from SPSS Program

4 rows and 4 columns so that there are $(4-1) \times (4-1) = 9$ *degrees of freedom. The chi square table in Appendix* **??** *for 0.05 significance and 9 degrees of freedom gives a critical* χ^2 *value of 16.919. At 0.05 significance, the region of rejection of* H_0 *is all chi square values of 16.919 or more. Since* $17.463 > 16.919$, *the chi square value from the table is in the region of rejection of the null hypothesis, and the null hypothesis of no relationship between age and attitude can be rejected. At the 0.05 level of significance, there is evidence of a relationship between age and attitude.*

With the SPSS program, the exact level of significance for the value of the chi square statistic is also given. This provides an alternative way of conducting the test, without having to use the chi sqaure table in Appendix **??**. *At the bottom right of the table, the exact significance is reported to be 0.042. This means that there is exactly 0.042 of the area under the chi square curve that lies to the right of* $\chi^2 = 17.463$. *This means that if the null hypothesis is true, the chance that the chi square value is as large as 17.463, or larger, is 0.042. Since this is a fairly small probability, the null hypothesis can be rejected. That is, at any level of significance as low as 0.042, the null hypothesis can be rejected. At any lower level of signficance than 0.042, the null hypothesis cannot be rejected. The reader can decide whether 0.042 is low enough to allow for rejection of the hypothesis of no relationship, and acceptence of the hypothesis of some relationship.*

Given the conclusion that age and attitude are related to each other, the nature of the relationship can be examined in Table 10.20. The pattern of the relationship appears to be that the youngest and the oldest age groups both tend to agree that it is too easy to get welfare assistance. The age groups which have larger percentages of respondents who disagree are the 30-44 and 45-59 age groups. This can be seen by comparing the observed and expected numbers of cases in each of the cells as follows.

Notice first that for the less than 30 age group, there are more observed than expected cases in each of the strongly agree and somewhat agree categories. That is, there are more of the youngest age group who agree, than what would be expected if there was no relationship between age and attitude. For the middle two age groups, this pattern is reversed, with there being fewer observed than expected cases. For these two age groups, 30-44 and 45-59, there are more observed than expected cases among the two disagree categories. Finally, for the oldest age group, there are considerably more observed than expected cases in the two agree categories. This means that older respondents were more likely to agree that it is too easy to get welfare assistance, than what would have been expected under the *assumption of no relationship between age and attitude.*

In summary, the pattern that appears in this table is that the middle aged groups express more disagreement, with the youngest and oldest age groups expressing more agreement. This pattern could be a consequence of the particular sample chosen, or the particular age grouping adopted here. But given these, this does appear to the pattern that exists in this table.

Example 10.4.5 Differences in Education and Attitudes Toward Welfare Spending

This example shows the relationship between the education of the respondent and attitudes toward welfare spending. The attitude question used in this example is the same question as in Example 10.4.4, and this variable was cross classified by education of respondent. The complete table is given in Table 10.21. This table is read in the same manner as Table 10.20. The first number in each cell is the observed number of respondents with each combination of characteristics. The second number is the expected number of respondents, assuming independence between education and attitude. Row, column and grand totals are also given.

At the bottom of the table, it can be seen that $\chi^2 = 35.107$, *but with 27 degrees of freedom, this result is not very significant statistically. The exact significance of 0.13618 means that there is 0.13618 of the area under the* χ^2 *curve to the right of the chi square value of 35.107. Alternatively stated, if the null hypothesis of no relationship between education and attitude is true, the probability of obtaining a chi square value of 35.107 or more is 0.13618. This is not a low probability, so that the null hypothesis of no relationship between education and attitude may be correct. This null hypothesis of no such relationship cannot rejected at any significance level below 0.136.*

In stating this conclusion, there are two problems, both caused by the size of the table. First, there is one cell with as few as 1.7 expected cases. This is rather low, being quite close to 1 case in that cell, and well below 5 expected cases per cell. In addition, over 20% of the cells have less than 5 expected cases per cell. The assumptions concerning the number of expected cases are not met in this table.

The second problem with so many cells in the table is that it is difficult to determine any patterns of relationship which may exist in the data. Perhaps there is little relationship between the variables, but even if there is, it would be difficult to discern what the nature of this relationship would be. In order to make the table a little easier to analyze, the categories of education were

Table 10.21: Relationship between Education and Attitude

regrouped, so that there were only 4 categories. This was done by grouping adjacent categories so the new categories are (11) less than high school, (12) completed high school, (13) some postsecondary education, but no degree, and (14) having a degree. Table 10.22 contains the SPSS table for the cross classification of education by attitude, where these new education categories are used.

Again the table is read in the same manner as earlier, but now all the assumptions are met and the table is considerably easier to examine. The relationships also become much clearer. If the null hypothesis is that there is no relationship between education and attitude, and the alternative hypothesis is that education and attitude are dependent variables, then the null hypothesis can be rejected at the 0.02 level of significance. That is, the chi square statistic for this table is 19.791 and with 9 degrees of freedom, there is 0.01925 of the area under the chi square curve to the right of $\chi^2 = 19.791$*. This means that the null hypothesis of no relationship can be rejected at all levels of significance down to 0.02. Since this is quite a low probability of Type I error, this is usually regarded as providing sufficient evidence for a statistical relationship between two variables.*

The pattern of the relationship between education and attitude can also be determined from the table. For education level 13, some postsecondary education, the observed and expected numbers of cases are very close to each other. But for those who have a degree, there are considerably more observed than expected cases in each of the two disagree categories. The opposite pattern prevails for those with high school education or less. There, the observed numbers of cases generally exceeds the expected numbers of cases on the agree side. The pattern which emerges then is that respondents with degrees tend toward disagreement, and those with high school or less education tend toward agreement. Those with some postsecondary education, but not a degree fall in between these two groups on both variables. The pattern is a fairly clear one that as education level increases, there is a greater tendency for respondents to agree, and less of a tendency to disagree.

By taking a little effort to organize the table in a manner which groups together respondents of similar levels of education, the relationship between education and attitude can be more clearly seen. In addition, this grouping solved the problem of the low numbers of expected cases in some of the cells of the cross classification table.

Example 10.4.6 Political Preference and Attitude Concerning Indians and Metis

EDUC RESPONDENT'S CURRENT EDUCATION
by WELF TOO EASY TO GET WELFARE ASSISTANCE?

EDUC	Count / Exp Val	WELF: STRONGLY AGREE 1	SOMEWHAT AGREE 2	SOMEWHAT DISAGRE 3	STRONGLY DISAGRE 4	Row Total
	11.00	64	51	38	20	173
LESS THAN HIGH		58.4	54.2	40.3	20.1	22.8%
	12.00	69	49	30	15	163
HIGH		55.0	51.1	38.0	18.9	21.5%
	13.00	77	80	50	25	232
HIGH PLUS		78.3	72.7	54.1	26.9	30.6%
	14.00	46	58	59	28	191
DEGREE		64.4	59.9	44.5	22.1	25.2%
	Column	256	238	177	88	759
	Total	33.7%	31.4%	23.3%	11.6%	100.0%

Chi-Square	Value	DF	Significance
Pearson	19.79131	9	.01925

Minimum Expected Frequency - 18.899

Table 10.22: Relation between Education (4 levels) and Attitude

Another question asked of respondents in the Regina Labour Force Survey was

> Now, I'd like to ask you some questions about Native Canadian Indians and Metis. Again, can you tell me if you strongly agree, agree somewhat, disagree somewhat or strongly disagree with these statements. First of all, Indians and Metis deserve to be better off than they are now ...

In order to determine whether there was any relationship between political preference of respondents and this attitude question, the cross classification of Table 10.23 was obtained. Political preference in that table is restricted to those who expressed a preference for either the NDP or PCs at the provincial level in Saskatchewan. The table gives the observed, expected, row, column and total numbers of cases, along with the chi square statistic for the test of independence, the degrees of freedom and the exact significance level.

If a 0.05 level of significance is adopted, the null hypothesis of no relationship between attitude and political preference can be rejected. This is because the exact significance printed on the table is $0.04150 < 0.05$ *so that* $\chi^2 = 8.229$ *is in the right 0.05 of the area under the chi square curve for 3 degrees of freedom.*

By examining the cells of the table, some of the patterns of the relationship can be seen, although the relationship is not as simple as might be imagined. There are considerably more NDP supporters (125) in the strongly agree category than expected (111.4) under the assumption of no relationship between attitude and political preference. For the PCs the pattern is reversed, with fewer observed than expected PCs in the strongly agree category. But in the agree somewhat category, the situation is again reversed, with more PCs than expected, but fewer NDPers. For the two disagree categories, the observed and expected numbers of cases are quite close for supporters of both parties. In summary, the difference between PC and NDP supporters is that NDPers tend to more strongly agree than do PCers that Indians and Metis deserve to be better off.

In order to test whether there is any real difference in the propensity of PCs and NDPers to agree or disagree, the two agree categories are grouped together, and the two disagree categories are grouped together. The fairly strong relationship of Table 10.23 disappears when the attitude categories are only agree and disagree. That is, by grouping all the agrees together, and all the disagrees together, the strength of agreement and disagreement is ignored. The new table is Table 10.24 and it can be seen there that the

AIM:INDIANS/METIS DESERVE TO BE BETTER OFF by
PROVINCIAL POLITICAL PREFERENCE

AIM	Count / Exp Val	PROVINCIAL POLITICAL PREFERENCE: NDP 2	P.C. 3	Row Total
1 STRONGLY AGREE	Count	125	36	161
	Exp Val	111.4	49.6	33.1%
2 AGREE SOMEWHAT	Count	114	64	178
	Exp Val	123.2	54.8	36.6%
3 DISAGREE SOMEWHA	Count	54	28	82
	Exp Val	56.7	25.3	16.8%
4 STRONGLY DISAGRE	Count	44	22	66
	Exp Val	45.7	20.3	13.6%
Column Total		337	150	487
		69.2%	30.8%	100.0%

Chi-Square	Value	DF	Significance
Pearson	8.22937	3	.04150

Minimum Expected Frequency - 20.329

Table 10.23: Relation between Political Preference and Attitude

```
AIM:INDIANS/METIS DESERVE TO BE BETTER OFF  by
PROVINCIAL POLITICAL PREFERENCE

                    PROVINCIAL POLITICAL
                    PREFERENCE
            Count  |
            Exp Val |NDP      P.C.
                    |                  Row
                    |     2  |     3  | Total
 AIM        --------+--------+--------+
                 1  |   239  |   100  |   339
   agree            | 234.6  | 104.4  | 69.6%
                    +--------+--------+
                 2  |    98  |    50  |   148
   disagree         | 102.4  |  45.6  | 30.4%
                    +--------+--------+
            Column     337      150      487
             Total   69.2%    30.8%   100.0%


Chi-Square          Value       DF     Significance
-----------      ----------    ----    ------------

 Pearson            .88761       1        .34613

 Minimum Expected Frequency -   45.585
```

Table 10.24: Relation between Political Preference and Attitude

chi square statistic of 0.888 is associated with an area of 0.34613 to the right of it. The probability of obtaining $\chi^2 = 0.888$ *under the assumption of no relationship between the two variables is 0.346. Since this is a large probability, the hypothesis of no relation between political preference and attitude is not rejected.*

Note that in the previous table, the NDP tended to strongly agree, and the PCs tended to agree somewhat. Once all the strongly agree and somewhat agree are grouped together as agree, and the same for the disagree side, the relationship between political preference and attitude disappears.

10.4.6 Summary

The chi square test for independence is an extremely flexible and useful test. The test can be used to examine the relationship between any two variables, with any types of measurement - nominal, ordinal, interval or ratio, and discrete or continuous. The only constraint on the use of the test is that there be sufficient numbers of cases in each cell of the table. The rule given on page 742 that there be no cell with less than one expected case, and no more than one fifth of the cells with less than 5 expected cases, is generally an adequate rule. If there are too few cases in some of the cells, then adjacent categories can usually be grouped together so that this rule can be satisfied. In doing this, the researcher must be careful concerning how the grouping is carried out.

Examples 10.4.5 and 10.4.6 demonstrate how different conclusions can be obtained from the same data. In each case, difference groupins of the data led to different conclusions. In the education example, a seemingly nonexistent relationship is made to appear as a relationship. In the case of the last example, there is some relationship between strength of attitude and political preference, but there is no relationship between overall agreement and disagreement and political preference.

While chi square tests of independence are very flexible, and can be used with any cross classification, a researcher must be careful not to either overemphasize or hide a relationship between two variables. In doing this, the chi square test itself is not the problem. The methodological problem is the difficulty of deciding the proper approach to the grouping of the data. There are no strict guidelines concerning how data is to be grouped properly. In many cases, the researcher may try several groupings, and observe what happens as these groupings change. If the results change little as the groupings change, then the relationship is likely to be quite apparent. Where the relationship seems to change as the grouping of the data changes, considerably more effort may have to be made to discern the exact nature of the relationship between the variables.

If a different grouping of the data provides too few expected cases, there may be little that the researcher can do without obtaining more data. One test that can be used for very small sample size is Fisher's Exact Test. This is not discussed in this textbook, but can be found in many statistics textbooks. This test is also provided in the SPSS program when the sample size in the cross classification table are small. In addition, some statistical programs provide corrections for continuity in the case of small sample sizes.

Chi square tests for independence are most commonly used in the course of conducting research. If a researcher has a large number of variables, some of which are measured at less than the interval level, then cross classification tables and chi square tests may be conducted for many pairs of variables. Once a chi square test for independence of two variables has been conducted, this may only be the beginning of the analysis of the relationship between these variables. If there is no relationship betweeen the two variables shown by the test, the researcher must make sure that this lack of relationship really does mean no statistical relation.

If the chi square test for independence provides some evidence that there is a relationship between the variables, then the researcher may want to find some way to describe the direction, pattern and strength of this relationship. There are many possible ways of doing this, and Chapter 11 contains a number of measures such as the correlation coefficient which can be used to describe a statistical relationship.

When reporting research results, cross classification tables, along with the chi square statistic, and the level of significance of the statistic may be reported. In addition, the researcher may provide some measure of association to summarize the relationship. But when examining any statistical relationship, there is no substitute for looking at the table itself, and examining how the two variables are related. Once the researcher understands the nature of the relationship between the variables, then the summary measures of the next chapter are more useful.

10.5 Conclusion

This chapter has presented two different types of chi square tests. The goodness of fit test allows the researcher to hypothesize that a distribution behaves in a particular manner. The chi square test is used to determine whether the observed distribution conforms to the model hypothesized by the researcher in the null hypothesis. In the test for independence, the chi square statistic and distribution are used to test whether there is a relationship between the two variables or not. For each of the two tests, the chi square statistic and distribution are the same. However, the two tests are quite different types of hypothesis tests, one examining only a single variable, and the other examining the relationship between two variables.

This chapter completes the introduction to hypothesis testing. The general principles involved in testing hypotheses were introduced in Chapter 9. There the tests involved tests of parameters such as the mean or proportion, or the difference in these between two populations. These are the most commonly used tests for these parameters. In this chapter, these principles of hypothesis testing were applied to whole distributions. The structure of the test is the same for each of the chi square tests, but the statistic used, and the fact that whole distributions rather than parameters are used, makes these tests appear somewhat different.

In the remainder of this book, methods of describing relationships among variables are discussed. In doing this, tests of hypothesis are an essential part but again the tests may appear quite different than the tests introduced so far. In the tests which follow in Chapter 11, the principles of hypothesis testing are the same as the principles introduced in Chapter 9. However, the emphasis in Chapter 11 is placed on determining the nature of the relationship among variables.