CHAPTER 12

Tests of Goodness of Fit and Independence

CONTENTS

STATISTICS *in* PRACTICE

UNITED WAY*
ROCHESTER, NEW YORK

United Way of Greater Rochester is a nonprofit organization dedicated to improving the quality of life for all people in the seven counties it serves by meeting the community's most important human care needs.

The annual United Way/Red Cross fund-raising campaign, conducted each spring, funds hundreds of programs offered by more than 200 service providers. These providers meet a wide variety of human needs—physical, mental, and social—and serve people of all ages, backgrounds, and economic means.

Because of enormous volunteer involvement, United Way of Greater Rochester is able to hold its operating costs at just eight cents of every dollar raised.

The United Way of Greater Rochester decided to conduct a survey to learn more about community perceptions of charities. Focus-group interviews were held with professional, service, and general worker groups to get preliminary information on perceptions. The information obtained was then used to help develop the questionnaire for the survey. The questionnaire was pretested, modified, and distributed to 440 individuals; 323 completed questionnaires were obtained.

A variety of descriptive statistics, including frequency distributions and crosstabulations, were provided from the data collected. An important part of the analysis involved the use of contingency tables and chi-square tests of independence. One use of such statistical tests was to determine whether perceptions of administrative expenses were independent of occupation.

The hypotheses for the test of independence were:

H_0: Perception of United Way administrative expenses is independent of the occupation of the respondent.

H_a: Perception of United Way administrative expenses is not independent of the occupation of the respondent.

United Way programs meet the needs of children as well as adults. © Ed Bock/CORBIS.

Two questions in the survey provided the data for the statistical test. One question obtained data on perceptions of the percentage of funds going to administrative expenses (up to 10%, 11–20%, and 21% or more). The other question asked for the occupation of the respondent.

The chi-square test at a .05 level of significance led to rejection of the null hypothesis of independence and to the conclusion that perceptions of United Way's administrative expenses did vary by occupation. Actual administrative expenses were less than 9%, but 35% of the respondents perceived that administrative expenses were 21% or more. Hence, many had inaccurate perceptions of administrative costs. In this group, production-line, clerical, sales, and professional-technical employees had more inaccurate perceptions than other groups.

The community perceptions study helped United Way of Rochester to develop adjustments to its programs and fund-raising activities. In this chapter, you will learn how a statistical test of independence, such as that described here, is conducted.

*The authors are indebted to Dr. Philip R. Tyler, marketing consultant to the United Way, for providing this Statistics in Practice.

In Chapter 11 we showed how the chi-square distribution could be used in estimation and in hypothesis tests about a population variance. In Chapter 12, we introduce two additional hypothesis testing procedures, both based on the use of the chi-square distribution. Like other hypothesis testing procedures, these tests compare sample results with those that are expected when the null hypothesis is true. The conclusion of the hypothesis test is based on how "close" the sample results are to the expected results.

In the following section we introduce a goodness of fit test for a multinomial population. Later we discuss the test for independence using contingency tables and then show goodness of fit tests for the Poisson and normal distributions.

12.1 Goodness of Fit Test: A Multinomial Population

In this section we consider the case in which each element of a population is assigned to one and only one of several classes or categories. Such a population is a **multinomial population**. The multinomial distribution can be thought of as an extension of the binomial distribution to the case of three or more categories of outcomes. On each trial of a multinomial experiment, one and only one of the outcomes occurs. Each trial of the experiment is assumed to be independent, and the probabilities of the outcomes remain the same for each trial.

The assumptions for the multinomial experiment parallel those for the binomial experiment with the exception that the multinomial has three or more outcomes per trial.

As an example, consider the market share study being conducted by Scott Marketing Research. Over the past year market shares stabilized at 30% for company A, 50% for company B, and 20% for company C. Recently company C developed a "new and improved" product to replace its current entry in the market. Company C retained Scott Marketing Research to determine whether the new product will alter market shares.

In this case, the population of interest is a multinomial population; each customer is classified as buying from company A, company B, or company C. Thus, we have a multinomial population with three outcomes. Let us use the following notation for the proportions.

$$p_{\text{A}} = \text{market share for company A}$$
$$p_{\text{B}} = \text{market share for company B}$$
$$p_{\text{C}} = \text{market share for company C}$$

Scott Marketing Research will conduct a sample survey and compute the proportion preferring each company's product. A hypothesis test will then be conducted to see whether the new product caused a change in market shares. Assuming that company C's new product will not alter the market shares, the null and alternative hypotheses are stated as follows.

$$H_0\colon p_{\text{A}} = .30,\ p_{\text{B}} = .50,\text{ and } p_{\text{C}} = .20$$
$$H_{\text{a}}\colon \text{The population proportions are not}$$
$$p_{\text{A}} = .30,\ p_{\text{B}} = .50,\text{ and } p_{\text{C}} = .20$$

If the sample results lead to the rejection of H_0, Scott Marketing Research will have evidence that the introduction of the new product affects market shares.

Let us assume that the market research firm has used a consumer panel of 200 customers for the study. Each individual was asked to specify a purchase preference among the three alternatives: company A's product, company B's product, and company C's new product. The 200 responses are summarized here.

The consumer panel of 200 customers in which each individual is asked to select one of three alternatives is equivalent to a multinomial experiment consisting of 200 trials.

Observed Frequency

Company A's Product	Company B's Product	Company C's New Product
48	98	54

We now can perform a **goodness of fit test** that will determine whether the sample of 200 customer purchase preferences is consistent with the null hypothesis. The goodness of fit test is based on a comparison of the sample of *observed* results with the *expected* results under the assumption that the null hypothesis is true. Hence, the next step is to compute expected purchase preferences for the 200 customers under the assumption that $p_A = .30$, $p_B = .50$, and $p_C = .20$. Doing so provides the expected results.

Expected Frequency

Company A's Product	Company B's Product	Company C's New Product
200(.30) = 60	200(.50) = 100	200(.20) = 40

Thus, we see that the expected frequency for each category is found by multiplying the sample size of 200 by the hypothesized proportion for the category.

The goodness of fit test now focuses on the differences between the observed frequencies and the expected frequencies. Large differences between observed and expected frequencies cast doubt on the assumption that the hypothesized proportions or market shares are correct. Whether the differences between the observed and expected frequencies are "large" or "small" is a question answered with the aid of the following test statistic.

TEST STATISTIC FOR GOODNESS OF FIT

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \tag{12.1}$$

where

f_i = observed frequency for category i
e_i = expected frequency for category i
k = the number of categories

Note: The test statistic has a chi-square distribution with $k - 1$ degrees of freedom provided that the expected frequencies are 5 *or more* for all categories.

Let us continue with the Scott Market Research example and use the sample data to test the hypothesis that the multinomial population retains the proportions $p_A = .30$, $p_B = .50$, and $p_C = .20$. We will use an $\alpha = .05$ level of significance. We proceed by using the observed and expected frequencies to compute the value of the test statistic. With the expected frequencies all 5 or more, the computation of the chi-square test statistic is shown in Table 12.1. Thus, we have $\chi^2 = 7.34$.

The test for goodness of fit is always a one-tailed test with the rejection occurring in the upper tail of the chi-square distribution.

We will reject the null hypothesis if the differences between the observed and expected frequencies are *large*. Large differences between the observed and expected frequencies will result in a large value for the test statistic. Thus the test of goodness of fit will always be an upper tail test. We can use the upper tail area for the test statistic and the *p*-value approach to determine whether the null hypothesis can be rejected. With $k - 1 = 3 - 1 = 2$ degrees of freedom, the chi-square table (Table 3 of Appendix B) provides the following:

An introduction to the chi-square distribution and the use of the chi-square table were presented in Section 11.1.

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**	**.005**
χ^2 Value (2 *df*)	4.605	5.991	7.378	9.210	10.597

$\chi^2 = 7.34$ (between 5.991 and 7.378)

TABLE 12.1 COMPUTATION OF THE CHI-SQUARE TEST STATISTIC FOR THE SCOTT MARKETING RESEARCH MARKET SHARE STUDY

Category	Hypothesized Proportion	Observed Frequency (f_i)	Expected Frequency (e_i)	Difference ($f_i - e_i$)	Squared Difference $(f_i - e_i)^2$	Squared Difference Divided by Expected Frequency $(f_i - e_i)^2/e_i$
Company A	.30	48	60	−12	144	2.40
Company B	.50	98	100	−2	4	0.04
Company C	.20	54	40	14	196	4.90
Total		200				$\chi^2 = 7.34$

The test statistic $\chi^2 = 7.34$ is between 5.991 and 7.378. Thus, the corresponding upper tail area or *p*-value must be between .05 and .025. With *p*-value $\leq \alpha = .05$, we reject H_0 and conclude that the introduction of the new product by company C will alter the current market share structure. Minitab or Excel procedures provided in Appendix F at the back of the book can be used to show $\chi^2 = 7.34$ provides a *p*-value $= .0255$.

Instead of using the *p*-value, we could use the critical value approach to draw the same conclusion. With $\alpha = .05$ and 2 degrees of freedom, the critical value for the test statistic is $\chi^2_{.05} = 5.991$. The upper tail rejection rule becomes

$$\text{Reject } H_0 \text{ if } \chi^2 \geq 5.991$$

With $7.34 > 5.991$, we reject H_0. The *p*-value approach and critical value approach provide the same hypothesis testing conclusion.

Although no further conclusions can be made as a result of the test, we can compare the observed and expected frequencies informally to obtain an idea of how the market share structure may change. Considering company C, we find that the observed frequency of 54 is larger than the expected frequency of 40. Because the expected frequency was based on current market shares, the larger observed frequency suggests that the new product will have a positive effect on company C's market share. Comparisons of the observed and expected frequencies for the other two companies indicate that company C's gain in market share will hurt company A more than company B.

Let us summarize the general steps that can be used to conduct a goodness of fit test for a hypothesized multinomial population distribution.

MULTINOMIAL DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.

   H_0: The population follows a multinomial distribution with specified probabilities for each of the *k* categories

   H_a: The population does not follow a multinomial distribution with the specified probabilities for each of the *k* categories

2. Select a random sample and record the observed frequencies f_i for each category.
3. Assume the null hypothesis is true and determine the expected frequency e_i in each category by multiplying the category probability by the sample size.

4. Compute the value of the test statistic.

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

5. Rejection rule:

p-value approach:	Reject H_0 if p-value $\leq \alpha$
Critical value approach:	Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

where α is the level of significance for the test and there are $k - 1$ degrees of freedom.

Exercises

Methods

SELF test

1. Test the following hypotheses by using the χ^2 goodness of fit test.

$$H_0\colon p_A = .40,\ p_B = .40, \text{ and } p_C = .20$$
$$H_a\colon \text{The population proportions are not } p_A = .40,\ p_B = .40, \text{ and } p_C = .20$$

A sample of size 200 yielded 60 in category A, 120 in category B, and 20 in category C. Use $\alpha = .01$ and test to see whether the proportions are as stated in H_0.

a. Use the p-value approach.

b. Repeat the test using the critical value approach.

2. Suppose we have a multinomial population with four categories: A, B, C, and D. The null hypothesis is that the proportion of items is the same in every category. The null hypothesis is

$$H_0\colon p_A = p_B = p_C = p_D = .25$$

A sample of size 300 yielded the following results.

A: 85 B: 95 C: 50 D: 70

Use $\alpha = .05$ to determine whether H_0 should be rejected. What is the p-value?

Applications

SELF test

3. During the first 13 weeks of the television season, the Saturday evening 8:00 P.M. to 9:00 P.M. audience proportions were recorded as ABC 29%, CBS 28%, NBC 25%, and independents 18%. A sample of 300 homes two weeks after a Saturday night schedule revision yielded the following viewing audience data: ABC 95 homes, CBS 70 homes, NBC 89 homes, and independents 46 homes. Test with $\alpha = .05$ to determine whether the viewing audience proportions changed.

4. M&M/MARS, makers of M&M® chocolate candies, conducted a national poll in which more than 10 million people indicated their preference for a new color. The tally of this poll resulted in the replacement of tan-colored M&Ms with a new blue color. In the

brochure "Colors," made available by M&M/MARS Consumer Affairs, the distribution of colors for the plain candies is as follows:

Brown	Yellow	Red	Orange	Green	Blue
30%	20%	20%	10%	10%	10%

In a follow-up study, samples of 1-pound bags were used to determine whether the reported percentages were indeed valid. The following results were obtained for one sample of 506 plain candies.

Brown	Yellow	Red	Orange	Green	Blue
177	135	79	41	36	38

Use $\alpha = .05$ to determine whether these data support the percentages reported by the company.

5. Where do women most often buy casual clothing? Data from the U.S. Shopper Database provided the following percentages for women shopping at each of the various outlets (*The Wall Street Journal,* January 28, 2004).

Outlet	Percentage	Outlet	Percentage
Wal-Mart	24	Kohl's	8
Traditional department stores	11	Mail order	12
JC Penney	8	Other	37

The other category included outlets such as Target, Kmart, and Sears as well as numerous smaller specialty outlets. No individual outlet in this group accounted for more than 5% of the women shoppers. A recent survey using a sample of 140 women shoppers in Atlanta, Georgia, found 42 Wal-Mart, 20 traditional department store, 8 JC Penney, 10 Kohl's, 21 mail order, and 39 other outlet shoppers. Does this sample suggest that women shoppers in Atlanta differ from the shopping preferences expressed in the U.S. Shopper Database? What is the p-value? Use $\alpha = .05$. What is your conclusion?

6. The American Bankers Association collects data on the use of credit cards, debit cards, personal checks, and cash when consumers pay for in-store purchases (*The Wall Street Journal,* December 16, 2003). In 1999, the following usages were reported.

In-Store Purchase	Percentage
Credit card	22
Debit card	21
Personal check	18
Cash	39

A sample taken in 2003 found that for 220 in-stores purchases, 46 used a credit card, 67 used a debit card, 33 used a personal check, and 74 used cash.

a. At $\alpha = .01$, can we conclude that a change occurred in how customers paid for in-store purchases over the four-year period from 1999 to 2003? What is the p-value?
b. Compute the percentage of use for each method of payment using the 2003 sample data. What appears to have been the major change or changes over the four-year period?
c. In 2003, what percentage of payments was made using plastic (credit card or debit card)?

7. *The Wall Street Journal's* Shareholder Scoreboard tracks the performance of 1000 major U.S. companies (*The Wall Street Journal,* March 10, 2003). The performance of each company is rated based on the annual total return, including stock price changes and the reinvestment of dividends. Ratings are assigned by dividing all 1000 companies into five groups from A (top 20%), B (next 20%), to E (bottom 20%). Shown here are the one-year ratings for a sample of 60 of the largest companies. Do the largest companies differ in performance from the performance of the 1000 companies in the Shareholder Scoreboard? Use $\alpha = .05$.

A	B	C	D	E
5	8	15	20	12

8. How well do airline companies serve their customers? A study showed the following customer ratings: 3% excellent, 28% good, 45% fair, and 24% poor (*BusinessWeek,* September 11, 2000). In a follow-up study of service by telephone companies, assume that a sample of 400 adults found the following customer ratings: 24 excellent, 124 good, 172 fair, and 80 poor. Is the distribution of the customer ratings for telephone companies different from the distribution of customer ratings for airline companies? Test with $\alpha = .01$. What is your conclusion?

12.2 Test of Independence

Another important application of the chi-square distribution involves using sample data to test for the independence of two variables. Let us illustrate the test of independence by considering the study conducted by the Alber's Brewery of Tucson, Arizona. Alber's manufactures and distributes three types of beer: light, regular, and dark. In an analysis of the market segments for the three beers, the firm's market research group raised the question of whether preferences for the three beers differ among male and female beer drinkers. If beer preference is independent of the gender of the beer drinker, one advertising campaign will be initiated for all of Alber's beers. However, if beer preference depends on the gender of the beer drinker, the firm will tailor its promotions to different target markets.

A test of independence addresses the question of whether the beer preference (light, regular, or dark) is independent of the gender of the beer drinker (male, female). The hypotheses for this test of independence are:

H_0: Beer preference is independent of the gender of the beer drinker

H_a: Beer preference is not independent of the gender of the beer drinker

Table 12.2 can be used to describe the situation being studied. After identification of the population as all male and female beer drinkers, a sample can be selected and each individual

TABLE 12.2 CONTINGENCY TABLE FOR BEER PREFERENCE AND GENDER OF BEER DRINKER

		Beer Preference		
		Light	**Regular**	**Dark**
Gender	**Male**	cell(1,1)	cell(1,2)	cell(1,3)
	Female	cell(2,1)	cell(2,2)	cell(2,3)

TABLE 12.3 SAMPLE RESULTS FOR BEER PREFERENCES OF MALE AND FEMALE BEER DRINKERS (OBSERVED FREQUENCIES)

		Beer Preference			
		Light	**Regular**	**Dark**	**Total**
Gender	**Male**	20	40	20	80
	Female	30	30	10	70
	Total	50	70	30	150

To test whether two variables are independent, one sample is selected and crosstabulation is used to summarize the data for the two variables simultaneously.

asked to state his or her preference for the three Alber's beers. Every individual in the sample will be classified in one of the six cells in the table. For example, an individual may be a male preferring regular beer (cell (1,2)), a female preferring light beer (cell (2,1)), a female preferring dark beer (cell (2,3)), and so on. Because we have listed all possible combinations of beer preference and gender or, in other words, listed all possible contingencies, Table 12.2 is called a **contingency table**. The test of independence uses the contingency table format and for that reason is sometimes referred to as a *contingency table test.*

Suppose a simple random sample of 150 beer drinkers is selected. After tasting each beer, the individuals in the sample are asked to state their preference or first choice. The crosstabulation in Table 12.3 summarizes the responses for the study. As we see, the data for the test of independence are collected in terms of counts or frequencies for each cell or category. Of the 150 individuals in the sample, 20 were men who favored light beer, 40 were men who favored regular beer, 20 were men who favored dark beer, and so on.

The data in Table 12.3 are the observed frequencies for the six classes or categories. If we can determine the expected frequencies under the assumption of independence between beer preference and gender of the beer drinker, we can use the chi-square distribution to determine whether there is a significant difference between observed and expected frequencies.

Expected frequencies for the cells of the contingency table are based on the following rationale. First we assume that the null hypothesis of independence between beer preference and gender of the beer drinker is true. Then we note that in the entire sample of 150 beer drinkers, a total of 50 prefer light beer, 70 prefer regular beer, and 30 prefer dark beer. In terms of fractions we conclude that $^{50}/_{150} = ^1/_3$ of the beer drinkers prefer light beer, $^{70}/_{150} = ^7/_{15}$ prefer regular beer, and $^{30}/_{150} = ^1/_5$ prefer dark beer. If the *independence* assumption is valid, we argue that these fractions must be applicable to both male and female beer drinkers. Thus, under the assumption of independence, we would expect the sample of 80 male beer drinkers to show that $(^1/_3)80 = 26.67$ prefer light beer, $(^7/_{15})80 = 37.33$ prefer regular beer, and $(^1/_5)80 = 16$ prefer dark beer. Application of the same fractions to the 70 female beer drinkers provides the expected frequencies shown in Table 12.4.

Let e_{ij} denote the expected frequency for the contingency table category in row i and column j. With this notation, let us reconsider the expected frequency calculation for males

TABLE 12.4 EXPECTED FREQUENCIES IF BEER PREFERENCE IS INDEPENDENT OF THE GENDER OF THE BEER DRINKER

		Beer Preference			
		Light	**Regular**	**Dark**	**Total**
Gender	**Male**	26.67	37.33	16.00	80
	Female	23.33	32.67	14.00	70
	Total	50.00	70.00	30.00	150

(row $i = 1$) who prefer regular beer (column $j = 2$), that is, expected frequency e_{12}. Following the preceding argument for the computation of expected frequencies, we can show that

$$e_{12} = (7/15)80 = 37.33$$

This expression can be written slightly differently as

$$e_{12} = (7/15)80 = (70/150)80 = \frac{(80)(70)}{150} = 37.33$$

Note that 80 in the expression is the total number of males (row 1 total), 70 is the total number of individuals preferring regular beer (column 2 total), and 150 is the total sample size. Hence, we see that

$$e_{12} = \frac{(\text{Row 1 Total})(\text{Column 2 Total})}{\text{Sample Size}}$$

Generalization of the expression shows that the following formula provides the expected frequencies for a contingency table in the test of independence.

EXPECTED FREQUENCIES FOR CONTINGENCY TABLES UNDER THE ASSUMPTION OF INDEPENDENCE

$$e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}} \tag{12.2}$$

Using the formula for male beer drinkers who prefer dark beer, we find an expected frequency of $e_{13} = (80)(30)/150 = 16.00$, as shown in Table 12.4. Use equation (12.2) to verify the other expected frequencies shown in Table 12.4.

The test procedure for comparing the observed frequencies of Table 12.3 with the expected frequencies of Table 12.4 is similar to the goodness of fit calculations made in Section 12.1. Specifically, the χ^2 value based on the observed and expected frequencies is computed as follows.

TEST STATISTIC FOR INDEPENDENCE

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \tag{12.3}$$

where

f_{ij} = observed frequency for contingency table category in row i and column j

e_{ij} = expected frequency for contingency table category in row i and column j based on the assumption of independence

Note: With n rows and m columns in the contingency table, the test statistic has a chi-square distribution with $(n - 1)(m - 1)$ degrees of freedom provided that the expected frequencies are five or more for all categories.

The double summation in equation (12.3) is used to indicate that the calculation must be made for all the cells in the contingency table.

By reviewing the expected frequencies in Table 12.4, we see that the expected frequencies are five or more for each category. We therefore proceed with the computation of the chi-square test statistic. The calculations necessary to compute the chi-square test statistic for determining whether beer preference is independent of the gender of the beer drinker are shown in Table 12.5. We see that the value of the test statistic is $\chi^2 = 6.12$.

The number of degrees of freedom for the appropriate chi-square distribution is computed by multiplying the number of rows minus 1 by the number of columns minus 1. With two rows and three columns, we have $(2 - 1)(3 - 1) = 2$ degrees of freedom. Just like the test for goodness of fit, the test for independence rejects H_0 if the differences between observed and expected frequencies provide a large value for the test statistic. Thus the test for independence is also an upper tail test. Using the chi-square table (Table 3 in Appendix B), we find the following information for 2 degrees of freedom.

The test for independence is always a one-tailed test with the rejection region in the upper tail of the chi-square distribution.

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**	**.005**
χ^2 Value (2 *df*)	4.605	5.991	7.378	9.210	10.597

$\chi^2 = 6.12$

The test statistic $\chi^2 = 6.12$ is between 5.991 and 7.378. Thus, the corresponding upper tail area or p-value is between .05 and .025. The Minitab or Excel procedures in Appendix F can be used to show p-value $= .0469$. With p-value $\leq \alpha = .05$, we reject the null hypothesis and conclude that beer preference is not independent of the gender of the beer drinker.

Computer software packages such as Minitab and Excel can be used to simplify the computations required for tests of independence. The input to these computer procedures is the contingency table of observed frequencies shown in Table 12.3. The software then computes the expected frequencies, the value of the χ^2 test statistic, and the p-value automatically. The Minitab and Excel procedures that can be used to conduct these tests of independence are presented in Appendixes 12.1 and 12.2. The Minitab output for the Alber's Brewery test of independence is shown in Figure 12.1.

Although no further conclusions can be made as a result of the test, we can compare the observed and expected frequencies informally to obtain an idea about the dependence between beer preference and gender. Refer to Tables 12.3 and 12.4. We see that male beer drinkers have higher observed than expected frequencies for both regular and dark beers, whereas female beer drinkers have a higher observed than expected frequency only for light

TABLE 12.5 COMPUTATION OF THE CHI-SQUARE TEST STATISTIC FOR DETERMINING WHETHER BEER PREFERENCE IS INDEPENDENT OF THE GENDER OF THE BEER DRINKER

Gender	**Beer Preference**	**Observed Frequency (f_{ij})**	**Expected Frequency (e_{ij})**	**Difference ($f_{ij} - e_{ij}$)**	**Squared Difference ($f_{ij} - e_{ij})^2$**	**Squared Difference Divided by Expected Frequency $(f_{ij} - e_{ij})^2/e_{ij}$**
Male	Light	20	26.67	−6.67	44.44	1.67
Male	Regular	40	37.33	2.67	7.11	0.19
Male	Dark	20	16.00	4.00	16.00	1.00
Female	Light	30	23.33	6.67	44.44	1.90
Female	Regular	30	32.67	−2.67	7.11	0.22
Female	Dark	10	14.00	−4.00	16.00	1.14
	Total	150				$\chi^2 = 6.12$

FIGURE 12.1 MINITAB OUTPUT FOR THE ALBER'S BREWERY TEST OF INDEPENDENCE

```
Expected counts are printed below observed counts

        Light    Regular     Dark    Total
   1       20         40       20       80
        26.67      37.33    16.00

   2       30         30       10       70
        23.33      32.67    14.00

Total      50         70       30      150

Chi-Sq = 6.122, DF = 2, P-Value = 0.047
```

beer. These observations give us insight about the beer preference differences between male and female beer drinkers.

Let us summarize the steps in a contingency table test of independence.

TEST OF INDEPENDENCE: A SUMMARY

1. State the null and alternative hypotheses.

H_0: The column variable is independent of the row variable

H_a: The column variable is not independent of the row variable

2. Select a random sample and record the observed frequencies for each cell of the contingency table.
3. Use equation (12.2) to compute the expected frequency for each cell.
4. Use equation (12.3) to compute the value of the test statistic.
5. Rejection rule:

p-value approach:	Reject H_0 if p-value $\leq \alpha$
Critical value approach:	Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

where α is the level of significance, with n rows and m columns providing $(n - 1)(m - 1)$ degrees of freedom.

NOTES AND COMMENTS

The test statistic for the chi-square tests in this chapter requires an expected frequency of five for each category. When a category has fewer than five, it is often appropriate to combine two adjacent categories to obtain an expected frequency of five or more in each category.

Exercises

Methods

SELF test

9. The following 2 × 3 contingency table contains observed frequencies for a sample of 200. Test for independence of the row and column variables using the χ^2 test with $\alpha = .05$.

	Column Variable		
Row Variable	A	B	C
P	20	44	50
Q	30	26	30

10. The following 3 × 3 contingency table contains observed frequencies for a sample of 240. Test for independence of the row and column variables using the χ^2 test with $\alpha = .05$.

	Column Variable		
Row Variable	A	B	C
P	20	30	20
Q	30	60	25
R	10	15	30

Applications

SELF test

11. One of the questions on the *BusinessWeek* Subscriber Study was, "In the past 12 months, when traveling for business, what type of airline ticket did you purchase most often?" The data obtained are shown in the following contingency table.

	Type of Flight	
Type of Ticket	Domestic Flights	International Flights
First class	29	22
Business/executive class	95	121
Full fare economy/coach class	518	135

Use $\alpha = .05$ and test for the independence of type of flight and type of ticket. What is your conclusion?

12. Visa Card USA studied how frequently consumers of various age groups use plastic cards (debit and credit cards) when making purchases (Associated Press, January 16, 2006). Sample data for 300 customers shows the use of plastic cards by four age groups.

	Age Group			
Payment	18–24	25–34	35–44	45 and over
Plastic	21	27	27	36
Cash or check	21	36	42	90

a. Test for the independence between method of payment and age group. What is the *p*-value? Using $\alpha = .05$, what is your conclusion?
b. If method of payment and age group are not independent, what observation can you make about how different age groups use plastic to make purchases?
c. What implications does this study have for companies such as Visa, MasterCard, and Discover?

13. With double-digit annual percentage increases in the cost of health insurance, more and more workers are likely to lack health insurance coverage (*USA Today,* January 23, 2004). The following sample data provide a comparison of workers with and without health insurance coverage for small, medium, and large companies. For the purposes of this study,

small companies are companies that have fewer than 100 employees. Medium companies have 100 to 999 employees, and large companies have 1000 or more employees. Sample data are reported for 50 employees of small companies, 75 employees of medium companies, and 100 employees of large companies.

	Health Insurance		
Size of Company	**Yes**	**No**	**Total**
Small	36	14	50
Medium	65	10	75
Large	88	12	100

a. Conduct a test of independence to determine whether employee health insurance coverage is independent of the size of the company. Use $\alpha = .05$. What is the p-value, and what is your conclusion?

b. The *USA Today* article indicated employees of small companies are more likely to lack health insurance coverage. Use percentages based on the preceding data to support this conclusion.

14. *Consumer Reports* measures owner satisfaction of various automobiles by asking the survey question, "Considering factors such as price, performance, reliability, comfort and enjoyment, would you purchase this automobile if you had it to do all over again?" (Consumer Reports website, January 2009). Sample data for 300 owners of four popular midsize sedans are as follows.

	Automobile				
Purchase Again	**Chevrolet Impala**	**Ford Taurus**	**Honda Accord**	**Toyata Camry**	**Total**
Yes	49	44	60	46	199
No	37	27	18	19	101

a. Conduct a test of independence to determine if the owner's intent to purchase again is independent of the automobile. Use a .05 level of significance. What is your conclusion?

b. *Consumer Reports* provides an owner satisfaction score for each automobile by reporting the percentage of owners who would purchase the same automobile if they could do it all over again. What are the *Consumer Reports* owner satisfaction scores for the Chevrolet Impala, Ford Taurus, Honda Accord, and Toyota Camry? Rank the four automobiles in terms of owner satisfaction.

c. Twenty-three different automobiles were reviewed in the *Consumer Reports* midsize sedan class. The overall owner satisfaction score for all automobiles in this class was 69. How do the United States manufactured automobiles (Impala and Taurus) compare to the Japanese manufactured automobiles (Accord and Camry) in terms of owner satisfaction? What is the implication of these findings on the future market share for these automobiles?

15. FlightStats, Inc., collects data on the number of flights scheduled and the number of flights flown at major airports throughout the United States. FlightStats data showed 56% of flights scheduled at Newark, La Guardia, and Kennedy airports were flown during a three-day snowstorm (*The Wall Street Journal,* February 21, 2006). All airlines say they always operate within set safety parameters—if conditions are too poor, they don't fly. The following data show a sample of 400 scheduled flights during the snowstorm.

	Airline				
Did It Fly?	**American**	**Continental**	**Delta**	**United**	**Total**
Yes	48	69	68	25	210
No	52	41	62	35	190

Use the chi-square test of independence with a .05 level of significance to analyze the data. What is your conclusion? Do you have a preference for which airline you would choose to fly during similar snowstorm conditions? Explain.

16. As the price of oil rises, there is increased worldwide interest in alternate sources of energy. A *Financial Times*/Harris Poll surveyed people in six countries to assess attitudes toward a variety of alternate forms of energy (Harris Interactive website, February 27, 2008). The data in the following table are a portion of the poll's findings concerning whether people favor or oppose the building of new nuclear power plants.

	Country					
Response	**Great Britain**	**France**	**Italy**	**Spain**	**Germany**	**United States**
Strongly favor	141	161	298	133	128	204
Favor more than oppose	348	366	309	222	272	326
Oppose more than favor	381	334	219	311	322	316
Strongly oppose	217	215	219	443	389	174

a. How large was the sample in this poll?
b. Conduct a hypothesis test to determine whether people's attitude toward building new nuclear power plants is independent of country. What is your conclusion?
c. Using the percentage of respondents who "strongly favor" and "favor more than oppose," which country has the most favorable attitude toward building new nuclear power plants? Which country has the least favorable attitude?

17. The National Sleep Foundation used a survey to determine whether hours of sleeping per night are independent of age (*Newsweek,* January 19, 2004). The following show the hours of sleep on weeknights for a sample of individuals age 49 and younger and for a sample of individuals age 50 and older.

	Hours of Sleep				
Age	**Fewer than 6**	**6 to 6.9**	**7 to 7.9**	**8 or more**	**Total**
49 or younger	38	60	77	65	240
50 or older	36	57	75	92	260

a. Conduct a test of independence to determine whether the hours of sleep on weeknights are independent of age. Use $\alpha = .05$. What is the p-value, and what is your conclusion?
b. What is your estimate of the percentage of people who sleep fewer than 6 hours, 6 to 6.9 hours, 7 to 7.9 hours, and 8 or more hours on weeknights?

18. Samples taken in three cities, Anchorage, Atlanta, and Minneapolis, were used to learn about the percentage of married couples with both the husband and the wife in the workforce (*USA Today,* January 15, 2006). Analyze the following data to see whether both the husband and wife being in the workforce is independent of location. Use a .05 level of

significance. What is your conclusion? What is the overall estimate of the percentage of married couples with both the husband and the wife in the workforce?

	Location		
In Workforce	**Anchorage**	**Atlanta**	**Minneapolis**
Both	57	70	63
Only one	33	50	90

19. On a syndicated television show the two hosts often create the impression that they strongly disagree about which movies are best. Each movie review is categorized as Pro ("thumbs up"), Con ("thumbs down"), or Mixed. The results of 160 movie ratings by the two hosts are shown here.

	Host B		
Host A	**Con**	**Mixed**	**Pro**
Con	24	8	13
Mixed	8	13	11
Pro	10	9	64

Use the chi-square test of independence with a .01 level of significance to analyze the data. What is your conclusion?

12.3 Goodness of Fit Test: Poisson and Normal Distributions

In Section 12.1 we introduced the goodness of fit test for a multinomial population. In general, the goodness of fit test can be used with any hypothesized probability distribution. In this section we illustrate the goodness of fit test procedure for cases in which the population is hypothesized to have a Poisson or a normal distribution. As we shall see, the goodness of fit test and the use of the chi-square distribution for the test follow the same general procedure used for the goodness of fit test in Section 12.1.

Poisson Distribution

Let us illustrate the goodness of fit test for the case in which the hypothesized population distribution is a Poisson distribution. As an example, consider the arrival of customers at Dubek's Food Market in Tallahassee, Florida. Because of some recent staffing problems, Dubek's managers asked a local consulting firm to assist with the scheduling of clerks for the checkout lanes. After reviewing the checkout lane operation, the consulting firm will make a recommendation for a clerk-scheduling procedure. The procedure, based on a mathematical analysis of waiting lines, is applicable only if the number of customers arriving during a specified time period follows the Poisson distribution. Therefore, before the scheduling process is implemented, data on customer arrivals must be collected and a statistical test conducted to see whether an assumption of a Poisson distribution for arrivals is reasonable.

We define the arrivals at the store in terms of the *number of customers* entering the store during 5-minute intervals. Hence, the following null and alternative hypotheses are appropriate for the Dubek's Food Market study.

H_0: The number of customers entering the store during 5-minute intervals has a Poisson probability distribution

H_a: The number of customers entering the store during 5-minute intervals does not have a Poisson distribution

If a sample of customer arrivals indicates H_0 cannot be rejected, Dubek's will proceed with the implementation of the consulting firm's scheduling procedure. However, if the sample leads to the rejection of H_0, the assumption of the Poisson distribution for the arrivals cannot be made, and other scheduling procedures will be considered.

To test the assumption of a Poisson distribution for the number of arrivals during weekday morning hours, a store employee randomly selects a sample of 128 5-minute intervals during weekday mornings over a three-week period. For each 5-minute interval in the sample, the store employee records the number of customer arrivals. In summarizing the data, the employee determines the number of 5-minute intervals having no arrivals, the number of 5-minute intervals having one arrival, the number of 5-minute intervals having two arrivals, and so on. These data are summarized in Table 12.6.

TABLE 12.6 OBSERVED FREQUENCY OF DUBEK'S CUSTOMER ARRIVALS FOR A SAMPLE OF 128 5-MINUTE TIME PERIODS

Number of Customers Arriving	Observed Frequency
0	2
1	8
2	10
3	12
4	18
5	22
6	22
7	16
8	12
9	6
Total	128

Table 12.6 gives the observed frequencies for the 10 categories. We now want to use a goodness of fit test to determine whether the sample of 128 time periods supports the hypothesized Poisson distribution. To conduct the goodness of fit test, we need to consider the expected frequency for each of the 10 categories under the assumption that the Poisson distribution of arrivals is true. That is, we need to compute the expected number of time periods in which no customers, one customer, two customers, and so on would arrive if, in fact, the customer arrivals follow a Poisson distribution.

The Poisson probability function, which was first introduced in Chapter 5, is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{12.4}$$

In this function, μ represents the mean or expected number of customers arriving per 5-minute period, x is the random variable indicating the number of customers arriving during a 5-minute period, and $f(x)$ is the probability that x customers will arrive in a 5-minute interval.

Before we use equation (12.4) to compute Poisson probabilities, we must obtain an estimate of μ, the mean number of customer arrivals during a 5-minute time period. The sample mean for the data in Table 12.6 provides this estimate. With no customers arriving in two 5-minute time periods, one customer arriving in eight 5-minute time periods, and so on, the total number of customers who arrived during the sample of 128 5-minute time periods is given by $0(2) + 1(8) + 2(10) + \cdots + 9(6) = 640$. The 640 customer arrivals over the sample of 128 periods provide a mean arrival rate of $\mu = 640/128 = 5$ customers per 5-minute period. With this value for the mean of the Poisson distribution, an estimate of the Poisson probability function for Dubek's Food Market is

$$f(x) = \frac{5^x e^{-5}}{x!} \tag{12.5}$$

This probability function can be evaluated for different values of x to determine the probability associated with each category of arrivals. These probabilities, which can also be found in Table 7 of Appendix B, are given in Table 12.7. For example, the probability of zero customers arriving during a 5-minute interval is $f(0) = .0067$, the probability of one customer arriving during a 5-minute interval is $f(1) = .0337$, and so on. As we saw in Section 12.1, the expected frequencies for the categories are found by multiplying the probabilities by the sample size. For example, the expected number of periods with zero arrivals is given by $(.0067)(128) = .86$, the expected number of periods with one arrival is given by $(.0337)(128) = 4.31$, and so on.

Before we make the usual chi-square calculations to compare the observed and expected frequencies, note that in Table 12.7, four of the categories have an expected

TABLE 12.7 EXPECTED FREQUENCY OF DUBEK'S CUSTOMER ARRIVALS, ASSUMING A POISSON DISTRIBUTION WITH $\mu = 5$

Number of Customers Arriving (x)	Poisson Probability $f(x)$	Expected Number of 5-Minute Time Periods with x Arrivals, $128f(x)$
0	.0067	0.86
1	.0337	4.31
2	.0842	10.78
3	.1404	17.97
4	.1755	22.46
5	.1755	22.46
6	.1462	18.71
7	.1044	13.36
8	.0653	8.36
9	.0363	4.65
10 or more	.0318	4.07
		Total 128.00

When the expected number in some category is less than five, the assumptions for the χ^2 test are not satisfied. When this happens, adjacent categories can be combined to increase the expected number to five.

frequency less than five. This condition violates the requirements for use of the chi-square distribution. However, expected category frequencies less than five cause no difficulty, because adjacent categories can be combined to satisfy the "at least five" expected frequency requirement. In particular, we will combine 0 and 1 into a single category and then combine 9 with "10 or more" into another single category. Thus, the rule of a minimum expected frequency of five in each category is satisfied. Table 12.8 shows the observed and expected frequencies after combining categories.

As in Section 12.1, the goodness of fit test focuses on the differences between observed and expected frequencies, $f_i - e_i$. Thus, we will use the observed and expected frequencies shown in Table 12.8, to compute the chi-square test statistic.

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

TABLE 12.8 OBSERVED AND EXPECTED FREQUENCIES FOR DUBEK'S CUSTOMER ARRIVALS AFTER COMBINING CATEGORIES

Number of Customers Arriving	Observed Frequency (f_i)	Expected Frequency (e_i)
0 or 1	10	5.17
2	10	10.78
3	12	17.97
4	18	22.46
5	22	22.46
6	22	18.72
7	16	13.37
8	12	8.36
9 or more	6	8.72
	Total 128	128.00

TABLE 12.9 COMPUTATION OF THE CHI-SQUARE TEST STATISTIC FOR THE DUBEK'S FOOD MARKET STUDY

Number of Customers Arriving (x)	Observed Frequency (f_i)	Expected Frequency (e_i)	Difference ($f_i - e_i$)	Squared Difference $(f_i - e_i)^2$	Squared Difference Divided by Expected Frequency $(f_i - e_i)^2/e_i$
0 or 1	10	5.17	4.83	23.28	4.50
2	10	10.78	−0.78	0.61	0.06
3	12	17.97	−5.97	35.62	1.98
4	18	22.46	−4.46	19.89	0.89
5	22	22.46	−0.46	0.21	0.01
6	22	18.72	3.28	10.78	0.58
7	16	13.37	2.63	6.92	0.52
8	12	8.36	3.64	13.28	1.59
9 or more	6	8.72	−2.72	7.38	0.85
Total	128	128.00			$\chi^2 = 10.96$

The calculations necessary to compute the chi-square test statistic are shown in Table 12.9. The value of the test statistic is $\chi^2 = 10.96$.

In general, the chi-square distribution for a goodness of fit test has $k - p - 1$ degrees of freedom, where k is the number of categories and p is the number of population parameters estimated from the sample data. For the Poisson distribution goodness of fit test, Table 12.9 shows $k = 9$ categories. Because the sample data were used to estimate the mean of the Poisson distribution, $p = 1$. Thus, there are $k - p - 1 = k - 2$ degrees of freedom. With $k = 9$, we have $9 - 2 = 7$ degrees of freedom.

Suppose we test the null hypothesis that the probability distribution for the customer arrivals is a Poisson distribution with a .05 level of significance. To test this hypothesis, we need to determine the p-value for the test statistic $\chi^2 = 10.96$ by finding the area in the upper tail of a chi-square distribution with 7 degrees of freedom. Using Table 3 of Appendix B, we find that $\chi^2 = 10.96$ provides an area in the upper tail greater than .10. Thus, we know that the p-value is greater than .10. Minitab or Excel procedures described in Appendix F can be used to show p-value $= .1404$. With p-value $> \alpha = .05$, we cannot reject H_0. Hence, the assumption of a Poisson probability distribution for weekday morning customer arrivals cannot be rejected. As a result, Dubek's management may proceed with the consulting firm's scheduling procedure for weekday mornings.

POISSON DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.

H_0: The population has a Poisson distribution

H_a: The population does not have a Poisson distribution

2. Select a random sample and

a. Record the observed frequency f_i for each value of the Poisson random variable.

b. Compute the mean number of occurrences μ.

3. Compute the expected frequency of occurrences e_i for each value of the Poisson random variable. Multiply the sample size by the Poisson probability of occurrence for each value of the Poisson random variable. If there are fewer than five expected occurrences for some values, combine adjacent values and reduce the number of categories as necessary.
4. Compute the value of the test statistic.

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

5. Rejection rule:

p-value approach:	Reject H_0 if *p*-value $\leq \alpha$
Critical value approach:	Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

where α is the level of significance and there are $k - 2$ degrees of freedom.

Normal Distribution

The goodness of fit test for a normal distribution is also based on the use of the chi-square distribution. It is similar to the procedure we discussed for the Poisson distribution. In particular, observed frequencies for several categories of sample data are compared to expected frequencies under the assumption that the population has a normal distribution. Because the normal distribution is continuous, we must modify the way the categories are defined and how the expected frequencies are computed. Let us demonstrate the goodness of fit test for a normal distribution by considering the job applicant test data for Chemline, Inc., listed in Table 12.10.

TABLE 12.10 CHEMLINE EMPLOYEE APTITUDE TEST SCORES FOR 50 RANDOMLY CHOSEN JOB APPLICANTS

71	66	61	65	54	93
60	86	70	70	73	73
55	63	56	62	76	54
82	79	76	68	53	58
85	80	56	61	61	64
65	62	90	69	76	79
77	54	64	74	65	65
61	56	63	80	56	71
79	84				

Chemline hires approximately 400 new employees annually for its four plants located throughout the United States. The personnel director asks whether a normal distribution applies for the population of test scores. If such a distribution can be used, the distribution would be helpful in evaluating specific test scores; that is, scores in the upper 20%, lower 40%, and so on, could be identified quickly. Hence, we want to test the null hypothesis that the population of test scores has a normal distribution.

Let us first use the data in Table 12.10 to develop estimates of the mean and standard deviation of the normal distribution that will be considered in the null hypothesis. We use the sample mean $\bar{x}$ and the sample standard deviation s as point estimators of the mean and standard deviation of the normal distribution. The calculations follow.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{3421}{50} = 68.42$$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{5310.0369}{49}} = 10.41$$

WEB file

Chemline

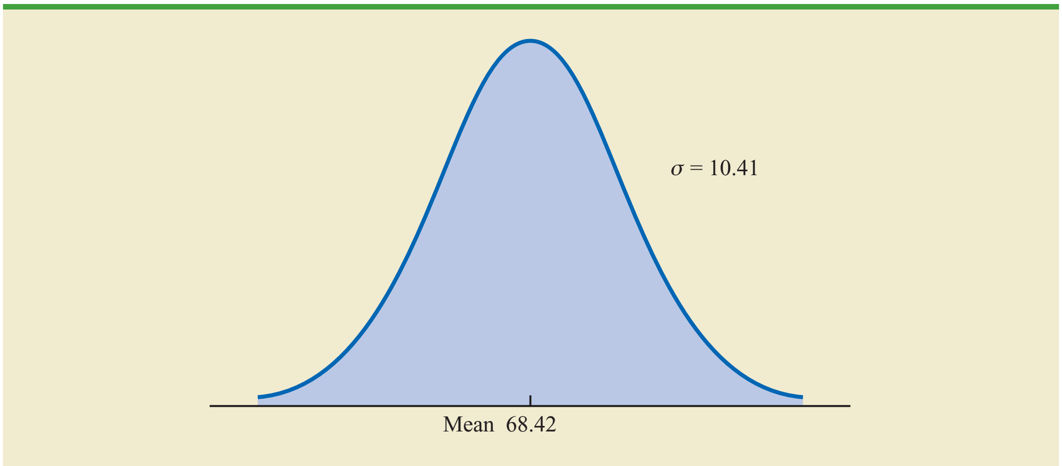
Using these values, we state the following hypotheses about the distribution of the job applicant test scores.

H_0: The population of test scores has a normal distribution with mean 68.42 and standard deviation 10.41

H_a: The population of test scores does not have a normal distribution with mean 68.42 and standard deviation 10.41

The hypothesized normal distribution is shown in Figure 12.2.

FIGURE 12.2 HYPOTHESIZED NORMAL DISTRIBUTION OF TEST SCORES FOR THE CHEMLINE JOB APPLICANTS



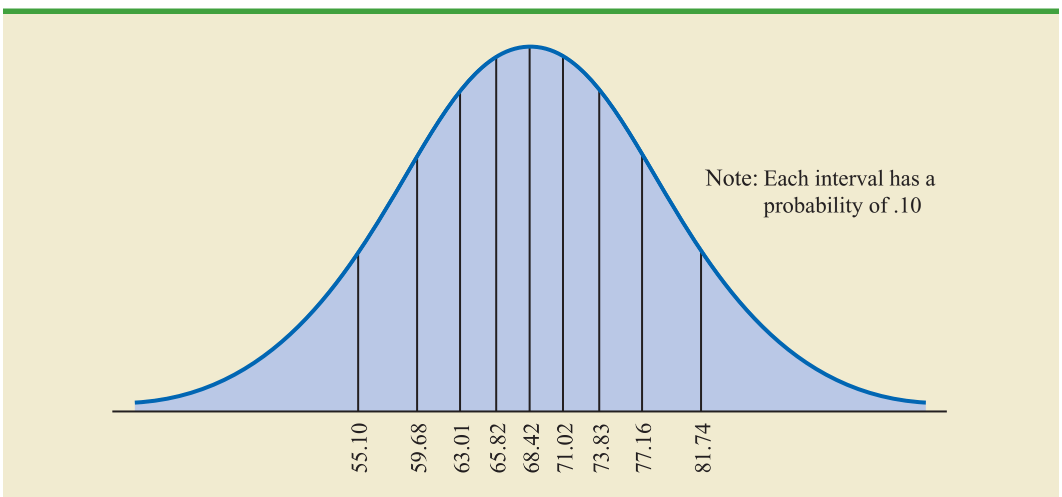
Now let us consider a way of defining the categories for a goodness of fit test involving a normal distribution. For the discrete probability distribution in the Poisson distribution test, the categories were readily defined in terms of the number of customers arriving, such as 0, 1, 2, and so on. However, with the continuous normal probability distribution, we must use a different procedure for defining the categories. We need to define the categories in terms of *intervals* of test scores.

Recall the rule of thumb for an expected frequency of at least five in each interval or category. We define the categories of test scores such that the expected frequencies will be at least five for each category. With a sample size of 50, one way of establishing categories is to divide the normal distribution into 10 equal-probability intervals (see Figure 12.3). With a sample size of 50, we would expect five outcomes in each interval or category, and the rule of thumb for expected frequencies would be satisfied.

With a continuous probability distribution, establish intervals such that each interval has an expected frequency of five or more.

Let us look more closely at the procedure for calculating the category boundaries. When the normal probability distribution is assumed, the standard normal probability tables can

FIGURE 12.3 NORMAL DISTRIBUTION FOR THE CHEMLINE EXAMPLE WITH 10 EQUAL-PROBABILITY INTERVALS



be used to determine these boundaries. First consider the test score cutting off the lowest 10% of the test scores. From Table 1 of Appendix B we find that the z value for this test score is -1.28. Therefore, the test score of $x = 68.42 - 1.28(10.41) = 55.10$ provides this cutoff value for the lowest 10% of the scores. For the lowest 20%, we find $z = -.84$, and thus $x = 68.42 - .84(10.41) = 59.68$. Working through the normal distribution in that way provides the following test score values.

Percentage	z	Test Score
10%	-1.28	$68.42 - 1.28(10.41) = 55.10$
20%	$-.84$	$68.42 - .84(10.41) = 59.68$
30%	$-.52$	$68.42 - .52(10.41) = 63.01$
40%	$-.25$	$68.42 - .25(10.41) = 65.82$
50%	$.00$	$68.42 + 0(10.41) = 68.42$
60%	$+.25$	$68.42 + .25(10.41) = 71.02$
70%	$+.52$	$68.42 + .52(10.41) = 73.83$
80%	$+.84$	$68.42 + .84(10.41) = 77.16$
90%	$+1.28$	$68.42 + 1.28(10.41) = 81.74$

These cutoff or interval boundary points are identified on the graph in Figure 12.3.

With the categories or intervals of test scores now defined and with the known expected frequency of five per category, we can return to the sample data of Table 12.10 and determine the observed frequencies for the categories. Doing so provides the results in Table 12.11.

With the results in Table 12.11, the goodness of fit calculations proceed exactly as before. Namely, we compare the observed and expected results by computing a χ^2 value. The computations necessary to compute the chi-square test statistic are shown in Table 12.12. We see that the value of the test statistic is $\chi^2 = 7.2$.

To determine whether the computed χ^2 value of 7.2 is large enough to reject H_0, we need to refer to the appropriate chi-square distribution tables. Using the rule for computing the number of degrees of freedom for the goodness of fit test, we have $k - p - 1 = 10 - 2 - 1 = 7$ degrees of freedom based on $k = 10$ categories and $p = 2$ parameters (mean and standard deviation) estimated from the sample data.

Suppose that we test the null hypothesis that the distribution for the test scores is a normal distribution with a .10 level of significance. To test this hypothesis, we need to determine the

TABLE 12.11 OBSERVED AND EXPECTED FREQUENCIES FOR CHEMLINE JOB APPLICANT TEST SCORES

Test Score Interval	Observed Frequency (f_i)	Expected Frequency (e_i)
Less than 55.10	5	5
55.10 to 59.68	5	5
59.68 to 63.01	9	5
63.01 to 65.82	6	5
65.82 to 68.42	2	5
68.42 to 71.02	5	5
71.02 to 73.83	2	5
73.83 to 77.16	5	5
77.16 to 81.74	5	5
81.74 and over	6	5
Total	50	50

TABLE 12.12 COMPUTATION OF THE CHI-SQUARE TEST STATISTIC FOR THE CHEMLINE JOB APPLICANT EXAMPLE

Test Score Interval	Observed Frequency (f_i)	Expected Frequency (e_i)	Difference ($f_i - e_i$)	Squared Difference ($f_i - e_i)^2$	Squared Difference Divided by Expected Frequency $(f_i - e_i)^2/e_i$
Less than 55.10	5	5	0	0	0.0
55.10 to 59.68	5	5	0	0	0.0
59.68 to 63.01	9	5	4	16	3.2
63.01 to 65.82	6	5	1	1	0.2
65.82 to 68.42	2	5	−3	9	1.8
68.42 to 71.02	5	5	0	0	0.0
71.02 to 73.83	2	5	−3	9	1.8
73.83 to 77.16	5	5	0	0	0.0
77.16 to 81.74	5	5	0	0	0.0
81.74 and over	6	5	1	1	0.2
Total	50	50			$\chi^2 = 7.2$

Estimating the two parameters of the normal distribution will cause a loss of two degrees of freedom in the χ^2 test.

p-value for the test statistic $\chi^2 = 7.2$ by finding the area in the upper tail of a chi-square distribution with 7 degrees of freedom. Using Table 3 of Appendix B, we find that $\chi^2 = 7.2$ provides an area in the upper tail greater than .10. Thus, we know that the p-value is greater than .10. Minitab or Excel procedures in Appendix F at the back of the book can be used to show $\chi^2 = 7.2$ provides a p-value $= .4084$. With p-value $> \alpha = .10$, the hypothesis that the probability distribution for the Chemline job applicant test scores is a normal distribution cannot be rejected. The normal distribution may be applied to assist in the interpretation of test scores. A summary of the goodness fit test for a normal distribution follows.

NORMAL DISTRIBUTION GOODNESS OF FIT TEST: A SUMMARY

1. State the null and alternative hypotheses.

H_0: The population has a normal distribution

H_a: The population does not have a normal distribution

2. Select a random sample and
   a. Compute the sample mean and sample standard deviation.
   b. Define intervals of values so that the expected frequency is at least five for each interval. Using equal probability intervals is a good approach.
   c. Record the observed frequency of data values f_i in each interval defined.

3. Compute the expected number of occurrences e_i for each interval of values defined in step 2(b). Multiply the sample size by the probability of a normal random variable being in the interval.

4. Compute the value of the test statistic.

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

5. Rejection rule:

p-value approach:	Reject H_0 if p-value $\leq \alpha$
Critical value approach:	Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

where α is the level of significance and there are $k - 3$ degrees of freedom.

Exercises

Methods

SELF test

20. Data on the number of occurrences per time period and observed frequencies follow. Use $\alpha = .05$ and the goodness of fit test to see whether the data fit a Poisson distribution.

Number of Occurrences	Observed Frequency
0	39
1	30
2	30
3	18
4	3

SELF test

21. The following data are believed to have come from a normal distribution. Use the goodness of fit test and $\alpha = .05$ to test this claim.

17	23	22	24	19	23	18	22	20	13	11	21	18	20	21
21	18	15	24	23	23	43	29	27	26	30	28	33	23	29

Applications

22. The number of automobile accidents per day in a particular city is believed to have a Poisson distribution. A sample of 80 days during the past year gives the following data. Do these data support the belief that the number of accidents per day has a Poisson distribution? Use $\alpha = .05$.

Number of Accidents	Observed Frequency (days)
0	34
1	25
2	11
3	7
4	3

23. The number of incoming phone calls at a company switchboard during 1-minute intervals is believed to have a Poisson distribution. Use $\alpha = .10$ and the following data to test the assumption that the incoming phone calls follow a Poisson distribution.

Number of Incoming Phone Calls During a 1-Minute Interval	Observed Frequency
0	15
1	31
2	20
3	15
4	13
5	4
6	2
Total	100

24. The weekly demand for a product is believed to be normally distributed. Use a goodness of fit test and the following data to test this assumption. Use $\alpha = .10$. The sample mean is 24.5 and the sample standard deviation is 3.

18	20	22	27	22
25	22	27	25	24
26	23	20	24	26
27	25	19	21	25
26	25	31	29	25
25	28	26	28	24

25. Use $\alpha = .01$ and conduct a goodness of fit test to see whether the following sample appears to have been selected from a normal distribution.

55	86	94	58	55	95	55	52	69	95	90	65	87	50	56
55	57	98	58	79	92	62	59	88	65					

After you complete the goodness of fit calculations, construct a histogram of the data. Does the histogram representation support the conclusion reached with the goodness of fit test? (*Note:* $\bar{x} = 71$ and $s = 17$.)

Summary

In this chapter we introduced the goodness of fit test and the test of independence, both of which are based on the use of the chi-square distribution. The purpose of the goodness of fit test is to determine whether a hypothesized probability distribution can be used as a model for a particular population of interest. The computations for conducting the goodness of fit test involve comparing observed frequencies from a sample with expected frequencies when the hypothesized probability distribution is assumed true. A chi-square distribution is used to determine whether the differences between observed and expected frequencies are large enough to reject the hypothesized probability distribution. We illustrated the goodness of fit test for multinomial, Poisson, and normal distributions.

A test of independence for two variables is an extension of the methodology employed in the goodness of fit test for a multinomial population. A contingency table is used to determine the observed and expected frequencies. Then a chi-square value is computed. Large chi-square values, caused by large differences between observed and expected frequencies, lead to the rejection of the null hypothesis of independence.

Glossary

Multinomial population A population in which each element is assigned to one and only one of several categories. The multinomial distribution extends the binomial distribution from two to three or more outcomes.
Goodness of fit test A statistical test conducted to determine whether to reject a hypothesized probability distribution for a population.
Contingency table A table used to summarize observed and expected frequencies for a test of independence.

Key Formulas

Test Statistic for Goodness of Fit

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \tag{12.1}$$

Expected Frequencies for Contingency Tables Under the Assumption of Independence

$$e_{ij} = \frac{(\text{Row } i \text{ Total})(\text{Column } j \text{ Total})}{\text{Sample Size}} \tag{12.2}$$

Test Statistic for Independence

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \tag{12.3}$$

Supplementary Exercises

26. In setting sales quotas, the marketing manager makes the assumption that order potentials are the same for each of four sales territories. A sample of 200 sales follows. Should the manager's assumption be rejected? Use $\alpha = .05$.

Sales Territories			
I	II	III	IV
60	45	59	36

27. Seven percent of mutual fund investors rate corporate stocks "very safe," 58% rate them "somewhat safe," 24% rate them "not very safe," 4% rate them "not at all safe," and 7% are "not sure." A *BusinessWeek*/Harris poll asked 529 mutual fund investors how they would rate corporate bonds on safety. The responses are as follows.

Safety Rating	Frequency
Very safe	48
Somewhat safe	323
Not very safe	79
Not at all safe	16
Not sure	63
Total	529

Do mutual fund investors' attitudes toward corporate bonds differ from their attitudes toward corporate stocks? Support your conclusion with a statistical test. Use $\alpha = .01$.

28. Since 2000, the Toyota Camry, Honda Accord, and Ford Taurus have been the three best-selling passenger cars in the United States. Sales data for 2003 indicated market shares among the top three as follows: Toyota Camry 37%, Honda Accord 34%, and Ford Taurus 29% (*The World Almanac,* 2004). Assume a sample of 1200 sales of passenger cars during the first quarter of 2004 shows the following.

Passenger Car	Units Sold
Toyota Camry	480
Honda Accord	390
Ford Taurus	330

Can these data be used to conclude that the market shares among the top three passenger cars have changed during the first quarter of 2004? What is the *p*-value? Use a .05 level of significance. What is your conclusion?

29. A regional transit authority is concerned about the number of riders on one of its bus routes. In setting up the route, the assumption is that the number of riders is the same on every day from Monday through Friday. Using the following data, test with $\alpha = .05$ to determine whether the transit authority's assumption is correct.

Day	Number of Riders
Monday	13
Tuesday	16
Wednesday	28
Thursday	17
Friday	16

30. The results of *Computerworld*'s Annual Job Satisfaction Survey showed that 28% of information systems (IS) managers are very satisfied with their job, 46% are somewhat satisfied, 12% are neither satisfied nor dissatisfied, 10% are somewhat dissatisfied, and 4% are very dissatisfied. Suppose that a sample of 500 computer programmers yielded the following results.

Category	Number of Respondents
Very satisfied	105
Somewhat satisfied	235
Neither	55
Somewhat dissatisfied	90
Very dissatisfied	15

Use $\alpha = .05$ and test to determine whether the job satisfaction for computer programmers is different from the job satisfaction for IS managers.

31. A sample of parts provided the following contingency table data on part quality by production shift.

Shift	Number Good	Number Defective
First	368	32
Second	285	15
Third	176	24

Use $\alpha = .05$ and test the hypothesis that part quality is independent of the production shift. What is your conclusion?

32. *The Wall Street Journal* Subscriber Study showed data on the employment status of subscribers. Sample results corresponding to subscribers of the eastern and western editions are shown here.

	Region	
Employment Status	**Eastern Edition**	**Western Edition**
Full-time	1105	574
Part-time	31	15
Self-employed/consultant	229	186
Not employed	485	344

Use $\alpha = .05$ and test the hypothesis that employment status is independent of the region. What is your conclusion?

33. A lending institution supplied the following data on loan approvals by four loan officers. Use $\alpha = .05$ and test to determine whether the loan approval decision is independent of the loan officer reviewing the loan application.

	Loan Approval Decision	
Loan Officer	**Approved**	**Rejected**
Miller	24	16
McMahon	17	13
Games	35	15
Runk	11	9

34. A Pew Research Center survey asked respondents if they would rather live in a place with a slower pace of life or a place with a faster pace of life (*USA Today*, February 13, 2009). Consider the following data showing a sample of preferences expressed by 150 men and 150 women.

	Preferred Pace of Life		
Respondent	**Slower**	**No Preference**	**Faster**
Men	102	9	39
Women	111	12	27

a. Combine the samples of men and women. What is the overall percentage of respondents who prefer to live in a place with a slower pace of life? What is the overall percentage of respondents who prefer to live in a place with a faster pace of life? What is your conclusion?

b. Is the preferred pace of life independent of the respondent? Use $\alpha = .05$. What is your conclusion? What is your recommendation?

35. Barna Research Group collected data showing church attendance by age group (*USA Today,* November 20, 2003). Use the sample data to determine whether attending church is independent of age. Use a .05 level of significance. What is your conclusion? What conclusion can you draw about church attendance as individuals grow older?

	Church Attendance		
Age	**Yes**	**No**	**Total**
20 to 29	31	69	100
30 to 39	63	87	150
40 to 49	94	106	200
50 to 59	72	78	150

36. The following data were collected on the number of emergency ambulance calls for an urban county and a rural county in Virginia.

		Day of Week							
		Sun	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**	**Sat**	**Total**
County	**Urban**	61	48	50	55	63	73	43	393
	Rural	7	9	16	13	9	14	10	78
	Total	68	57	66	68	72	87	53	471

Conduct a test for independence using $\alpha = .05$. What is your conclusion?

37. A random sample of final examination grades for a college course follows.

55	85	72	99	48	71	88	70	59	98	80	74	93	85	74
82	90	71	83	60	95	77	84	73	63	72	95	79	51	85
76	81	78	65	75	87	86	70	80	64					

Use $\alpha = .05$ and test to determine whether a normal distribution should be rejected as being representative of the population's distribution of grades.

38. The office occupancy rates were reported for four California metropolitan areas. Do the following data suggest that the office vacancies were independent of metropolitan area? Use a .05 level of significance. What is your conclusion?

Occupancy Status	Los Angeles	San Diego	San Francisco	San Jose
Occupied	160	116	192	174
Vacant	40	34	33	26

39. A salesperson makes four calls per day. A sample of 100 days gives the following frequencies of sales volumes.

Number of Sales	Observed Frequency (days)
0	30
1	32
2	25
3	10
4	3
Total	100

Records show sales are made to 30% of all sales calls. Assuming independent sales calls, the number of sales per day should follow a binomial distribution. The binomial probability function presented in Chapter 5 is

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

For this exercise, assume that the population has a binomial distribution with $n = 4$, $p = .30$, and $x = 0, 1, 2, 3$, and 4.

a. Compute the expected frequencies for $x = 0, 1, 2, 3$, and 4 by using the binomial probability function. Combine categories if necessary to satisfy the requirement that the expected frequency is five or more for all categories.
b. Use the goodness of fit test to determine whether the assumption of a binomial distribution should be rejected. Use $\alpha = .05$. Because no parameters of the binomial distribution were estimated from the sample data, the degrees of freedom are $k - 1$ when k is the number of categories.

Case Problem A Bipartisan Agenda for Change

In a study conducted by Zogby International for the *Democrat and Chronicle,* more than 700 New Yorkers were polled to determine whether the New York state government works. Respondents surveyed were asked questions involving pay cuts for state legislators, restrictions on lobbyists, term limits for legislators, and whether state citizens should be able to put matters directly on the state ballot for a vote (*Democrat and Chronicle,* December 7, 1997). The results regarding several proposed reforms had broad support, crossing all demographic and political lines.

Suppose that a follow-up survey of 100 individuals who live in the western region of New York was conducted. The party affiliation (Democrat, Independent, Republican) of each individual surveyed was recorded, as well as their responses to the following three questions.

1. Should legislative pay be cut for every day the state budget is late?
   Yes ____ No ____
2. Should there be more restrictions on lobbyists?
   Yes ____ No ____
3. Should there be term limits requiring that legislators serve a fixed number of years?
   Yes ____ No ____

WEB file NYReform

The responses were coded using 1 for a Yes response and 2 for a No response. The complete data set is available in the file named NYReform.

Managerial Report

1. Use descriptive statistics to summarize the data from this study. What are your preliminary conclusions about the independence of the response (Yes or No) and party affiliation for each of the three questions in the survey?
2. With regard to question 1, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = .05$.
3. With regard to question 2, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = .05$.
4. With regard to question 3, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = .05$.
5. Does it appear that there is broad support for change across all political lines? Explain.

Appendix 12.1 Tests of Goodness of Fit and Independence Using Minitab

Goodness of Fit Test

This Minitab procedure can be used for a goodness of fit test of a multinomial population in Section 12.1. The user must obtain the observed frequency and the hypothesized proportion for each of the *k* categories. The observed frequencies are entered in Column C1 and the hypothesized proportions are entered in Column C2. Using the Scott Marketing Research example presented in Section 12.1, Column C1 is labeled Observed and Column C2 is labeled Proportion. Enter the observed frequencies 48, 98, and 54 in Column C1 and enter the hypothesized proportions .30, .50, and .20 in Column C2. The Minitab steps for the goodness of fit test follow.

Step 1. Select the **Stat** menu
Step 2. Select **Tables**
Step 3. Choose **Chi-Square Goodness of Fit Test (One Variable)**
Step 4. When the Chi-Square Goodness of Fit Test dialog box appears;
Select **Observed counts**
Enter Cl in the **Observed counts** box
Select **Specific proportions**
Enter C2 in the **Specific proportions** box
Click **OK**

Test of Independence

We begin with a new Minitab worksheet and enter the observed frequency data for the Alber's Brewery example from Section 12.2 into columns 1, 2, and 3, respectively. Thus, we entered the observed frequencies corresponding to a light beer preference (20 and 30) in C1, the observed frequencies corresponding to a regular beer preference (40 and 30) in C2, and the observed frequencies corresponding to a dark beer preference (20 and 10) in C3. The Minitab steps for the test of independence are as follows.

Step 1. Select the **Stat** menu
Step 2. Select **Tables**
Step 3. Choose **Chi-Square Test (Two-Way Table in Worksheet)**
Step 4. When the Chi-Square Test dialog box appears:
Enter C1-C3 in the **Columns containing the table** box
Click **OK**

Appendix 12.2 Tests of Goodness of Fit and Independence Using Excel

Goodness of Fit Test

WEB file

FitTest

This Excel procedure can be used for a goodness of fit test for the multinomial distribution in Section 12.1 and the Poisson and normal distributions in Section 12.3. The user must obtain the observed frequencies, calculate the expected frequencies, and enter both the observed and expected frequencies in an Excel worksheet.

The observed frequencies and expected frequencies for the Scott Market Research example presented in Section 12.1 are entered in columns A and B as shown in Figure 12.4. The test statistic $\chi^2 = 7.34$ is calculated in column D. With $k = 3$ categories, the user enters the degrees of freedom $k - 1 = 3 - 1 = 2$ in cell D11. The CHIDIST function provides the p-value in cell D13. The background worksheet shows the cell formulas.

Test of Independence

WEB file

Independence

The Excel procedure for the test of independence requires the user to obtain the observed frequencies and enter them in the worksheet. The Alber's Brewery example from Section 12.2 provides the observed frequencies, which are entered in cells B7 to D8 as shown in the worksheet in Figure 12.5. The cell formulas in the background worksheet show the procedure used to compute the expected frequencies. With two rows and three columns, the user enters the degrees of freedom $(2 - 1)(3 - 1) = 2$ in cell E22. The CHITEST function provides the p-value in cell E24.

FIGURE 12.4 EXCEL WORKSHEET FOR THE SCOTT MARKETING RESEARCH GOODNESS OF FIT TEST

	A	B	C	D	E
1	**Goodness of Fit Test**				
2					
3	Observed	Expected			
4	Frequency	Frequency		Calculations	
5	48	60		=(A5-B5)^2/B5	
6	98	100		=(A6-B6)^2/B6	
7	54	40		=(A7-B7)^2/B7	
8					
9			Test Statistic	=SUM(D5:D7)	
10					
11			Degrees of Freedom	2	
12					
13			p-Value	=CHIDIST(D9,D11)	
14					

	A	B	C	D	E
1	**Goodness of Fit Test**				
2					
3	Observed	Expected			
4	Frequency	Frequency		Calculations	
5	48	60		2.40	
6	98	100		0.04	
7	54	40		4.90	
8					
9			Test Statistic	7.34	
10					
11			Degrees of Freedom	2	
12					
13			p-Value	0.0255	
14					

FIGURE 12.5 EXCEL WORKSHEET FOR THE ALBER'S BREWERY TEST OF INDEPENDENCE

	A	B	C	D	E	F
1	**Test of Independence**					
2						
3	**Observed Frequencies**					
4						
5			**Beer Peference**			
6	**Gender**	Light	Regular	Dark	Total	
7	Male	20	40	20	=SUM(B7:D7)	
8	Female	30	30	10	=SUM(B8:D8)	
9	Total	=SUM(B7:B8)	=SUM(C7:C8)	=SUM(D7:D8)	=SUM(E7:E8)	
10						
11						
12	**Expected Frequencies**					
13						
14			**Beer Peference**			
15	**Gender**	Light	Regular	Dark	Total	
16	Male	=E7*B$9/$E$9	=E7*C$9/$E$9	=E7*D$9/$E$9	=SUM(B16:D16)	
17	Female	=E8*B$9/$E$9	=E8*C$9/$E$9	=E8*D$9/$E$9	=SUM(B17:D17)	
18	Total	=SUM(B16:B17)	=SUM(C16:C17)	=SUM(D16:D17)	=SUM(E16:E17)	
19						
20				**Test Statistic**	=CHIINV(E24,E22)	
21						
22				**Degrees of Freedom**	2	
23						
24				**p-value**	=CHITEST(B7:D8,B16:D17)	
25						

	A	B	C	D	E	F
1	**Test of Independence**					
2						
3	**Observed Frequencies**					
4						
5			**Beer Peference**			
6	**Gender**	Light	Regular	Dark	Total	
7	Male	20	40	20	80	
8	Female	30	30	10	70	
9	Total	50	70	30	150	
10						
11						
12	**Expected Frequencies**					
13						
14			**Beer Peference**			
15	**Gender**	Light	Regular	Dark	Total	
16	Male	26.67	37.33	16	80	
17	Female	23.33	32.67	14	70	
18	Total	50	70	30	150	
19						
20				**Test Statistic**	6.12	
21						
22				**Degrees of Freedom**	2	
23						
24				**p-value**	0.0468	
25						