Contents 42

Goodness of Fit and Contingency Tables

Learning outcomes

You will learn how to decide whether a set of data fits a particular distribution. You will also learn about a situation in which hypothesis tests are applied to non-numeric data.

Goodness of Fit

42.1

Introduction

If you are applying statistics to practical problems in industry, you may find that much of your work is concerned with making decisions concerning probability distributions. Sometimes it is advantageous to be able to describe the approximate probability distribution followed by a data set obtained experimentally. For example you may be asked to decide whether a data set is approximately normal. In order to make such decisions, you will find that you may use the chi-squared test provided that certain conditions are satisfied. On other occasions you may be given data concerning non-numeric variables in the form of a contingency table. This is one of those occasions when hypothesis tests can be applied to non-numeric variables.

Prerequisites

Before starting this Section you should ...

- understand how to find probabilities for a chi-squared distribution (HELM 40)
- understand the principles of hypothesis testing (HELM 41)

Learning Outcomes

On completion you should be able to ...

- explain the term goodness-of-fit
- perform hypothesis tests based on the chi-squared distribution

1. Goodness-of-fit tests

The aim of a goodness-of-fit test is to determine the underlying nature of the probability distribution describing the population from which a random sample has been drawn. For example, we may wish to determine whether the population from which a sample has been drawn has a normal, binomial or Poisson distribution. While a variety of goodness-of-fit tests exist, the test described here depends on the χ^2-distribution and is usually called the chi-squared test.

We assume that a random sample of size n has been drawn from a population with an unknown probability distribution and that we wish to determine the nature of that distribution.

- Firstly, if the data are continuous we organize the data into k intervals (often equal but not necessarily so) in order that we can write down the observed frequency, say O_i, of the ith interval for $1 \leq i \leq k$.

- Secondly, we form a hypothesis about the nature of the unknown distribution. That is, we assume that it is normal, binomial, Poisson or some other appropriate probability distribution.

- Thirdly, we calculate, on the basis of the hypothesis outlined above, the expected frequency, say E_i, of the ith interval for $1 \leq i \leq k$. The values of E_i are calculated using the formula

  $$E_i = nP_i$$

  where P_i is the probability associated with the interval i.

- Fourthly, we calculate the goodness-of-fit statistic as defined in Key Point 1.

Key Point 1

The goodness-of-fit statistic is given by

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

It can be shown that, if the assumption made about the nature of the population (normal, binomial, Poisson etc.) is true then W follows (approximately) a chi-squared distribution with $k-p-1$ degrees of freedom. Note that p represents the number of parameters needed to describe the probability distribution of the population which we have to estimate from the data. For example the normal distribution has two parameters μ and σ, the binomial distribution has two parameters n and p but we usually only need to estimate p, while the Poisson distribution has one parameter, μ.

- Fifthly, we reject the hypothesis concerning the nature of the underlying probability distribution if the calculated value of W exceeds the value of $\chi^2_{\alpha,k-p-1}$ where α is the area in the tail of the χ^2-distribution, typically 5% or 1%.

Notes

(a) The larger the sample, the more reliable the result since the assertion that W follows (approximately) a chi-squared distribution improves with increasing sample size.

(b) The size of the expected frequencies should be monitored carefully. Various authors recommend that minimum expected frequencies of 3, 4 or 5 are acceptable. It is reasonably safe to accept expected frequencies provided that they are greater than 5 and 10 is certainly acceptable.

(c) Some authors recommend that the k intervals into which the data are organized are chosen so that the frequencies in each interval are roughly equal - remember that equal intervals are not necessary for the test to be performed.

We will now look at two examples of goodness-of-fit tests, the first uses a (discrete) Poisson distribution and the second uses a (continuous) normal distribution. Each worked Example is immediately followed by a Task for you to do.

Example 1

A manufacturer produces high-quality sheet aluminium for use in highly stressed aircraft wings. A random sample of 100 sheets is inspected and the number of faults per sheet recorded. The results are given in the table below.

Number of Faults per Sheet	Frequency of Occurrence
0	50
1	24
2	14
3	8
4	4

Suggest a possible probability distribution from which the sample may have been drawn and perform a chi-squared test to determine the validity of your suggestion.

Solution

The data are already given in 5 classes with observed frequencies as shown. We will assume that the underlying distribution is Poisson and calculate the expected frequencies accordingly using the Poisson formula $\quad \mathsf{P}(X=r)=\dfrac{e^{-\mu}\mu^r}{r!}\quad$ We need the value of the mean.

This is calculated as $\quad \mu=\dfrac{50\times 0+24\times 1+14\times 2+8\times 3+4\times 4}{100}=0.92$

Hence the Poisson probabilities and the corresponding expected frequencies are:

$$p_0=\mathsf{P}(X=0)=\frac{e^{-\mu}\mu^0}{0!}=e^{-0.92}=0.399,\ E_0=39.9$$

$$p_1=\mathsf{P}(X=1)=\frac{e^{-\mu}\mu^1}{1!}=e^{-0.92}\times 0.92=0.367,\ E_1=36.7$$

Solution (contd.)

$$p_2 = \mathsf{P}(X = 2) = \frac{e^{-\mu}\mu^2}{2!} = \frac{e^{-0.92} \times 0.92^2}{2} = 0.169,\ E_2 = 16.9$$

$$p_3 = \mathsf{P}(X = 3) = \frac{e^{-\mu}\mu^3}{3!} = \frac{e^{-0.92} \times 0.92^3}{6} = 0.052,\ E_3 = 5.2$$

$$p_4 = \mathsf{P}(X \geq 4) = 1 - (0.399 + 0.367 + 0.169 + 0.052) = 0.013,\ E_4 = 1.3$$

Note that in calculating p_4 we have ensured that our probabilities sum to unity.

Since the last frequency is very small we will combine the last two and use 4 classes so that $O_3 = 12$ and $E_3 = 6.5$.

The test statistic is

$$W = \sum_{i=0}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(50 - 39.9)^2}{39.9} + \frac{(24 - 36.7)^2}{36.7} + \frac{(14 - 16.9)^2}{16.9} + \frac{(12 - 6.5)^2}{6.5} = 12.103$$

and the number of degrees of freedom is $k - p - 1 = 4 - 1 - 1 = 2$ so that the critical value from Table 1 (at the end of the Workbook) is $\chi^2_{0.05,2} = 5.99$. Clearly $12.103 > 5.99$ and we must reject the null hypothesis that the underlying distribution is Poisson.

Task

A manufacturer produces electronic components for use in computer controlled monitoring systems. A random sample of 100 components is inspected and the number of faults per component recorded. The results are given in the table below.

Number of Faults per Component	Frequency of Occurrence
0	45
1	35
2	16
3	4

Perform a chi-squared test to determine the validity of the assumption that the occurrence of faults in the components is Poisson.

Your solution

Answer

The data are given in 4 classes with observed frequencies as shown. The expected frequencies using the Poisson formula with a mean

$$\mu = \frac{45 \times 0 + 35 \times 1 + 16 \times 2 + 4 \times 3}{100} = 0.79$$

are

$$p_0 = \mathsf{P}(X = 0) = \frac{e^{-\mu}\mu^0}{0!} = e^{-0.79} = 0.454, \;\; E_0 = 45.4$$

$$p_1 = \mathsf{P}(X = 1) = \frac{e^{-\mu}\mu^1}{1!} = e^{-0.79} \times 0.79 = 0.359, \;\; E_1 = 35.9$$

$$p_2 = \mathsf{P}(X = 2) = \frac{e^{-\mu}\mu^2}{2!} = \frac{e^{-0.79} \times 0.79^2}{2} = 0.142, \;\; E_2 = 14.2$$

$$p_3 = \mathsf{P}(X \geq 3) = 1 - (0.454 + 0.359 + 0.142) = 0.045, \;\; E_3 = 4.5$$

The last frequency is small but since it is greater than 3 we will allow its use.

The test statistic is

$$W = \sum_{i=0}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(45 - 45.4)^2}{45.4} + \frac{(35 - 35.9)^2}{35.9} + \frac{(16 - 14.2)^2}{14.2} + \frac{(4 - 4.5)^2}{4.5} = 0.310$$

and the number of degrees of freedom is $k - p - 1 = 4 - 1 - 1 = 2$ so that the critical value from tables is $\chi^2_{0.05,2} = 5.99$. Clearly $0.310 < 5.99$ and we accept the null hypothesis that the underlying distribution is Poisson. Note that the decision to accept the value $E_3 = 4.5$ is fairly marginal and that some personal judgement in such situations as to whether such values should be accepted or combined with another class is unavoidable.

Task

Using the data of the previous Task but combining the expected frequencies of the last two classes, perform a chi-squared test to determine the validity of the assumption that the occurrence of faults in the components is Poisson.

Your solution

Answer
The data are given in 4 classes with observed frequencies as shown. The expected frequencies using the Poisson formula with a mean

$$\mu = \frac{45 \times 0 + 35 \times 1 + 16 \times 2 + 4 \times 3}{100} = 0.79$$

are

$$p_0 = \mathrm{P}(X = 0) = \frac{e^{-\mu}\mu^0}{0!} = e^{-0.79} = 0.454, \; E_0 = 45.4$$

$$p_1 = \mathrm{P}(X = 1) = \frac{e^{-\mu}\mu^1}{1!} = e^{-0.79} \times 0.79 = 0.359, \; E_1 = 35.9$$

$$p_2 = \mathrm{P}(X = 2) = \frac{e^{-\mu}\mu^2}{2!} = \frac{e^{-0.79} \times 0.79^2}{2} = 0.142, \; E_2 = 14.2$$

$$p_3 = \mathrm{P}(X \geq 3) = 1 - (0.454 + 0.359 + 0.142) = 0.045, \; E_3 = 4.5$$

We will combine the expected frequencies of the last two classes and use 3 classes in total with expected frequencies of $E_0 = 45.4, E_1 = 35.9, E_2 = 18.7$.

The test statistic is

$$W = \sum_{i=0}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(45 - 45.4)^2}{45.4} + \frac{(35 - 35.9)^2}{35.9} + \frac{(20 - 18.7)^2}{18.7} = 0.113$$

and the number of degrees of freedom is $k - p - 1 = 4 - 1 - 1 = 2$ so that the critical value from Table 1 (at the end of the Workbook) is $\chi^2_{0.05,2} = 5.99$. Clearly $0.113 < 5.99$ and we accept the null hypothesis that the underlying distribution is Poisson. Note that the decision to combine the last two classes has not, in this case, affected the acceptance of the null hypothesis.

Example 2

A quality control engineer is given the job of checking the voltage output characteristics of a circuit component in a CD player. After checking 100 randomly selected components and plotting a histogram of the results, the engineer concludes that the mean output of the 100 checked components is $\bar{x} = 6.12$ volts, that the standard deviation is $s = 0.1$ volts and that the voltage distribution is probably normal. Choose a suitable test to decide whether the assumption of normality is valid at the 5% level of significance.

Solution

The engineer decides to use a chi-squared test to test the assumption of normality and follow the (common) practice of ensuring that the *expected* frequencies are equal. To do this, the data are put into eight equal classes and the class boundaries calculated as follows.

From the standard normal distribution the Z values corresponding to class boundaries giving a probability of 0.125 (i.e. 1/8) may be read off from tables as $0, 0.32, 0.675, 1.15$ and ∞ for positive values and $0, -0.32, -0.675, -1.15$ and $-\infty$ for negative values. Using

$$Z = \frac{x - \bar{x}}{s} \to x = \bar{x} + Z.s$$

the class boundaries are calculated to be: $6.005, 6.053, 6.088, 6.120, 6.152, 6.188, 6.235$. This gives the eight classes, the observed frequencies found by the engineer (you are given this information here), and the expected frequencies as:

Classes	Observed Frequencies O_i	Expected Frequencies E_i
$x < 6.005$	8	12.5
$6.005 \leq x < 6.053$	11	12.5
$6.053 \leq x < 6.088$	16	12.5
$6.088 \leq x < 6.120$	19	12.5
$6.120 \leq x < 6.152$	18	12.5
$6.152 \leq x < 6.188$	13	12.5
$6.188 \leq x < 6.235$	9	12.5
$6.235 \leq x$	6	12.5

The hypotheses are: H_0: distribution is normal, H_1 : distribution is not normal

The test statistic is

$$\begin{aligned}
W &= \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} \\
&= \frac{(8-12.5)^2}{12.5} + \frac{(11-12.5)^2}{12.5} + \frac{(16-12.5)^2}{12.5} + \frac{(19-12.5)^2}{12.5} + \frac{(18-12.5)^2}{12.5} \\
&+ \frac{(13-12.5)^2}{12.5} + \frac{(9-12.5)^2}{12.5} + \frac{(16-12.5)^2}{12.5} \\
&= 1.62 + 0.18 + 0.98 + 3.38 + 2.42 + 0.02 + 0.98 + 3.38 = 12.96
\end{aligned}$$

and the number of degrees of freedom is $k - p - 1 = 8 - 2 - 1 = 5$ so that the critical value from Table 1 is $\chi^2_{0.05,5} = 11.07$.

Since $11.07 < 12.96$ we have sufficient evidence to reject the null hypothesis and so the engineer should conclude that the distribution of voltages is not normal.

Task

An electrical engineer working for a Health and Safety Executive measures the radiation emitted through the closed doors of 100 used microwave ovens. The measurements, in mw cm^{-2}, are given in the table below.

0.19	0.16	0.14	0.20	0.17	0.21	0.18	0.22	0.26	0.23
0.13	0.17	0.16	0.21	0.18	0.22	0.20	0.23	0.16	0.26
0.19	0.16	0.14	0.20	0.18	0.21	0.19	0.22	0.27	0.24
0.12	0.17	0.15	0.20	0.18	0.22	0.19	0.23	0.29	0.25
0.06	0.16	0.14	0.20	0.17	0.21	0.18	0.22	0.26	0.23
0.13	0.17	0.16	0.20	0.18	0.22	0.19	0.23	0.30	0.25
0.19	0.17	0.14	0.20	0.18	0.21	0.19	0.22	0.27	0.24
0.11	0.17	0.15	0.20	0.18	0.21	0.19	0.23	0.27	0.24
0.13	0.17	0.16	0.21	0.18	0.22	0.19	0.23	0.33	0.25
0.13	0.17	0.16	0.21	0.18	0.22	0.19	0.23	0.36	0.26

The mean radiation of the checked ovens is $\bar{x} = 0.20$ mw cm^{-2}, and the standard deviation is $s = 0.05$ mw cm^{-2}. Verify that the table below giving the eight classes corresponding to the observed and expected frequencies shown is correct.

Classes	Observed Frequencies O_i	Expected Frequencies E_i
$x < 0.143$	11	12.5
$0.143 \le x < 0.166$	10	12.5
$0.166 \le x < 0.184$	19	12.5
$0.184 \le x < 0.200$	10	12.5
$0.200 \le x < 0.216$	16	12.5
$0.216 \le x < 0.234$	17	12.5
$0.234 \le x < 0.258$	6	12.5
$0.258 \le x$	11	12.5

Use a chi-squared test to decide whether the radiation readings obtained from the ovens are normally distributed at the 5% level of significance.

Your solution

Answer

Although the choice of class boundaries is arbitrary, for convenience we choose boundaries to make eight classes with equal probabilities of 0.125.

From the standard normal distribution the Z values corresponding to class boundaries giving a probability of 0.125 may be read off from tables as $0, 0.32, 0.675, 1.15$ and ∞ for positive values and $0, -0.32, -0.675, -1.15$ and $-\infty$ for negative values. Using

$$Z = \frac{x - \bar{x}}{s} \rightarrow x = \bar{x} + Z.s$$

the class boundaries are calculated to be:

$$0.143,\ 0.166,\ 0.184,\ 0.200,\ 0.216,\ 0.234,\ 0.258$$

This gives the eight classes, the observed frequencies found by the engineer and the expected frequencies as given in the table above.

The hypotheses are: H_0: distribution is normal, H_1 : distribution is not normal.

The test statistic is

$$\begin{aligned} W &= \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(11-12.5)^2}{12.5} + \frac{(10-12.5)^2}{12.5} + \frac{(19-12.5)^2}{12.5} + \frac{(10-12.5)^2}{12.5} + \frac{(16-12.5)^2}{12.5} \\ &+ \frac{(17-12.5)^2}{12.5} + \frac{(6-12.5)^2}{12.5} + \frac{(11-12.5)^2}{12.5} \\ &= 0.18 + 0.5 + 3.38 + 0.5 + 0.98 + 1.62 + 3.38 + 0.18 = 10.72 \end{aligned}$$

and the number of degrees of freedom is $k - p - 1 = 8 - 2 - 1 = 5$ so that the critical value from Table 1 is $\chi^2_{0.05,5} = 11.07$.

Since $10.72 < 11.07$ we do not have sufficient evidence to reject the null hypothesis and so the engineer should conclude that the distribution of microwave radiation readings taken from the ovens is normal.

Exercises

1. A factory produces portable CD players. Every week a sample of ten players is selected and subjected to 100 hours of continuous use. At the end of this time the players are tested and the number not reaching a specified standard is recorded. The numbers recorded in 100 consecutive weeks are given below. Test the hypothesis that the data come from a binomial distribution. Use the 5% level of significance.

Number failing standard	0	1	2	3	4	5
Number of weeks	34	24	19	14	9	0

2. A highway engineer records the numbers of vehicles passing a point in a road in 120 consecutive one-minute intervals, as follows. Test the hypothesis that the data come from a Poisson distribution. Use the 5% level of significance.

Number of vehicles	0	1	2	3	4	5	6	7	8	9	10	11
Number of intervals	0	5	10	20	30	20	15	7	6	4	2	1

3. In a test of a device to generate electricity from wave power at sea, 60 observations are made of the root mean square bending moment Y of a component (in newton metres). The data are summarised as follows. The sample mean is 5.08 and the sample variance is 3.29. Test the hypothesis that Y has a normal distribution. Use the 5% level of significance.

Class	Frequency	Class	Frequency
$Y \leq 2$	1	$6 < Y \leq 7$	5
$2 < Y \leq 3$	4	$7 < Y \leq 8$	4
$3 < Y \leq 4$	12	$8 < Y \leq 9$	2
$4 < Y \leq 5$	18	$9 < Y \leq 10$	2
$5 < Y \leq 6$	11	$10 < Y$	1

4. Eighty aircraft components are tested until they fail. The failure times T in hours are summarised as follows. The sample mean is 6434. Test the hypothesis that the distribution of T is exponential. Use the 5% level of significance.

Class	Frequency	Class	Frequency
$0 < T \leq 2000$	11	$10000 < T \leq 12000$	3
$2000 < T \leq 4000$	21	$12000 < T \leq 14000$	5
$4000 < T \leq 6000$	19	$14000 < T \leq 16000$	1
$6000 < T \leq 8000$	9	$16000 < T \leq 18000$	3
$8000 < T \leq 10000$	4	$18000 < T$	4

Answers

1. Total number of failures: $0 \times 34 + 1 \times 24 + \cdots + 4 \times 9 = 140.$

Mean number of failures per week: $140/100 = 1.4.$

Estimate of $p : 1.4/5 = 0.28.$

Use binomial$(5, 0.28)$ distribution.

$$\mathsf{P}(X = j) = \begin{pmatrix} 5 \\ j \end{pmatrix} 0.28^j 0.72^{5-j}$$

No. failing	Probability	Frequency	
		Expected	Observed
0	0.1935	19.35	34
1	0.3762	37.62	24
2	0.2926	29.26	19
3	0.1138	11.38	14
4	0.0221	2.21	9
5	0.0017	0.17	0

Some expected frequencies are too small so we combine neighbouring classes.

No. failing	Probability	Frequency	
		Expected	Observed
0	0.1935	19.35	34
1	0.3762	37.62	24
2	0.2926	29.26	19
3,4,5	0.1376	13.76	23

Test statistic:

$$W = \frac{(34 - 19.35)^2}{19.35} + \cdots + \frac{(23 - 13.76)^2}{13.76} = 25.825.$$

Degrees of freedom: $4 - 1 - 1 = 2$ (4 classes, 1 estimated parameter).

Critical value: $\chi^2_2(5\%) = 5.991.$

The test statistic is significant at the 5% level. We reject the null hypothesis. We conclude that the data do not come from a binomial distribution. There seems to be an excess of large and small counts.

Answers

2. Total number of vehicles: $0 \times 0 + 1 \times 5 + 2 \times 10 + \cdots + 11 \times 1 = 559$.

Mean number of vehicles per minute: $559/120 = 4.658$.

Use Poisson(4.658) distribution.

$$\mathsf{P}(X = j) = \frac{e^{-4.658} 4.658^j}{j!}.$$

No. vehicles	Probability	Frequency	
		Expected	Observed
0	0.00949	1.14	0
1	0.04418	5.30	5
2	0.10290	12.35	10
3	0.15977	19.17	20
4	0.18606	22.33	30
5	0.17333	20.80	20
6	0.13456	16.15	15
7	0.08954	10.74	7
8	0.05214	6.26	6
9	0.02698	3.24	4
10	0.01257	1.51	2
≥ 11	0.00848	1.02	1

Some expected frequencies are too small so we combine neighbouring classes.

No. vehicles	Probability	Frequency	
		Expected	Observed
0,1	0.05367	6.44	5
2	0.10290	12.35	10
3	0.15977	19.17	20
4	0.18606	22.33	30
5	0.17333	20.80	20
6	0.13456	16.15	15
7	0.08954	10.74	7
8	0.05214	6.26	6
≥ 9	0.04803	5.76	7

Test statistic:

$$W = \frac{(5 - 6.44)^2}{6.44} + \cdots + \frac{(7 - 5.76)^2}{5.76} = 5.132.$$

Degrees of freedom: $9 - 1 - 1 = 7$ (9 classes, 1 estimated parameter).

Critical value: $\chi^2_7(5\%) = 14.07$.

The test statistic is not significant at the 5% level. We do not reject the null hypothesis. There is insufficient evidence to conclude that the data do not come from a Poisson distribution.

Answers

3. Using a $N(5.08, 3.29)$ distribution we can calculate the probabilities for the various class intervals. For example,

$$\begin{aligned}
\mathsf{P}(5 < Y \leq 6) &= \Phi\left(\frac{6-5.08}{\sqrt{3.29}}\right) - \Phi\left(\frac{5-5.08}{\sqrt{3.29}}\right) \\
&= \Phi(0.5072) - \Phi(-0.0441) \\
&= \Phi(0.5072) + \Phi(0.0441) - 1 \\
&= 0.694 + 0.518 - 1 = 0.694 - 0.482 = 0.212.
\end{aligned}$$

Bending moment Y	Probability	Frequency	
		Expected	Observed
$Y \leq 2$	0.045	2.70	1
$2 < Y \leq 3$	0.081	4.86	4
$3 < Y \leq 4$	0.150	9.00	12
$4 < Y \leq 5$	0.206	12.36	18
$5 < Y \leq 6$	0.212	12.72	11
$6 < Y \leq 7$	0.161	9.66	5
$7 < Y \leq 8$	0.091	5.46	4
$8 < Y \leq 9$	0.039	2.34	2
$9 < Y \leq 10$	0.012	0.72	2
$10 < Y$	0.003	0.18	1

Some expected frequencies are too small so we combine neighbouring classes.

Bending moment Y	Probability	Frequency	
		Expected	Observed
$Y \leq 3$	0.126	7.56	5
$3 < Y \leq 4$	0.150	9.00	12
$4 < Y \leq 5$	0.206	12.36	18
$5 < Y \leq 6$	0.212	12.72	11
$6 < Y \leq 7$	0.161	9.66	5
$7 < Y \leq 8$	0.091	5.46	4
$8 < Y$	0.054	3.24	5

Test statistic:

$$W = \frac{(5-7.56)^2}{7.56} + \cdots + \frac{(5-3.24)^2}{3.24} = 8.267.$$

Degrees of freedom: $7 - 2 - 1 = 4$ (7 classes, 2 estimated parameters).

Critical value: $\chi^2_4(5\%) = 9.488$.

The test statistic is not significant at the 5% level. We do not reject the null hypothesis. There is insufficient evidence to conclude that the data do not come from a normal distribution.

Answers

4. The sample mean is 6434. We estimate λ using $1/6434 = 1.554 \times 10^{-4}$. We use an exponential$(1.554 \times 10^{-4})$ distribution. For example

$$\begin{aligned}
\mathrm{P}(2000 < T \leq 4000) &= \{1 - \exp(-4000 \times 1.544 \times 10^{-4})\} - \{1 - \exp(-2000 \times 1.544 \times 10^{-4})\} \\
&= \exp(-2000/6434) - \exp(-4000/6434) \\
&= \exp(-0.3108) - \exp(-0.6217) \\
&= 0.733 - 0.537 = 0.196
\end{aligned}$$

Failure time T	Probability	Frequency	
		Expected	Observed
$0 < T \leq 2000$	0.267	21.36	11
$2000 < T \leq 4000$	0.196	15.68	21
$4000 < T \leq 6000$	0.143	11.44	19
$6000 < T \leq 8000$	0.106	8.48	9
$8000 < T \leq 10000$	0.077	6.16	4
$10000 < T \leq 12000$	0.056	4.48	3
$12000 < T \leq 14000$	0.041	3.28	5
$14000 < T \leq 16000$	0.031	2.48	1
$16000 < T \leq 18000$	0.022	1.76	3
$18000 < T$	0.061	4.88	4

Some expected frequencies are too small so we combine neighbouring classes.

Failure time T	Probability	Frequency	
		Expected	Observed
$0 < T \leq 2000$	0.267	21.36	11
$2000 < T \leq 4000$	0.196	15.68	21
$4000 < T \leq 6000$	0.143	11.44	19
$6000 < T \leq 8000$	0.106	8.48	9
$8000 < T \leq 10000$	0.077	6.16	4
$10000 < T \leq 12000$	0.056	4.48	3
$12000 < T \leq 14000$	0.041	3.28	5
$14000 < T \leq 18000$	0.053	4.24	4
$18000 < T$	0.061	4.88	4

Test statistic:

$$W = \frac{(11 - 21.36)^2}{21.36} + \cdots + \frac{(4 - 4.88)^2}{4.88} = 14.18.$$

Degrees of freedom: $9 - 1 - 1 = 7$ (9 classes, 1 estimated parameter).

Critical value: $\chi^2_7(5\%) = 14.07$.

The test statistic is significant at the 5% level. We reject the null hypothesis. We conclude that the data do not come from an exponential distribution. The observed frequency in the first class seems to be too small.

Contingency Tables 42.2

Introduction

The practical application of statistics to engineering problems met in industry often concerns making decisions concerning probability distributions. For example you may be asked to decide whether a data set is approximately normal since much of the statistics you may apply makes this assumption. On occasions you may have to make such decisions given data concerning non-numeric variables in the form of a contingency table. Contingency tables are described in detail in this Workbook. This is one of the relatively rare occasions when hypothesis tests can be applied to non-numeric variables.

Prerequisites

Before starting this Section you should ...

- understand thoroughly what is meant by the term degrees of freedom
- have knowledge of the chi-squared distribution described in HELM 40

Learning Outcomes

On completion you should be able to ...

- explain the term contingency table
- perform hypothesis tests involving data given as a contingency table

1. Contingency tables

On occasions, it is possible that the members of a sample taken from a population can be classified by two different methods. Examples of this are:

(a) articles produced by three machines running during two shifts on a production line;

(b) the failure of electronic components and the position in which they are mounted in a machine;

(c) the failure under compression testing of steel-alloy components and the rate of cooling applied during their production.

We can represent the information obtained by observation in such situations in a *contingency table*. By using the observed data to estimate expected data on the assumption that the classification methods are independent, we can use the chi-squared test to investigate the statistical independence (or otherwise) of the classification methods.

Consider the following contingency table with r rows and c columns. Such a table is referred to as an $r \times c$ contingency table.

	1	2	3	...	c	Row Totals
1	O_{11}	O_{12}	O_{13}	...	O_{1c}	R_1
2	O_{21}	O_{22}	O_{23}	...	O_{2c}	R_2
3	O_{31}	O_{32}	O_{33}	...	O_{3c}	R_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	O_{r1}	O_{r2}	O_{r3}	...	O_{rc}	R_r
Column Totals	C_1	C_2	C_3	...	C_c	N

Note that N is the total of the row totals and is the same as the total of the column totals, that is, N is the number of members of the sample taken from a population.

On the basis of the observed data we can estimate the expected frequency, say E_{ij} corresponding to the observed frequency O_{ij}. This is done as follows.

The probability that a randomly chosen element of the sample appears in row class i and column class j is given by p_{ij} where

$$p_{ij} = \frac{R_i}{N} \times \frac{C_j}{N}$$

Hence the required expected frequency is given by E_{ij} which is defined in Key Point 2.

Key Point 2

Expected Frequencies in Contingency Tables

$$E_{ij} = N \times p_{ij} = N \times \frac{R_i}{N} \times \frac{C_j}{N} = \frac{R_i \times C_j}{N}$$

Using this formula repeatedly, we can calculate the expected frequencies corresponding to the observed frequencies and hence calculate a test statistic W where

$$W = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

This formula tells you to calculate $\dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$ for every cell in the contingency table and sum them.

It can be shown that, provided N is large, and none of the expected frequencies are too small, say less than 3, then the quantity

$$W = \sum_{i=1}^{c} \sum_{j=1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

follows approximately a chi-squared distribution with $(r-1) \times (c-1)$ degrees of freedom when the null hypothesis is true. This number of degrees of freedom arises since each row has $r-1$ independent entries and each column has $c-1$ independent entries.

Notes

The above statements are correct provided that we can calculate the expected frequencies without knowing the population parameters. If we have to estimate the population parameters, the number of degrees of freedom becomes $(r-1)\times(c-1)-m$ where m is the number of population parameters estimated. In the examples given here we shall not need to estimate the population parameters.

To complete the test procedure we note that the null hypothesis assumes class independence. For example, referring back to Example 2 given at the start of this Section, the null hypothesis would assume that the failure of electronic components and the position in which they are mounted in a machine are independent.

Should the test statistic exceed the critical value of χ^2 read from Table 1 at (say) the 5% level of significance, we would reject the null hypothesis and conclude that a relationship of some kind exists between the classes.

It is worth noting that in some cases (such as the following Example 3) one classification is chosen deliberately but the other is random while in other cases, both classifications are random. The same test applies in both cases.

Example 3

In an experiment to determine the most advantageous position in a machine to mount an electronic component which may be prone to failure due to excessive heat build-up, 300 machines are tested with 100 randomly chosen examples of the component in each of 3 positions. The results obtained were as follows.

Position	1	2	3	Row Totals
Failure	40	30	50	120
Non-failure	60	70	50	180
Column Totals	100	100	100	300

Use a χ^2-test at the 5% level of significance to determine whether component failure is related to mounting position.

Solution

The hypotheses are:

H_0: component failure is independent of position,

H_1: component failure is not independent of position

The expected frequencies are calculated are follows:

$$E_{11} = \frac{120 \times 100}{300} = 40, \quad E_{12} = \frac{120 \times 100}{300} = 40, \quad E_{13} = \frac{120 \times 100}{300} = 40$$

$$E_{21} = \frac{180 \times 100}{300} = 60, \quad E_{22} = \frac{180 \times 100}{300} = 60, \quad E_{23} = \frac{180 \times 100}{300} = 60$$

The test statistic is

$$\begin{aligned} W &= \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{(40-40)^2}{40} + \frac{(30-40)^2}{40} + \frac{(50-40)^2}{40} + \frac{(60-60)^2}{60} + \frac{(70-60)^2}{60} + \frac{(50-60)^2}{60} \\ &= 0 + 2.5 + 2.5 + 0 + 1.67 + 1.67 = 8.34 \end{aligned}$$

and the number of degrees of freedom is $(r-1) \times (c-1) = (2-1) \times (3-1) = 2$ so that the critical value from tables is $\chi^2_{0.05,2} = 5.99$.

Since $5.99 < 8.34$ we reject the null hypothesis and so we should conclude that there is a relationship between component failure and mounting position. Position 2 seems to be the most favourable and position 3 the least.

Washing machines are made on three production lines in a factory. A record is kept of faults reported, during the guarantee period, in machines produced by each of the three lines. The faults are classified into three types A, B and C. The results are given in the table below.

	Fault type			
Production line	A	B	C	Row Totals
1	40	28	34	102
2	27	39	32	98
3	45	26	29	100
Column Totals	112	93	95	300

Use a χ^2-test at the 5% level of significance to determine whether fault type is related to the production line on which the machine was produced.

Your solution

Answer

The hypotheses are:

H_0 : fault type is independent of production line,

H_1 : fault type is not independent of production line

The expected frequencies are calculated are follows:

$$E_{11} = \frac{102 \times 112}{300} = 38.08, \quad E_{12} = \frac{102 \times 93}{300} = 31.62, \quad E_{13} = \frac{102 \times 95}{300} = 32.30$$

$$E_{21} = \frac{98 \times 112}{300} = 36.59, \quad E_{22} = \frac{98 \times 93}{300} = 30.38, \quad E_{23} = \frac{98 \times 95}{300} = 31.03$$

$$E_{31} = \frac{100 \times 112}{300} = 37.30, \quad E_{32} = \frac{100 \times 93}{300} = 31.00, \quad E_{33} = \frac{100 \times 95}{300} = 31.70$$

The test statistic is

$$\begin{aligned} W &= \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{(40-38.08)^2}{38.08} + \frac{(28-31.62)^2}{31.62} + \frac{(34-32.30)^2}{32.30} + \frac{(27-36.59)^2}{36.59} + \frac{(39-30.38)^2}{30.38} \\ &+ \frac{(32-31.03)^2}{31.03} + \frac{(45-37.30)^2}{37.30} + \frac{(26-31.00)^2}{31} + \frac{(29-31.70)^2}{31.7} \\ &= 0.097 + 0.414 + 0.089 + 2.512 + 2.446 + 0.030 + 1.590 + 0.806 + 0.230 = 8.214 \end{aligned}$$

and the number of degrees of freedom is $(r-1) \times (c-1) = (3-1) \times (3-1) = 4$ so that the critical value from tables is $\chi^2_{0.05,4} = 9.49$.

Since $8.214 < 9.49$ we do not have sufficient evidence to reject the null hypothesis and so we should conclude that there is no evidence that the distribution of fault types differs between production lines.

Exercises

1. A new compound for the drive belt of domestic vacuum cleaners is tested. Twenty cleaners are fitted with belts made from the new material and twenty are fitted with standard belts. The cleaners are run for a fixed period after which the belts are examined for signs of wear. The numbers showing significant wear are counted. The data are as follows.

	Wear	No wear
Standard	12	8
New compound	6	14

   Test the hypothesis that there is no difference between the standard belts and those made with the new compound in terms of the probability of showing wear. Use the 5% level of significance.

2. Electronic devices are made on three production lines. Records are kept of faults found on devices made on each line. Faults are classified as "electronics", "power supply" or "mechanical". The data are as follows.

	Production Line		
	1	2	3
Electronic	13	33	15
Power supply	7	4	11
Mechanical	18	10	14

   Test the hypothesis that there is no association between production line and type of fault. Use the 5% level of significance.

Answers

1. Observed frequencies:

	Wear	No wear	Total
Standard	12	8	20
New compound	6	14	20
Total	18	22	40

Expected frequencies: $20 \times 18/40 = 9$, $20 \times 22/40 = 11$.

	Wear	No wear	Total
Standard	9	11	20
New compound	9	11	20
Total	18	22	40

Test statistic
$$W = \sum \frac{(O-E)^2}{E} = \frac{(12-9)^2}{9} + \frac{(8-11)^2}{11} + \frac{(6-9)^2}{9} + \frac{(14-11)^2}{11}$$
$$= 1 + 0.82 + 1 + 0.82 = 3.636$$

Degrees of freedom: $(2-1) \times (2-1) = 1$.

Critical value: $\chi^2_1(5\%) = 3.841$.

The result is not significant at the 5% level. There is insufficient evidence to conclude that there is a difference between the wear rates.

Answers

2. Observed frequencies:

	Production Line			
	1	2	3	Total
Electronic	13	33	15	61
Power supply	7	4	11	22
Mechanical	18	10	14	42
Total	38	47	40	125

Expected frequencies, e.g. $61 \times 38/125 = 18.544$.

	Production Line			
	1	2	3	Total
Electronic	18.544	22.936	19.520	61
Power supply	6.688	8.272	7.040	22
Mechanical	12.768	15.792	13.440	42
Total	38.000	47.000	40.000	125

Test statistics

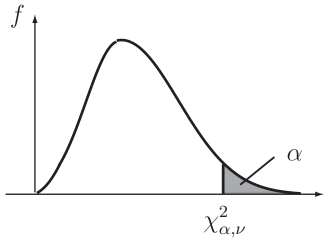
$$W = \sum \frac{(O-E)^2}{E} = \frac{(13-18.544)^2}{18.544} + \cdots + \frac{(14-13.440)^2}{13.440} = 15.860.$$

Degrees of freedom: $(3-1) \times (3-1) = 4$.

Critical value: $\chi^2_4(5\%) = 9.488$.

The test statistic is significant at the 5% level. We reject the null hypothesis and conclude that there is an association between fault type and production line. In particular there seems to be an excess of electronic faults on Line 2.

Table 1: Percentage Points $\chi^2_{\alpha,\nu}$ of the χ^2 distribution



α	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
v											
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.01	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.28
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	31.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17