Enter **25** and **15** (degrees of freedom) as the second argument and third arguments (second and third question marks) – the **Numeric Expression** box should now have **CDF.CHISQ(Fratio,25,15)**
Click **OK**

The value 0.9594 will be returned in the first row of column 4 in the Data Editor, now named **Cumprob**. This is a cumulative probability, i.e. the area under the curve to the left of $F = 2.40$. The upper tail area is $1 - 0.9594 = 0.0406$. For a two-tailed test, as we did in Section 11.2, the p-value is twice this area, p-value $= 0.081$.

Chapter 12

Tests of Goodness of Fit and Independence

Learning objectives

After studying this chapter and doing the exercises, you should be able to construct and interpret the results of goodness of fit tests, using the chi-squared distribution, for several situations:

1 A multinomial population with given probabilities.
2 A test of independence in a two-way contingency table.
3 A Poisson distribution.
4 A normal distribution.

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

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

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 of all others, and the probabilities of the outcomes remain the same at each trial.

As an example, consider a market share study being conducted by Scott Market Research. Over the past year market shares stabilized at 30 per cent for company A, 50 per cent for company B and 20 per cent for company C. Recently company C developed a 'new and improved' product to replace its current offering in the market. Company C retained Scott Market Research to assess 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. So we have a multinomial population with three possible outcomes. We use the following notation:

$$\pi_A = \text{market share for company A}$$
$$\pi_B = \text{market share for company B}$$
$$\pi_C = \text{market share for company C}$$

Scott Market Research will conduct a sample survey and find the sample proportion preferring each company's product. A hypothesis test will then be done to assess whether the new product will lead to a change in market shares. The null and alternative hypotheses are:

$$H_0\text{: } \pi_A = 0.30,\ \pi_B = 0.50 \text{ and } \pi_C = 0.20$$
$$H_1\text{: The population proportions are not } \pi_A = 0.30,\ \pi_B = 0.50 \text{ and } \pi_C = 0.20$$

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

Statistics in Practice

National lotteries

On 23 December 2008, amid global financial upheavals, the *Guardian* newspaper in the UK had a headline announcing **El Gordo brings £2 bn sparkle to Spain**. The Spanish national Christmas lottery, Lotería de Navidad or El Gordo ('The Fat One'), is traditionally drawn on 22 December. El Gordo claims to be the only lottery in the world that distributes the equivalent of more than one billion US dollars as a result of a single draw.

Many countries have government-sponsored national lotteries, some drawn on a weekly rather than an annual basis. Ireland and the UK, for example, each have a twice-weekly Lotto draw. In the Irish Lotto game, each ticket-holder selects six different numbers in the range 1 to 45. If these numbers match those selected randomly when the draw takes place, the ticket-holder wins a share in the jackpot prize. The UK Lotto game is similar, but the six numbers are chosen in the range 1 to 49. The chance of winning a jackpot share with a single ticket in the UK Lotto is therefore smaller than in the Irish Lotto: about 1 in 14 million for the UK draw compared to about 1 in 8 million in the Irish draw.

It is important that any national lottery is conducted fairly and transparently. The monitoring of the UK lottery is under the supervision of a public body called the National Lottery Commission (NLC). As part of its monitoring role, the NLC commissions statistical analyses to check that the lottery games are being conducted fairly and that numbers are being drawn randomly. The Centre for the Study of Gambling at the University of Salford, UK has carried out recent commissioned analyses. In its 2004 report on the main Lotto draw, one of the specific objectives was to report on tests for equality of frequency for each Lotto number drawn.

The objective of testing for equality of frequency for each Lotto number drawn was met using a statistical test known as a chi-squared test. This test compares the actual frequency with which the Lotto numbers were drawn with those 'expected' assuming equal probabilities of selection for all 49 numbers, but allowing for possible variation caused by the randomness of the selection process. Other specific aspects of randomness, such as independence between draws, were tested similarly using chi-squared tests, by comparing observed frequencies with those expected assuming randomness to prevail. In no case was any evidence of non-randomness found.

In this chapter you will learn how chi-squared tests like those described here are done.

A lottery hawker arranges the traditional El Gordo's tickets as he waits for customers in Madrid, Spain. © Fernando Alvarado/epa/Corbis.

The market research firm has used a consumer panel of 200 customers for the study, in which each individual is asked to specify a purchase preference for one of three alternatives: company A's product, company B's product and company C's new product. This is equivalent to a multinomial experiment consisting of 200 trials. The 200 responses are summarized here.

Observed frequency		
Company A's product	Company B's product	Company C's new product
48	98	54

We now can do a **goodness of fit test** to assess 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. The next step is therefore to compute expected purchase preferences for the 200 customers under the assumption that $\pi_A = 0.30$, $\pi_B = 0.50$ and $\pi_C = 0.20$. The expected frequency for each category is found by multiplying the sample size of 200 by the hypothesized proportion for the category.

Expected frequency		
Company A's product	Company B's product	Company C's new product
200(0.30) = 60	200(0.50) = 100	200(0.20) = 40

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

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

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-squared distribution with $k - 1$ degrees of freedom provided that the expected frequencies are five or more for all categories.

In the Scott Market Research example we use the sample data to test the hypothesis that the multinomial population has the proportions $\pi_A = 0.30$, $\pi_B = 0.50$ and $\pi_C = 0.20$. We shall use level of significance $\alpha = 0.05$. The computation of the chi-squared test statistic is shown in Table 12.1, giving $\chi^2 = 7.34$.

Table 12.1 Computation of the chi-squared test statistic for the Scott Market Research market share study

	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	0.30	48	60	−12	144	2.40
Company B	0.50	98	100	−2	4	0.04
Company C	0.20	54	40	14	196	4.90
Total		200				$\chi^2 = 7.34$

We shall reject the null hypothesis if the differences between the observed and expected frequencies are large, which in turn will result in a large value for the test statistic. Hence the test of goodness of fit will always be an upper tail test. With $k - 1 = 3 - 1 = 2$ degrees of freedom, the chi-squared table (Table 3 of Appendix B) provides the following (an introduction to the chi-squared distribution and the use of the chi-squared table were presented in Section 11.1).

Area in upper tail	0.10	0.05	0.025	0.01
χ^2 value (2 df)	4.605	5.991	7.378	9.210
			↑ $\chi^2 = 7.34$	

The test statistic $\chi^2 = 7.34$ is between 5.991 and 7.378 (very close to 7.378), so the corresponding upper tail area or p-value must be between 0.05 and 0.025 (very close to 0.025). With p-value $< \alpha = 0.05$, we reject H_0 and conclude that the introduction of the new product by company C may alter the current market share structure. MINITAB, PASW or EXCEL can be used to show that $\chi^2 = 7.34$ provides a p-value $= 0.0255$ (see the Software Section at the end of the chapter).

Instead of using the p-value, we could use the critical value approach to draw the same conclusion. With $\alpha = 0.05$ and 2 degrees of freedom, the critical value for the test statistic is $\chi^2 = 5.991$. The upper tail rejection rule becomes

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

With $\chi^2 = 7.34 > 5.991$, we reject H_0. The p-value approach and critical value approach provide the same conclusion.

Although the test itself does not directly tell us about *how* market shares may change, we can compare the observed and expected frequencies descriptively to get an idea of the change in market structure. We see that the observed frequency of 54 for company C is larger than the expected frequency of 40. Because the latter 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. Similar comparisons for the other two companies suggest that company C's gain in market share will hurt company A more than company B.

Here are the steps for doing 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_1: 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.

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

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

H_0: $\pi_A = 0.40$, $\pi_B = 0.40$, $\pi_C = 0.20$

H_1: The population proportions are not $\pi_A = 0.40$, $\pi_B = 0.40$, $\pi_C = 0.20$

A sample of size 200 yielded 60 in category A, 120 in category B, and 20 in category C. Use $\alpha = 0.01$ 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, i.e.

$$H_0: \pi_A = \pi_B = \pi_C = \pi_D = 0.25$$

A sample of size 300 yielded the following results.

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

Use $\alpha = 0.05$ to determine whether H_0 should be rejected. What is the p-value?

Applications

3 One of the questions on the *Business Week* Subscriber Study was, 'When making investment purchases, do you use full service or discount brokerage firms?' Survey results showed that 264 respondents use full service brokerage firms only, 255 use discount brokerage firms only and 229 use both full service and discount firms. Use $\alpha = 0.10$ to determine whether there are any differences in preference among the three service choices.

4 How well do airline companies serve their customers? A study by *Business Week* showed the following customer ratings: 3 per cent excellent, 28 per cent good, 45 per cent fair and 24 per cent poor. 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. Taking the figures from the *Business Week* study as 'population' values, is the distribution of the customer ratings for telephone companies different from the distribution of customer ratings for airline companies? Test with $\alpha = 0.01$. What is your conclusion?

5 In setting sales quotas, the marketing manager of a multinational company makes the assumption that order potentials are the same for each of four sales territories in Africa. A sample of 200 sales follows. Should the manager's assumption be rejected? Use $\alpha = 0.05$.

Sales territories			
1	2	3	4
60	45	59	36

6 A community park will open soon in a large European city. A sample of 210 individuals are asked to state their preference for when they would most like to visit the park. The sample results follow.

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
20	30	30	25	35	20	50

In developing a staffing plan, should the park manager plan on the same number of individuals visiting the park each day? Support your conclusion with a statistical test. Use $\alpha = 0.05$.

7 The results of *ComputerWorld's* Annual Job Satisfaction Survey showed that 28 per cent of information systems (IS) managers are very satisfied with their job, 46 per cent are somewhat satisfied, 12 per cent are neither satisfied or dissatisfied, 10 per cent are somewhat dissatisfied and 4 per cent 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

Taking the *ComputerWorld* figures as 'population' values, use $\alpha = 0.05$ and test to determine whether the job satisfaction for computer programmers is different from the job satisfaction for IS managers.

12.2 Test of independence

Another important application of the chi-squared distribution involves testing for the independence of two variables. Consider a study conducted by the Real Ale Brewery, which manufactures and distributes three types of beer: light ale, lager and best bitter. 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 between male and female beer drinkers. If beer preference is independent of gender, a single advertising campaign will be initiated for all of the Real Ale 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 ale, lager or best bitter) is independent of the gender of the beer drinker (male, female). The hypotheses for this test are:

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

H_1: Beer preference is not independent of the gender of the beer drinker

Table 12.2 Contingency table for beer preference and gender of beer drinker

	Beer preference		
Gender	Light ale	Lager	Best bitter
Male	cell(1,1)	cell(1,2)	cell(1,3)
Female	cell(2,1)	cell(2,2)	cell(2,3)

Table 12.2 can be used to describe the situation. The population under study is all male and female beer drinkers. A sample can be selected from this population and each individual asked to state his or her preference among the three Real Ale 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 lager (cell (1,2)), a female preferring light ale (cell (2,1)), a female preferring best bitter (cell (2,3)) and so on. Because we have listed all possible combinations of beer preference and gender – in other words, listed all possible contingencies – Table 12.2 is called a **contingency table**. The test of independence 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 first-choice preference. The cross-tabulation in Table 12.3 summarizes the responses. 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 favoured light ale, 40 were men who favoured lager, 20 were men who favoured best bitter 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-squared 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. 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 ale, 70 prefer lager and 30 prefer best bitter. In terms of fractions we conclude that 50/150 of the beer drinkers prefer light ale, 70/150 prefer lager and 30/150 prefer best bitter. If the *independence* assumption is valid, we argue that these fractions must be applicable to both male and female beer drinkers. So we would expect the sample of 80 male beer drinkers to show that $(50/150)80 = 26.67$ prefer light ale, $(70/150)80 = 37.33$ prefer lager, and $(30/150)80 = 16$ prefer best bitter. Application of the same fractions to the 70 female beer drinkers provides the expected frequencies shown in Table 12.4.

Table 12.3 Sample results for beer preferences of male and female beer drinkers (observed frequencies)

	Beer preference			
Gender	Light ale	Lager	Best bitter	Total
Male	20	40	20	80
Female	30	30	10	70
Total	50	70	30	150

Table 12.4 Expected frequencies if beer preference is independent of the gender of the beer drinker

	Beer preference			
Gender	Light ale	Lager	Best bitter	Total
Male	26.67	37.33	16.00	80
Female	23.33	32.67	14.00	70
Total	50.00	70.00	30.00	150

Let e_{ij} denote the expected frequency for the contingency table category in row i and column j. With this notation, consider the expected frequency calculation for males (row $i = 1$) who prefer lager (column $j = 2$): that is, expected frequency e_{12}. The argument above showed that

$$e_{12} = \left(\frac{70}{150}\right) 80 = 37.33$$

This expression can be written slightly differently as

$$e_{12} = \left(\frac{70}{150}\right) 80 = \frac{(80)(70)}{(150)} = 37.33$$

Note that the 80 in the expression is the total number of males (row 1 total), 70 is the total number of individuals preferring lager (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 this 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

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

Using this formula for male beer drinkers who prefer best bitter, 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

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

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-squared 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.

The expected frequencies are five or more for each category. We therefore proceed with the computation of the chi-squared test statistic, as 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-squared distribution is computed by multiplying the number of rows minus one by the number of columns minus one. With two rows and three columns, we have $(2 - 1)(3 - 1) = 3$ 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. So the test for independence is also an upper tail test. Using the chi-squared table (Table 3 of Appendix B), we find that the upper tail area or *p*-value at $\chi^2 = 6.12$ is between 0.025 and 0.05. At the 0.05 level of significance, *p*-value $< \alpha = 0.05$. We reject the null hypothesis of independence and conclude that beer preference is not independent of the gender of the beer drinker.

Computer software packages such as PASW, MINITAB and EXCEL can simplify the computations for a test of independence and provide the *p*-value for the test (see the

Table 12.5 Computation of the chi-squared 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 ale	20	26.67	−6.67	44.44	1.67
Male	Lager	40	37.33	2.67	7.11	0.19
Male	Best bitter	20	16.00	4.00	16.00	1.00
Female	Light ale	30	23.33	6.67	44.44	1.90
Female	Lager	30	32.67	−2.67	7.11	0.22
Female	Best bitter	10	14.00	−4.00	16.00	1.14
	Total	**150**				$\chi^2 = 6.12$

Software Section at the end of the chapter). In the Real Ale Brewery example, EXCEL, MINITAB or PASW shows *p*-value = 0.0468.

The test itself does not tell us directly about the nature of the dependence between beer preference and gender, but we can compare the observed and expected frequencies descriptively to get an idea. Refer to Tables 12.3 and 12.4. Male beer drinkers have higher observed than expected frequencies for both lager and best bitter beers, whereas female beer drinkers have a higher observed than expected frequency only for light ale beer. These observations give us insight about the beer preference differences between male and female beer drinkers.

Here are 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_1: 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 for the test, with *n* rows and *m* columns providing $(n - 1) \times (m - 1)$ degrees of freedom.

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

Exercises

Methods

8 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 $\alpha = 0.05$.

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

9 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 $\alpha = 0.05$.

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

Applications

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

Type of ticket	Type of flight: Domestic flights	International flights
First class	29	22
Business class	95	121
Economy class	518	135

Use $\alpha = 0.05$ and test for the independence of type of flight and type of ticket. What is your conclusion?

11 First-destination jobs for business and engineering graduates are classified by industry as shown in the following table.

Degree major	Industry: Oil	Chemical	Electrical	Computer
Business	30	15	15	40
Engineering	30	30	20	20

Use $\alpha = 0.01$ and test for independence of degree major and industry type.

12 Businesses are increasingly placing orders online. The Performance Measurement Group collected data on the rates of correctly filled electronic orders by industry. Assume a sample of 700 electronic orders provided the following results.

Order	Industry: Pharmaceutical	Consumer	Computers	Telecommunications
Correct	207	136	151	178
Incorrect	3	4	9	12

a. Test whether order fulfillment is independent of industry. Use $\alpha = 0.05$. What is your conclusion?

b. Which industry has the highest percentage of correctly filled orders?

13 Three suppliers provide the following data on defective parts.

Supplier	Part quality: Good	Minor defect	Major defect
A	90	3	7
B	170	18	7
C	135	6	9

Use $\alpha = 0.05$ and test for independence between supplier and part quality. What does the result of your analysis tell the purchasing department?

14 A sample of parts taken in a machine shop in Karachi 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 $\alpha = 0.05$ and test the hypothesis that part quality is independent of the production shift. What is your conclusion?

15 Visa Card studied how frequently consumers of various age groups use plastic cards (debit and credit cards) when making purchases. Sample data for 300 customers shows the use of plastic cards by four age groups.

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

a. Test for the independence between method of payment and age group. What is the *p*-value? Using $\alpha = 0.05$, 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 and MasterCard?

16 The following cross-tabulation shows industry type and P/E ratio for 100 companies in the consumer products and banking industries.

Industry	P/E ratio: 5–9	10–14	15–19	20–24	25–29	Total
Consumer	4	10	18	10	8	50
Banking	14	14	12	6	4	50
Total	**18**	**24**	**30**	**16**	**12**	**100**

Does there appear to be a relationship between industry type and P/E ratio? Support your conclusion with a statistical test using $\alpha = 0.05$.

12.3 Goodness of fit test: Poisson and normal distributions

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

Poisson distribution

Consider the arrival of customers at the Mediterranean Food Market. Because of recent staffing problems, the Mediterranean's managers asked a local consultancy to assist with the scheduling of checkout assistants. After reviewing the checkout operation, the consultancy will make a recommendation for a scheduling procedure. The procedure, based on a mathematical analysis of waiting times, 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 done 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 five-minute intervals. The following null and alternative hypotheses are appropriate:

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

H_1: The number of customers entering the store during five-minute intervals does not have a Poisson distribution

If a sample of customer arrivals indicates H_0 cannot be rejected, the Mediterranean will proceed with the implementation of the consultancy'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 assistant randomly selects a sample, $n = 128$, of five-minute intervals during weekday mornings over a three-week period. For each five-minute interval in the sample, the store employee records the number of customer arrivals. In summarizing the data, the store assistant determines the number of five-minute intervals having no arrivals, the number of five-minute intervals having one arrival, the number of five-minute intervals having two arrivals, and so on. These data are summarized in Table 12.6, which gives the observed frequencies for the ten categories.

To do the goodness of fit test, we need to consider the expected frequency for each of the ten categories, under the assumption that the Poisson distribution of arrivals is true. The Poisson probability function, first introduced in Chapter 5, is

$$p(X = x) = \frac{\mu^x e^{-\mu}}{x!} \quad \textbf{(12.4)}$$

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

Table 12.6 Observed frequency of the Mediterranean's customer arrivals for a sample of 128 five-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

To use (12.4), we must obtain an estimate of μ, the mean number of customer arrivals during a five-minute time period. The sample mean for the data in Table 12.6 provides this estimate. With no customers arriving in two five-minute time periods, one customer arriving in eight five-minute time periods and so on, the total number of customers who arrived during the sample of 128 five-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 an estimated mean arrival rate of $640/128 = 5$ customers per five-minute period. With this value for the mean of the distribution, an estimate of the Poisson probability function for the Mediterranean Food Market is

$$p(X = x) = \frac{5^x e^{-5}}{x!} \quad \textbf{(12.5)}$$

This probability function can be evaluated for different values 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 five-minute interval is $p(0) = 0.0067$, the probability of one customer arriving during a five-minute interval is $p(1) = 0.0337$ 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 $(0.0067)(128) = 0.86$, the expected number of periods with one arrival is given by $(0.0337)(128) = 4.31$, and so on.

Note that in Table 12.7, four of the categories have an expected frequency less than five. This condition violates the requirements for use of the chi-squared distribution. However, adjacent categories can be combined to satisfy the 'at least five' expected frequency requirement. In particular, we shall combine 0 and 1 into a single category, and then combine 9 with '10 or more' into another single category. 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, $f_i - e_i$. The calculations are shown in Table 12.8. The value of the test statistic is $\chi^2 = 10.96$.

Table 12.7 Expected frequency of Mediterranean's customer arrivals, assuming a Poisson distribution with $\mu = 5$

Number of customers arriving (x)	Poisson probability $p(x)$	Expected number of five-minute time periods with x arrivals, $128p(x)$
0	0.0067	0.86
1	0.0337	4.31
2	0.0842	10.78
3	0.1404	17.97
4	0.1755	22.46
5	0.1755	22.46
6	0.1462	18.71
7	0.1044	13.36
8	0.0653	8.36
9	0.0363	4.65
10 or more	0.0318	4.07
	Total	128.00

In general, the chi-squared 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. Table 12.8 shows $k = 9$ categories. Because the sample data were used to estimate the mean of the Poisson distribution, $p = 1$. Hence, there are $k - p - 1 = 9 - 1 - 1 = 7$ degrees of freedom.

Suppose we test the null hypothesis with a 0.05 level of significance. 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-squared distribution with 7 degrees of freedom. Using Table 3 of Appendix B,

Table 12.8 Observed and expected frequencies for the Mediterranean's customer arrivals after combining categories, and computation of the chi-squared test statistic

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			$\chi^2 = 10.96$

we find that $\chi^2 = 10.96$ provides an area in the upper tail greater than 0.10. So we know that the p-value is greater than 0.10. MINITAB, PASW or EXCEL shows p-value = 0.1403. With p-value > $\alpha = 0.10$, we cannot reject H_0. The assumption of a Poisson probability distribution for weekday morning customer arrivals cannot be rejected. As a result, the Mediterranean'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_1: 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.

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 for the test, and there are $k - 2$ degrees of freedom.

Normal distribution

A goodness of fit test for a normal distribution can also be based on the use of the chi-squared 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.

Consider the job applicant test data for Pharmaco plc, listed in Table 12.9. Pharmaco hires approximately 400 new employees annually for its four plants located throughout Europe. 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 per cent, lower 40 per cent 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.

We first use the data in Table 12.9 to calculate estimates of the mean and standard deviation of the normal distribution that will be considered in the null hypothesis. We use

Table 12.9 Pharmaco employee aptitude test scores for 50 randomly chosen job applicants

71	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	64	65	62	90
69	76	79	77	54	64	74	65	65	61
56	63	80	56	71	79	84	66	61	61

the sample mean and the sample standard deviation as point estimators of the mean and standard deviation of the normal distribution. The calculations follow.

PHARMACO

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

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

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_1: The population of test scores does not have a normal distribution with mean 68.42 and standard deviation 10.41.

Now we look at how to define 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 ten equal-probability intervals (see Figure 12.2). 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.

When the normal probability distribution is assumed, the standard normal distribution tables can be used to determine the category boundaries. First consider the test score cutting off the lowest 10 per cent 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 $x = 68.42 - 1.28\,(10.41) = 55.10$ provides this cut-off value for the lowest 10 per cent of the scores. For the lowest 20 per cent, we find $z = -0.84$, and so $x = 68.42 - 0.84(10.41) = 59.68$. Working through the normal distribution in that way provides the following test score values.

Lower 10%: $68.42 - 1.28(10.41) = 55.10$
Lower 20%: $68.42 - 0.84(10.41) = 59.68$

Lower 30%: $68.42 - 0.52(10.41) = 63.01$
Lower 40%: $68.42 - 0.25(10.41) = 65.82$
Mid-score: $68.42 - 0(10.41) = 68.42$
Upper 40%: $68.42 + 0.25(10.41) = 71.02$
Upper 30%: $68.42 + 0.52(10.41) = 73.83$
Upper 20%: $68.42 + 0.84(10.41) = 77.16$
Upper 10%: $68.42 + 1.28(10.41) = 81.74$

These cutoff or interval boundary points are identified on the graph in Figure 12.1.

We can now return to the sample data of Table 12.9 and determine the observed frequencies for the categories. The results are in Table 12.10. The goodness of fit calculations now proceed exactly as before. Namely, we compare the observed and expected results by computing a χ^2 value. The computations are also shown in Table 12.10. 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-squared 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 we do the test with a 0.10 level of significance. To test this hypothesis, we need to determine the p-value for the test statistic $\chi^2 = 7.2$ by finding the area in the upper tail of a chi-squared 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 0.10. So we know that the p-value is greater than 0.10. EXCEL, PASW or MINITAB shows p-value $= 0.4084$.

With p-value $> \alpha = 0.10$, the hypothesis that the probability distribution for the Pharmaco 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.

Figure 12.1 Normal distribution for the Pharmaco example with ten equal-probability intervals

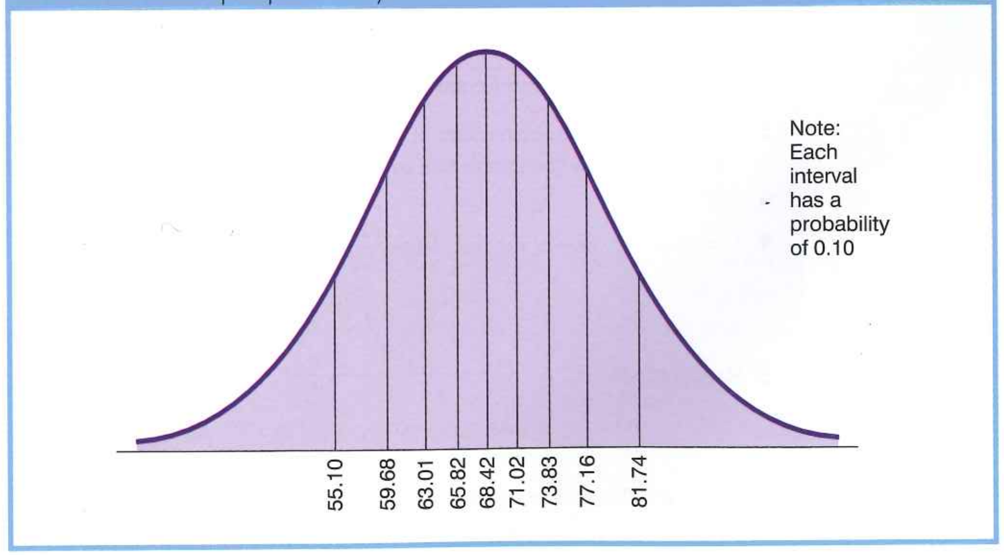


Table 12.10 Observed and expected frequencies for Pharmaco job applicant test scores, and computation of the chi-squared test statistic

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 59.67	5	5	0	0	0.0
59.68 to 63.00	9	5	4	16	3.2
63.01 to 65.81	6	5	1	1	0.2
65.82 to 68.41	2	5	3	9	1.8
68.42 to 71.01	5	5	0	0	0.0
71.02 to 73.82	2	5	3	9	1.8
73.83 to 77.15	5	5	0	0	0.0
77.16 to 81.73	5	5	0	0	0.0
81.74 and over	6	5	1	1	0.2
Total	50	50			$\chi^2 = 7.2$

Normal distribution goodness of fit test: a summary

1 State the null and alternative hypotheses.

H_0: The population has a normal distribution
H_1: 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 $\le \alpha$
Critical value approach: Reject H_0 if $\chi^2 \ge \chi^2_\alpha$

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

Exercises

Methods

17 The following data are believed to have come from a normal distribution. Use a goodness of fit test and $\alpha = 0.05$ 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

18 Data on the number of occurrences per time period and observed frequencies follow. Use $\alpha = 0.05$ and a 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

Applications

19 The number of incoming phone calls to a small call centre in Mumbai, during one-minute intervals, is believed to have a Poisson distribution. Use $\alpha = 0.10$ and the following data to test the assumption that the incoming phone calls follow a Poisson distribution.

Number of incoming phone calls during a one-minute interval	Observed frequency
0	15
1	31
2	20
3	15
4	13
5	4
6	2
Total	100

20 The weekly demand for a particular product in a white-goods store is thought to be normally distributed. Use a goodness of fit test and the following data to test this assumption. Use $\alpha = 0.10$. The sample mean is 24.5 and the sample standard deviation is 3.0.

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

21 A random sample of final examination grades for a college course in Middle-East studies 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 $\alpha = 0.05$ and test to determine whether a normal distribution should be rejected as being representative of the population's distribution of grades.

22 The number of car 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 $\alpha = 0.05$.

Number of accidents	Observed frequency (days)
0	34
1	25
2	11
3	7
4	3

Summary

The purpose of a 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-squared distribution is used to determine whether the differences between observed and expected frequencies are large enough to reject the hypothesized probability distribution.

In this chapter we introduced the goodness of fit test for a multinomial distribution. A test of independence for two variables is an extension of the methodology used in the goodness of fit test for a multinomial population. A contingency table is used to set out the observed and expected frequencies. Then a chi-squared value is computed.

We also illustrated the goodness of fit test for Poisson and normal distributions.

Key terms

Contingency table
Goodness of fit test
Multinomial population

Key formulae

Test statistic for goodness of fit

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

Expected frequencies for contingency tables under the assumption of independence

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

Test statistic for independence

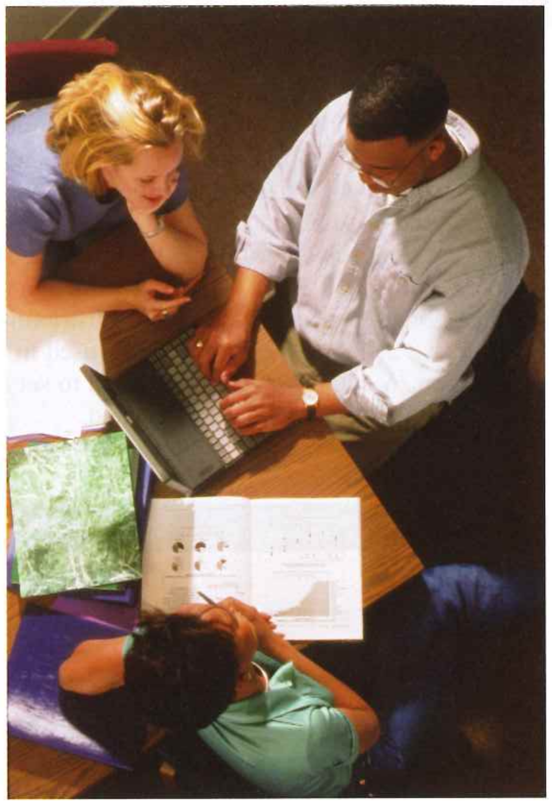
$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \qquad (12.3)$$

Case problem 1 Evaluation of Management School website pages

A group of MSc students at an international university conducted a survey to assess the students' views regarding the web pages of the university's Management School. Among the questions in the survey were items that asked respondents to express agreement or disagreement with the following statements.

1 The Management School web pages are attractive for prospective students.

2 I find it easy to navigate the Management School web pages.



Students using a computer to research business schools. © John Henley/CORBIS.

3 There is up-to-date information about courses on the Management School web pages.

4 If I were to recommend the university to someone else, I would suggest that he/she goes to the Management School web pages.

Responses were originally given on a five-point scale, but in the data file on CD that accompanies this case problem ('Web Pages'), the responses have been recoded as binary variables. For each questionnaire item, those who agreed or agreed strongly with the statement have been grouped into one category (Agree). Those who disagreed, disagreed strongly, were indifferent or opted for a 'Don't know' response, have been grouped into a second category (Don't Agree). The data file also contains particulars of respondent gender and level of study (undergraduate or postgraduate). A screenshot of the first few rows of the data file is shown below.

Managerial report

1 Use descriptive statistics to summarize the data from this study. What are your preliminary conclusions about the independence of the response (Agree or Don't Agree) and gender for each of the four items? What are your preliminary conclusions about the independence of the response (Agree or Don't Agree) and level of study for each of the four items?

2 With regard to each of the four items, test for the independence of the response (Agree or Don't Agree) and gender. Use $\alpha = 0.05$.

3 With regard to each of the four items, test for the independence of the response (Agree or Don't Agree) and level of study. Use $\alpha = 0.05$.

4 Does it appear that views regarding the web pages are consistent for students of both genders and both levels of study? Explain.

Gender	Study level	Attractiveness	Navigation	Up-to-date	Referrals
Female	Undergraduate	Don't Agree	Agree	Agree	Agree
Female	Undergraduate	Agree	Agree	Agree	Agree
Male	Undergraduate	Don't Agree	Don't Agree	Don't Agree	Don't Agree
Male	Undergraduate	Agree	Agree	Agree	Agree
Male	Undergraduate	Agree	Agree	Agree	Agree
Female	Undergraduate	Don't Agree	Don't Agree	Agree	Agree
Male	Undergraduate	Don't Agree	Agree	Agree	Agree
Male	Undergraduate	Agree	Agree	Agree	Agree
Male	Undergraduate	Don't Agree	Agree	Agree	Agree

Case problem 2 Checking for randomness in Lotto draws

National Lottery play slips. © Positive Image/Alamy.

In the main Lotto game of the UK National Lottery, six balls are randomly selected from a set of balls numbered 1, 2, . . . , 49. The file 'Lotto' on the accompanying CD contains details of the numbers drawn in the main Lotto game from early January 1995 up to early March 2008.

A screen shot of the first few rows of the data file is shown below. In addition to showing the six numbers drawn in the game each time, and the order in which they were drawn, the file also gives details of the day on which the draw took place, the machine that was used to do the draw, and the set of balls that was used. In recent years, Lotto draws have taken place on both Wednesday and Saturday each week. A number of similar machines are used for the draws: Sapphire, Topaz, etc, and eight sets of balls are used.

Analyst's report

1 Use an appropriate hypothesis test to assess whether there is any evidence of non-randomness in the first ball drawn. Similarly, test for non-randomness in the second ball drawn, third ball drawn, . . . , sixth ball drawn.

2 Use an appropriate hypothesis test to assess whether there is any evidence of non-randomness overall in the drawing of the 49 numbers (regardless of the order of selection).

3 Use an appropriate hypothesis test to assess whether there is evidence of any dependence between the numbers drawn and the day on which the draw is made.

4 Use an appropriate hypothesis test to assess whether there is evidence of any dependence between the numbers drawn and the machine on which the draw is made.

5 Use an appropriate hypothesis test to assess whether there is evidence of any dependence between the numbers drawn and the set of balls that is used.

No.	Day	DD	MMM	YYYY	N1	N2	N3	N4	N5	N6	Machine	Set
1378	Sat	7	Mar	2009	30	33	32	44	21	6	Sapphire	3
1377	Wed	4	Mar	2009	46	3	32	37	25	21	Topaz	2
1376	Sat	28	Feb	2009	43	6	33	37	25	42	Sapphire	1
1375	Wed	25	Feb	2009	19	6	39	32	25	37	Topaz	6
1374	Sat	21	Feb	2009	18	42	45	46	26	22	Sapphire	6
1373	Wed	18	Feb	2009	24	44	29	5	23	39	Topaz	3
1372	Sat	14	Feb	2009	17	45	4	2	19	38	Topaz	2
1371	Wed	11	Feb	2009	46	38	14	9	47	16	Sapphire	1
1370	Sat	7	Feb	2009	41	26	7	3	24	25	Sapphire	5

Software Section for Chapter 12

Tests of goodness of fit and independence using MINITAB

Goodness of fit test

This MINITAB procedure can be used for goodness of fit tests 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 a MINITAB worksheet. Using the Scott Market Research example presented in Section 12.1, open a MINITAB worksheet, enter the observed frequencies 48, 98 and 54 in column C1 and enter the expected frequencies 60, 100 and 40 in column C2. The MINITAB steps for the goodness of fit test follow.

Step 1 **Calc > Calculator** [Main menu bar]

Step 2 Enter **ChiSquared** in the **Store result in variable** box [**Calculator** panel]
Enter **Sum((C1-C2)**2/C2)** in the **Expression** box
Click **OK**

Step 3 **Calc > Probability Distributions > Chi-square** [Main menu bar]

Step 4 Select **Cumulative probability** [**Chi-square Distribution** panel]
Enter **2** in the **Degrees of freedom** box
Select **Input column** and enter **ChiSquared** in the adjacent box
Click **OK**

The MINITAB output provides the chi-squared statistic $\chi^2 = 7.34$, and the cumulative probability 0.9745, which is the area under the curve to the left of $\chi^2 = 7.34$. The area remaining in the upper tail is the p-value. We have p-value $= 1 - 0.9745 = 0.0255$.

Test of independence

We begin with a new MINITAB worksheet and enter the observed frequency data for the Real Ale Brewery example from Section 12.2 into columns 1, 2 and 3, respectively. We enter the observed frequencies corresponding to a light ale preference (20 and 30) in C1, the observed frequencies corresponding to a lager preference (40 and 30) in C2, and the observed frequencies corresponding to a best bitter preference (20 and 10) in C3. The MINITAB steps for the test of independence are given below. The screen shot in Figure 12.2 shows the MINITAB output, with χ^2 (2 df) $= 6.122$, $p = 0.047$.

Figure 12.2 MINITAB output for the Real Ale Brewery test of independence

Chi-Square Test: Light Ale, Lager, Best Bitter

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

                               Best
        Light Ale   Lager    Bitter   Total
   1           20      40        20      80
            26.67   37.33     16.00
            1.667   0.190     1.000

   2           30      30        10      70
            23.33   32.67     14.00
            1.905   0.218     1.143

Total          50      70        30     150

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

Step 1 **Stat > Tables > Chi-square Test (Two-Way Table in Worksheet)**
[Main menu bar]

Step 2 Enter **C1-C3** in the **Columns containing the table** box
[**Chi-square Test (Table in Worksheet)** panel]
Click **OK**

Tests of goodness of fit and independence using EXCEL

Goodness of fit test

This EXCEL procedure can be used for goodness of fit tests 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 12.3 (refer to the file 'FitTest.XLS' on the accompanying CD). 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 left-hand panel shows the cell formulae.

Test of independence

The EXCEL procedure for the test of independence requires the user to obtain the observed frequencies and enter them in the worksheet. The Real Ale 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 12.4 (refer to the file 'Independence.XLS' on the

Figure 12.3 EXCEL worksheet for the Scott Market Research goodness of fit test

	C	D
1		
2		
3		
4		Calculations
5		=(A5-B5)^2/B5
6		=(A6-B6)^2/B6
7		=(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
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				

accompanying CD). The cell formulae in the left-hand panel 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 Real Ale Brewery test of independence

	B	C	D	E
1				
2				
3				
4				
5		Beer Peference		
6	Light	Regular	Dark	Total
7	20	40	20	=SUM(B7:D7)
8	30	30	10	=SUM(B8:D8)
9	=SUM(B7:B8)	=SUM(C7:C8)	=SUM(D7:D8)	=SUM(E7:E8)
10				
11				
12				
13				
14		Beer Peference		
15	Light	Regular	Dark	Total
16	=E7*B$9/$E$9	=E7*C$9/$E$9	=E7*D$9/$E$9	=SUM(B16:D16)
17	=E8*B$9/$E$9	=E8*C$9/$E$9	=E8*D$9/$E$9	=SUM(B17:D17)
18	=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:D
25				

	A	B	C	D	E
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					

Tests of goodness of fit and independence using PASW

PASW

REALALE

A PASW data file suitable for the Real Ale Brewery example is on the accompanying CD ('RealAle.SAV'). There are three columns in the data file, which between them define the contingency table for this example. The first column contains the row indices for the table (1, 2 for Male, Female), the second column contains the column indices (1, 2, 3 for Light Ale, Lager, Best Bitter) and the third column contains the frequency counts. The data file has six rows corresponding to the six cells of the contingency table. The PASW steps for the test of independence are as follows.

Step 1 **Data > Weight Cases** [Main menu bar]

Step 2 Select **Weight cases by** [**Weight Cases** panel]
Transfer **Frequency count** (third column) to the **Frequency Variable** box
Click **OK**

Step 3 **Analyze > Descriptive Statistics > Crosstabs** [Main menu bar]

Step 4 Transfer **Gender** (first column) to the **Row(s)** box [**Crosstabs** panel]
Transfer **Type of beer** (second column) to the Column(s) box
Click on the **Statistics** button

Step 5 Check **Chi-square** [**Crosstabs: Statistics** panel]
Click **Continue**

Step 6 Click **OK** [**Crosstabs** panel]

The PASW output is shown below in Figure 12.5. The first table is the contingency table. Expected values, together with row and/or column percentages, can be added to this table using the **Cells** button on the **Crosstabs** dialogue panel. The result of the chi-squared test is shown in the first row of the second table, χ^2 (2 df) = 6.122, p = 0.047.

Figure 12.5 PASW output for the Real Ale Brewery test of independence

Gender * Type of beer Crosstabulation

Count

		Type of beer			
		Light Ale	Lager	Best Bitter	Total
Gender	Male	20	40	20	80
	Female	30	30	10	70
Total		50	70	30	150

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.122[a]	2	.047
Likelihood Ratio	6.178	2	.046
Linear-by-Linear Association	5.872	1	.015
N of Valid Cases	150		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.00.

Chapter 13

Analysis of Variance and Experimental Design