Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 1

Question:

An octagonal die is thrown 500 times and the results are noted. It is assumed that the die is unbiased. A test is to be done to see whether the observed results differ from the expected ones. Write down a null hypothesis and an alternative hypothesis that can be used.

Solution:

H_0 : There is no difference between the observed and expected distributions.

H_1: There is a difference between the observed and expected distributions.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 2

Question:

For 5 degrees of freedom find the critical value of χ^2 which is exceeded with a probability of 5%.

Solution:

11.070

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 3

Question:

Find the values of the following from the table on page 137.

a $\chi_5^2(5\%)$

b $\chi_8^2(1\%)$

c $\chi_{10}^2(10\%)$

Solution:

i 11.070
ii 20.090
iii 15.987

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 4

Question:

With $\nu = 10$ find the value of χ^2 that is exceeded with 0.05 probability.

Solution:

18.307

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 5

Question:

With $\nu = 8$ find the value of χ^2 that is exceeded with 0.10 probability.

Solution:

13.362

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 6

Question:

The random variable Y has a χ^2 distribution with 8 degrees of freedom. Find y such that $P(Y < y) = 0.99$.

Solution:

20.090

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 7

Question:

The random variable X has a χ^2 distribution with 5 degrees of freedom. Find x such that $P(X < x) = 0.95$.

Solution:

11.070

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise A, Question 8

Question:

The random variable Y has a χ^2 distribution with 12 degrees of freedom.
Find:

a y such that $P(Y < y) = 0.05$,

b y such that $P(Y < y) = 0.95$.

Solution:

a 5.226

b 21.026

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 1

Question:

The following table shows observed values for what is thought to be a discrete uniform distribution.

x	1	2	3	4	5	6	7	8
Frequency of x	12	24	18	20	25	17	21	23

a Calculate the expected frequencies and, using a 5% significance level, conduct a goodness of fit test.
b State your conclusions.

Solution:

a Total frequency $= 160$

Expected frequency for each value of $x = \frac{160}{8} = 20$

H_0: A discrete uniform distribution is a suitable model.

H_1: A discrete uniform distribution is *not* a suitable model.

Test statistic $(\chi^2) = \frac{(12-20)^2}{20} + \frac{(24-20)^2}{20} + \cdots + \frac{(23-20)^2}{20}$

t.s. $(\chi^2) = 6.4$

Degrees of freedom $= 8 - 1 = 7$

Critical value (c.v.) $= \chi^2_7(5\%) = 14.067$

$\chi^2 <$ c.v. so accept H_0

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

b There is no reason to doubt that the data could be from a discrete uniform distribution. A discrete uniform distribution is a good model.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 2

Question:

The following tables show observed values (O) and expected values (E) for a goodness of fit test of a binomial distribution model. The probability used in calculating the expected values has not been found from the observed values.

O	17	28	32	15	5	3
E	19.69	34.74	27.59	12.98	4.01	0.99

a Conduct the test using a 5% significance level and state your conclusions.
b Suggest how the model might be improved.

Solution:

a H_0 : A binomial distribution is a good model.
H_1: A binomial distribution is *not* a good model.

Test statistic $\chi^2 = \frac{(17-19.69)^2}{19.69} + \frac{(28-34.74)^2}{34.74} + \cdots + \frac{(8-5.00)^2}{5.00}$

t.s. $\chi^2 = 4.49$

The last 2 groups must be pooled to get E as 5 or more

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

There are 5 classes after pooling.

Degrees of freedom $= 5 - 1 = 4$

c.v. $= \chi_4^2(5\%) = 9.488$

$\chi^2 <$ c.v. so accept H_0

We only have 1 restriction, the total frequencies agree.

Conclude that there is no reason to doubt the data could be modelled by a binomial distribution.

b The mean of the original data could have been calculated and then used to find p using mean $= np$.

This would increase the number of restrictions by 1 and hence reduce the degrees of freedom by 1.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 3

Question:

The following table shows observed values for a distribution which it is thought may be modelled by a Poisson distribution.

x	0	1	2	3	4	5	> 5
Frequency of x	12	23	24	24	12	5	0

A possible model is thought to be Po(2). From tables, the expected values are found to be as shown in the following table.

X	0	1	2	3	4	5	> 5
Expected frequency of x	13.53	27.07	27.07	18.04	9.02	3.61	1.66

a Conduct a goodness of fit test at the 5% significance level.
b It is suggested that the model could be improved by estimating the value of λ from the observed results. What effect would this have on the number of constraints placed upon the degrees of freedom?

Solution:

a H_0: The data can be modelled by Po(2).
H_1: The data cannot be modelled by Po(2).

x	0	1	2	3	4	5 or more
O	12	23	24	24	12	5
E	13.53	27.07	27.07	18.04	9.02	5.27

Pool the last 2 classes to get all the E values to be 5 or more.

Test statistic $\left(\chi^2\right) = \frac{(12-13.53)^2}{13.53} + \frac{(23-27.07)^2}{27.07} + \cdots + \frac{(5-5.27)^2}{5.27}$

t.s. $\left(\chi^2\right) = 4.10$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Classes after pooling.

Degrees of freedom $= 6 - 1 = 5$

One restriction, the total frequencies agree.

critical value $= \chi_5^2(5\%) = 11.070$

$\chi^2 <$ c.v. so accept H_0.

Conclude that there is no reason to doubt the data can be modelled by Po(2).
There is no reason to reject H_0.

b If λ is calculated then this becomes another restriction.
Degrees of freedom $= 6 - 2 = 4$

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 4

Question:

A mail order firm receives packets every day through the mail.
They think that their deliveries are uniformly distributed throughout the week. Tes this assertion, given that their deliveries over a 4-week period were as follows. Use a 0.05 significance level.

Day	Mon	Tues	Wed	Thurs	Fri	Sat
Frequency	15	23	19	20	14	11

Solution:

H_0: The data can be modelled by a discrete uniform distribution.
H_1: The data cannot be modelled by a discrete uniform distribution.

$$\text{Expected frequency} = \frac{102}{6} = 17$$

$$\text{Test statistic } (\chi^2) = \frac{(15-17)^2}{17} + \frac{(23-17)^2}{17} + \cdots + \frac{(11-17)^2}{17}$$

$$\text{t.s. } (\chi^2) = 5.765$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Degrees of freedom $= 6 - 1 = 5$

One restriction, total frequencies agree.

critical value $= \chi_5^2\,(5\%) = 11.070$

$\chi^2 <$ c.v. so accept H_0.

Conclude that there is no reason to doubt the data can be modelled by a discrete uniform distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 5

Question:

Over a period of 50 weeks the number of road accidents reported to a police station were as shown.

Number of accidents	0	1	2	3	4
Number of weeks	15	13	9	13	0

a Find the mean number of accidents per week.

b Using this mean and a 0.10 significance level, test the assertion that these data are from a population with a Poisson distribution.

Solution:

a $\bar{X} = 1.4$

This is one restriction in part **b**.

b H_0: The data can be modelled by Po(1.4).
H_1: The data cannot be modelled by Po(1.4).

A Poisson distribution has no upper limit

x	0	1	2	3	4	5 or more
O	15	13	9	13	0	0
E	12.330	17.262	12.083	5.639	1.974	0.712

50 – [total of other E's in list]

Total frequency = 50

This is one restriction in part **b**.

$$E(0) = \left(e^{-1.4} \times \frac{1.4^0}{0!}\right) \times 50$$

The last 3 classes must be pooled so *all* the E's are 5 or more.

x	0	1	2	3 or more
O	15	13	9	13
E	12.330	17.262	12.083	8.325

$$\text{Test statistic } (\chi^2) = \frac{(15-12.330)^2}{12.330} + \frac{(13-17.262)^2}{17.262} + \frac{(9-12.083)^2}{12.083} + \frac{(13-8.325)^2}{8.325}$$

$$\text{t.s. } (\chi^2) = 5.04$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Number of classes after pooling.

Degrees of freedom $= 4 - 2$
$= 2$

Number of restrictions.

critical value $= \chi^2_2(10\%) = 4.605$

$\chi^2 >$ c.v. so reject H_0.

Conclude that there is evidence to suggest that the data cannot be modelled by Po(1.4).

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 6

Question:

A marksman fires 6 shots at a target and records the number r of bull's-eyes hit. After a series of 100 such trials he analyses his scores, the frequencies being as follows.

r	0	1	2	3	4	5	6
Frequency	0	26	36	20	10	6	2

a Estimate the probability of hitting a bull's-eye.

b Use a test at the 0.05 significance level to see if these results are consistent with the assumption of a binomial distribution.

Solution:

In part **b** this is 1 restriction since a parameter has been estimated by calculation in order to find P.

a $\bar{r} = 2.4, n = 6$

mean $= np$ $\quad 2.4 = 6 \times p$

$\therefore p = 0.4$

b Total frequency $= 100$

This is 1 restriction.

H_0: The data can be modelled by B(6, 0.4).

H_1: The data cannot be modelled by B(6, 0.4).

$$E(X=0) = 100 \times P(X=0) = 100 \times \binom{6}{0} \times 0.4^0 \times 0.6^6 = 4.666$$

$$E(X=1) = 100 \times P(X=1) = 100 \times \binom{6}{1} \times 0.4^1 \times 0.6^5 = 18.662$$

Pool to get $E \geq 5$.

$$E(X=2) = 100 \times P(X=2) = 100 \times \binom{6}{2} \times 0.4^2 \times 0.6^4 = 31.104$$

$$E(X=3) = 100 \times P(X=3) = 100 \times \binom{6}{3} \times 0.4^3 \times 0.6^3 = 27.648$$

$$E(X=4) = 100 \times P(X=4) = 100 \times \binom{6}{4} \times 0.4^4 \times 0.6^2 = 13.824$$

$$E(X=5) = 100 \times P(X=5) = 100 \times \binom{6}{5} \times 0.4^5 \times 0.6^1 = 3.6864$$

$$E(X=6) = 100 \times P(X=6) = 100 \times \binom{6}{6} \times 0.4^6 \times 0.6^0 = 0.4096$$

Pool to get $E \geq 5$.

x	1 or less	2	3	4 or more
O	26	36	20	18
E	23.328	31.104	27.648	17.92

$$\text{Test statistic } (\chi^2) = \frac{(26-23.328)^2}{23.328} + \frac{(36-31.104)^2}{31.104} + \frac{(20-27.648)^2}{27.648} + \frac{(18-17.92)^2}{17.92}$$

t.s. $(\chi^2) = 3.19$

Degrees of freedom $= 4 - 2$

$= 2$

Number of classes after pooling.

Number of restrictions.

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_2^2(5\%) = 5.991$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data can be modelled by B(6, 0.4).

There is no reason to reject H_0.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 7

Question:

The table below shows the number of employees in thousands at five factories and the number of accidents in 3 years.

Factory	A	B	C	D	E
Employees (thousands)	4	3	5	1	2
Accidents	22	14	25	8	12

Using a 0.05 significance level, test the hypothesis that the number of accidents per 1000 employees is constant at each factory.

Solution:

H_0: The rate of accidents is constant at the factories.
H_1: The rate of accidents isn't constant at the factories.

Total number of accidents $= 81$
Total number of employees $= 15$(thousand)
$\therefore$ mean rate of accidents $= \frac{81}{15} = 5.4$ (per thousand)

This is 1 restriction, calculation of the mean rate.

Factory	A	B	C	D	E
O	22	14	25	8	12
E	21.6	16.2	27	5.4	10.8

$$\text{Test statistic } (\chi^2) = \frac{(22-21.6)^2}{21.6} + \frac{(14-16.2)^2}{16.2} + \cdots + \frac{(12-10.8)^2}{10.8}$$
$$= 1.84$$

Degrees of freedom $= 5 - 1 = 4$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi^2_4(5\%) = 9.488$
$\chi^2 <$ c.v. so accept H_0.
Conclude there is no reason to doubt that accidents occur at a constant rate at the factories.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 8

Question:

In a test to determine the red blood cell count in a patient's blood sample, the number of cells in each of 80 squares is counted with the following results.

Number of cells per square, x	0	1	2	3	4	5	6	7	8
Frequency, f	2	8	15	18	14	13	7	3	0

It is assumed that these will fit a Poisson distribution. Test this assertion at the 0.05 significance level.

Solution:

H_0: The data can be modelled by a Poisson distribution.

H_1: The data cannot be modelled by a Poisson distribution.

Total frequency = 80

This is 1 restriction.

Mean = $\lambda = 3.45$

The mean has been calculated so it is 1 restriction.

$$\left.\begin{aligned} E(X=0) &= 80\times P(0) = 80\times\left(e^{-3.45}\times\frac{3.45^0}{0!}\right) = 2.540 \\ E(X=1) &= 80\times P(1) = 80\times\left(e^{-3.45}\times\frac{3.45^1}{1!}\right) = 8.762 \end{aligned}\right\}$$

Pool to get E's 5 or more.

Similarly $E(2) = 15.114, E(3) = 17.381, E(4) = 14.991, E(5) = 10.344$

$E(6) = 5.948, E(7) = 2.931$

$E(8 \text{ or more}) = 80 - [E(0) + E(1) + \cdots + E(7)] = 1.989$

The last 3 classes must be combined to get E to be 5 or more.

X	1 or less	2	3	4	5	6 or more
O	10	15	18	14	13	10
E	11.302	15.114	17.381	14.991	10.344	10.868

$$\text{Test statistic } (\chi^2) = \frac{(10-11.302)^2}{11.302} + \frac{(15-15.114)^2}{15.114} + \cdots + \frac{(10-10.868)^2}{10.868}$$

$$\text{t.s. } (\chi^2) = 0.990$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Degrees of freedom $= 6 - 2$
$= 4$

Number of classes after pooling.

critical value $= \chi^2_4(5\%) = 9.488$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt the data can be modelled by a Poisson distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 9

Question:

A factory has a machine. The number of times it broke down each week was recorded over 100 weeks with the following results.

Number of times broken down	0	1	2	3	4	5
Frequency	50	24	12	9	5	0

It is thought that the distribution is Poisson.

a Give reasons why this assumption might be made.

b Conduct a test at the 0.05 level of significance to see if the assumption is reasonable.

Solution:

a Breakdowns occur singly, independently and at random. They occur at a constant average rate.

b mean $= 0.95$ so let $\lambda = 0.95$

The mean has been calculated so it is 1 restriction.

H_0: The data can be modelled by Po(0.95).

H_1: The data cannot be modelled by Po(0.95).

Total frequency $= 100$

This is 1 restriction. Total frequencies agree.

$$E(X=0) = 100 \times P(X=0) = 100 \times \left[e^{-0.95} \times \frac{0.95^0}{0!} \right] = 38.674$$

$$E(X=1) = 100 \times P(X=1) = 100 \times \left[e^{-0.95} \times \frac{0.95^1}{1!} \right] = 36.740$$

Similarly $E(X=2) = 17.452, E(X=3) = 5.526, E(X=4) = 1.3125$.

There is no need to go further, as further terms are extremely small. Here we find $E(X$ is 3 or more) to get all the E's to be 5 or more.

X	0	1	2	3 or more
O	50	24	12	14
E	38.674	36.740	17.452	7.134

$$\text{Test statistic } \left(\chi^2\right) = \frac{(50-38.674)^2}{38.674} + \frac{(24-36.740)^2}{36.740} + \frac{(12-17.452)^2}{17.452} + \frac{(14-7.134)^2}{7.134}$$

$$\text{t.s. } \left(\chi^2\right) = 16.0$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Degrees of freedom $= 4 - 2$

$= 2$

Number of classes after pooling.

Number of restrictions.

critical value $= \chi_2^2(5\%) = 5.991$

$\chi^2 >$ c.v. so reject H_0.

Conclude that breakdowns cannot modelled by Po(0.95).

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 10

Question:

In a lottery there are 505 prizes, and it is assumed that they will be uniformly distributed throughout the numbered tickets. An investigation gave the following:

Ticket number	1–1000	1001–2000	2001–3000	3001–4000	4001–5000	5001–6000	6001–7000	7001–8000	8001–9000	9001–10000
Frequency	56	49	35	47	63	58	44	52	51	50

Using a suitable test with a 0.05 significance level, and stating your null and alternative hypotheses, see if the assumption is reasonable.

Solution:

H_0: The prizes are uniformly distributed.

H_1: The prizes are not uniformly distributed.

$$\text{Total frequency} = 505$$

$$\therefore \text{ Expected frequency for each class} = \frac{505}{10} = 50.5$$

This is 1 restriction, total frequencies agree.

$$\therefore \text{ Test statistic } \left(\chi^2\right) = \frac{(56-50.5)^2}{50.5} + \frac{(49-50.5)^2}{50.5} + \cdots + \frac{(50-50.5)^2}{50.5}$$

$$\text{t.s. } \left(\chi^2\right) = 10.7$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

$$\text{Degrees of freedom} = 10-1 = 9$$

$$\text{critical value} = \chi^2_9\,(5\%) = 16.919$$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt that prizes are distributed uniformly.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 11

Question:

Data were collected on the number of female puppies born in 200 litters of size 8. It was decided to test whether or not a binomial model with parameters $n = 8$ and $p = 0.5$ is a suitable model for these data. The following table shows the observed frequencies and the expected frequencies, to 2 decimal places, obtained in order to carry out this test.

Number of females	Observed number of litters	Expected number of litters
0	1	0.78
1	9	6.25
2	27	21.88
3	46	R
4	49	S
5	35	T
6	26	21.88
7	5	6.25
8	2	0.78

a Find the values of R, S and T.
b Carry out the test to determine whether or not this binomial model is a suitable one.
State your hypotheses clearly and use a 5% level of significance.

An alternative test might have involved estimating p rather than assuming $p = 0.5$.

c Explain how this would have affected the test.

Solution:

a B(8, 0.5)

Total frequency = 200

This is 1 restriction, total frequencies agree.

$R = 200 \times P(X = 3) = 200 \times \left[\binom{8}{3} \times 0.5^3 \times 0.5^5\right]$

$R = 43.75$

$S = 200 \times P(X = 4) = 200 \times \left[\binom{8}{4} \times 0.5^4 \times 0.5^4\right]$

$S = 54.69$

$T = 200 \times P(X = 5) = 200 \times \left[\binom{8}{5} \times 0.5^5 \times 0.5^3\right]$

$T = 43.75$

b H_0: B(8, 0.5) is a suitable model.
H_1: B(8, 0.5) is not a suitable model.

X	1 or fewer	2	3	4	5	6	7 or more
O	10	27	46	49	35	26	7
E	7.03	21.88	43.75	54.69	43.75	21.88	7.03

Pool the first 2 classes to get E to be 5 or more.

Pool the last 2 classes to get E to be 5 or more.

Test statistic $(\chi^2) = \dfrac{(10-7.03)^2}{7.03} + \dfrac{(27-21.88)^2}{21.88} + \cdots + \dfrac{(7-7.03)^2}{7.03}$

t.s. $(\chi^2) = 5.69$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Number of classes after pooling.

Degrees of freedom $= 7 - 1$
$= 6$

Only 1 restriction.

critical value $= \chi^2_6\ (5\%) = 10.645$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt the data could be modelled by B(8, 0.5).

c We would have to calculate the mean which would give an extra 1 restriction. This would reduce the degrees of freedom by 1 so $\nu = 5$. The critical value would become $\chi^2_5\ (5\%) = 11.070$.

$\chi^2 <$ c.v. so no change in conclusion.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 12

Question:

A random sample of 300 football matches was taken and the number of goals scored in each match was recorded. The results are given in the table below.

Number of goals	0	1	2	3	4	5	6	7
Frequency	33	55	80	56	56	11	5	4

a Show that an unbiased estimate of the mean number of goals scored in a football match is 2.4 and find an unbiased estimate of the variance.

It is thought that a Poisson distribution might provide a good model for the number of goals per match.

b State briefly an implication of a Poisson model on goal scoring at football matches.

Using a Poisson distribution, with mean 2.4, expected frequencies were calculated as follows:

Number of goals	0	1	2	3	4	5	6	7
Expected frequency	s	65.3	t	62.7	37.6	18.1	7.2	2.5

c Find the values of s and t.

d State clearly the hypotheses required to test whether or not a Poisson distribution provides a suitable model for these data.

In order to carry out this test the class for 7 goals is redefined as 7 or more goals.

e Find the expected frequency for this class.

The test statistic for the test in part **d** is 15.6 and the number of degrees of freedom used is 5.

f Explain fully why there are 5 degrees of freedom.

g Stating clearly the critical value used, carry out the test in part **d**, using a 5% level of significance.

Solution:

a $\text{mean} = \dfrac{[0\times33+1\times55+2\times80+3\times56+4\times56+5\times11+6\times5+7\times4]}{[33+55+80+56+56+11+5+4]}$

$\text{mean} = 2.4$

This is 1 restriction in part **d**.

Unbiased estimate of variance = $\dfrac{2426-300\times2.4^2}{299} = 2.4\ldots$ (a Poisson situation $\lambda = \sigma^2$)

b It assumes goals are scored independently and at random, at a constant average rate.

c Total frequency = 300

This is 1 restriction in part **d**.

$s = \mathrm{E}(X=0) = 300\times\mathrm{P}(X=0) = 300\times\left[e^{-2.4}\times\dfrac{2.4^0}{0!}\right] = 27.2$

$t = \mathrm{E}(X=2) = 300\times\mathrm{P}(X=2) = 300\times\left[e^{-2.4}\times\dfrac{2.4^2}{2!}\right] = 78.4$

d H_0: Po(2.4) is a suitable model.

H_1: Po(2.4) is not a suitable model.

e Expected frequency for '7 or more goals'

$= 300-[\mathrm{E}(0)+\mathrm{E}(1)+\mathrm{E}(2)+\mathrm{E}(3)+\mathrm{E}(4)+\mathrm{E}(5)+\mathrm{E}(6)]$

$= 300-[27.2+65.3+78.4+62.7+37.6+18.1+7.2]$

$= 3.5$

f This expected frequency is less than 5 so must be combined with $\mathrm{E}(X=6)$ to give the class '6 or more goals' which now has expected frequency $7.2+3.5=10.7$.
We now have 7 classes after pooling and 2 restrictions so degrees of freedom $= 7-2=5$.

g Test statistic $(\chi^2) = 15.6$

Given in the question.

critical value = $\chi^2_5(5\%) = 11.070$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence the data cannot be modelled by Po(2.4).

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise B, Question 13

Question:

A student of botany believed that *multifolium uniflorum* plants grow in random positions in grassy meadowland. He recorded the number of plants in one square metre of grassy meadow, and repeated the procedure to obtain the 148 results in the table.

Number of plants	0	1	2	3	4	5	6	7 or greater
Frequency	9	24	43	34	21	15	2	0

a Show that, to two decimal places, the mean number of plants in one square metre is 2.59.

b Give a reason why the Poisson distribution might be an appropriate model for these data.

Using the Poisson model with mean 2.59, expected frequencies corresponding to the given frequencies were calculated, to two decimal places, and are shown in the table below.

Number of plants	0	1	2	3	4	5	6	7 or greater
Expected frequencies	11.10	28.76	s	32.15	20.82	10.78	4.65	t

c Find the values of s and t to two decimal places.

d Stating clearly your hypotheses, test at the 5% level of significance whether or not this Poisson model is supported by these data.

Solution:

a $\text{mean} = \dfrac{[0\times 9 + 1\times 24 + 2\times 43 + 3\times 34 + 4\times 21 + 5\times 15 + 6\times 2]}{[9+24+43+34+21+15+2]}$

$\text{mean} = 2.5878\ldots$

$\text{mean} = 2.59$ (2 d.p.)

This is 1 restriction in part **d**.

b It is assumed plants occur at a constant average rate and occur independently and at random in the meadow.

Total frequency $= 148$

This is 1 restriction in part **c**.

c $s = \text{E}(2) = 148\times \text{P}(X=2) = 148\times\left[e^{-2.59}\times\dfrac{2.59^2}{2!}\right]$

$\therefore s = 37.24$ (2 d.p.)

$t = 148 - [11.10 + 28.76 + 37.24 + 32.15 + 20.82 + 10.78 + 4.65]$

$t = 2.50$

d H_0: Po(2.59) is a suitable model.

H_1: Po(2.59) is not a suitable model.

x	0	1	2	3	4	5	6 or more
O	9	24	43	34	21	15	2
E	11.10	28.76	37.24	32.15	20.82	10.78	7.15

Pool the last 2 classes to get E to be 5 or more.

Test statistic $(\chi^2) = \dfrac{(9-11.10)^2}{11.10} + \dfrac{(24-28.76)^2}{28.76} + \cdots + \dfrac{(2-7.15)^2}{7.15}$

t.s. $(\chi^2) = 7.55$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Number of classes after pooling.

Degrees of freedom $= 7 - 2$

$= 5$

Number of restrictions.

critical value $= \chi^2_5(5\%) = 11.070$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt the data can be modelled by Po(2.59).

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 1

Question:

The diameters of a random sample of 30 mass-produced components were measured by checking their diameter with gauges. Of the 30, 18 passed through a 4.0 mm gauge and of these 6 failed to pass through a 3.5 mm gauge. Test the hypothesis that the diameters of the components were a sample of a normal population with mean 3.8 mm and standard deviation 0.5 mm. Use a 5% significance level for your test.

Solution:

H_0: The data can be modelled by $N(3.8, 0.5^2)$.

H_1: The data cannot be modelled by $N(3.8, 0.5^2)$.

x	$x < 3.5$	$3.5 \le x < 4.0$	$x \ge 4.0$
0	12	6	12

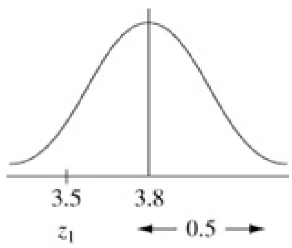
You need to put the given data into a table as shown first.

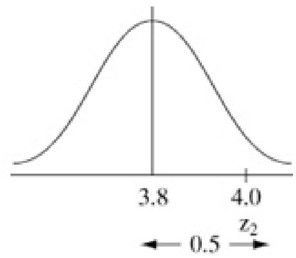
Total frequency = 30

This is 1 restriction.

$$z_1 = \frac{(3.5 - 3.8)}{0.5} = -0.6$$

$$z_2 = \frac{(4.0 - 3.8)}{0.5} = 0.4$$





$$\begin{aligned}
\therefore P(X < 3.5) &= \Phi(-0.6) = 0.2743 \\
\text{so } E(X < 3.5) &= 30 \times 0.2743 = 8.229 \\
P(3.5 \le x \le 4.0) &= \Phi(0.4) - \Phi(-0.6) \\
&= 0.6554 - 0.2743 \\
&= 0.3811 \\
\therefore E(3.5 \le x \le 4.0) &= 30 \times 0.3811 = 11.433 \\
\therefore E(X > 4.0) &= 30 - [8.229 + 11.433] = 10.338
\end{aligned}$$

The total frequency must total 30.

x	$x < 3.5$	$3.5 \le x < 4.0$	$x \ge 4.0$
0	12	6	12
E	8.229	11.433	10.338

$$\text{Test statistic } (\chi^2) = \frac{(6 - 8.229)^2}{8.229} + \frac{(12 - 11.433)^2}{11.433} + \frac{(12 - 10.338)^2}{10.338}$$

$$\text{t.s. } (\chi^2) = 0.899$$

Number of classes.

$$\text{Degrees of freedom } = 3 - 1 = 2$$

Number of restrictions.

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

$$\text{critical value } = \chi^2_2(5\%) = 5.991$$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $N(3.8, 0.5^2)$.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 2

Question:

An egg producer takes a sample of 150 eggs from his flock of chickens and grades them into classes according to their weights as follows.

Class	2	3	4	5	6
Weight	66–70	61–65	56–60	51–55	46–50
Frequency	10	32	67	29	12

Does this distribution fit a normal distribution of mean 58 g and standard deviation 4 g? Use a 5% significance level for your test.

Solution:

H_0: The data can be modelled by $N(58, 4^2)$.
H_1: The data cannot be modelled by $N(58, 4^2)$.

Total frequency $= 150$ ← This is 1 restriction.

Class a	Class b	$z = \dfrac{b-\mu}{\sigma}$ (rounded values)	$P(a \le X \le b)$	Expected frequency (E)
	< 45.5	-3.13	$1 - 0.9991 = 0.0009$	0.135
45.5	50.5	-1.88	$= 0.0291$	4.365
50.5	55.5	-0.63	$= 0.2343$	35.145
55.5	60.5	0.63	$0.7357 - 0.2643 = 0.4714$	70.710
60.5	65.5	1.88	$0.9700 - 0.7357 = 0.2343$	35.145
65.5	70.5	3.13	$0.9991 - 0.9900 = 0.0291$	4.365
> 70.5			$1 - 0.9991 = 0.0009$	0.135

These 3 classes must be pooled to get E to be 5 or more.

These 3 classes must be combined to get E to be 5 or more.

The Z-scores are symmetrical for the classes so we can deduce the probabilities for the first 3 classes since they are the same as for the last 3 classes. This is also true for the expected frequencies.

Class	55.5 or less	55.5 to 65.5	65.5 or more
O	42	67	41
E	39.645	70.710	39.645

Test statistic $(\chi^2) = \dfrac{(42-39.645)^2}{39.645} + \dfrac{(67-70.710)^2}{70.710} + \dfrac{(41-39.645)^2}{39.645}$

t.s. $(\chi^2) = 0.381$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

A χ^2 value in the range 0.28–0.39 would be accepted.

Number of classes after pooling.

Degrees of freedom $= 3 - 1$ ← Number of restrictions.
$= 2$

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

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $N(58, 4^2)$.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 3

Question:

A sample of 100 apples is taken from a load. The apples have the following distribution of sizes.

Diameter (cm)	≤ 6	7	8	9	≥ 10
Frequency	8	29	38	16	9

It is thought that they come from a normal distribution with mean diameter of 8 cm and a standard deviation of 0.9 cm. Test this assertion using a 0.05 level of significance.

Solution:

H_0: The data can be modelled by $N(8, 0.9^2)$.

H_1: The data cannot be modelled by $N(8, 0.9^2)$.

Total frequency $= 100$

This is 1 restriction.

Class a	Class b	$z = \frac{b-\mu}{\sigma}$ (rounded values)	$P(a \le X \le b)$	E
	≤ 6.5	-1.67	0.0475	4.75
6.5	7.5	-0.56	0.2402	24.02
7.5	8.5	0.56	$0.7123 - 0.2877 = 0.4246$	42.46
8.5	9.5	1.67	$0.9525 - 0.7123 = 0.2402$	24.02
≥ 9.5			$1 - 0.9525 = 0.0475$	4.75

These classes must be pooled to get all the E's to be 5 or more.

Same here.

Because the Z-scores are symmetrical we can fill in the probabilities for the first two classes with the values from the last 2 classes. We can do the same with the E's.

Class	7 or less	8	9 or more
O	37	38	25
E	28.77	42.46	28.77

$$\text{Test statistic } (\chi^2) = \frac{(37-28.77)^2}{28.77} + \frac{(38-42.46)^2}{42.46} + \frac{(25-28.77)^2}{28.77}$$
$$= 3.32$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

χ^2 values accepted in a range 3.2–3.33.

Degrees of freedom $= 3 - 1$

$= 2$

Number of classes after pooling.

Restriction.

critical value $= \chi_2^2\,(5\%) = 5.991$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $N(8, 0.9^2)$.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 4

Question:

A shop owner found that the number of cans of a particular drink sold per day during 100 days in summer was as follows.

Drinks per day, *d*	0–	10–	20–	30–	40–
Frequency of *d*	10	24	45	14	7

It is thought that these data can be modelled by a normal distribution.

a Estimate values of μ and σ and conduct a goodness of fit test using 1% significance level.

b Explain how the shopkeeper might use this model.

Solution:

a

Mid-interval value	5	15	25	35	45
Frequency	10	24	45	14	7

mean $= \bar{x} = 23.4$
standard deviation $= S = 10.222$
Total frequency $= 100$

Get these directly from the calculator.

This is 1 restriction.

Because the mean *and* standard deviation have been estimated by *calculation* they give a further 2 restrictions. Here we have 3 restrictions in total.

H_0: The data can may be modelled by $N(23.4, 10.222^2)$.

H_1: The data cannot may be modelled by $N(23.4, 10.222^2)$.

Class a	Class b	$z = \frac{b-\mu}{\sigma}$ (rounded values)	$P(a \le X \le b)$	E
	≤ 10	-1.31	0.0951	9.51
10	20	-0.33	$0.3707 - 0.0951 = 0.2756$	27.56
20	30	0.65	$0.7422 - 0.3707 = 0.3715$	37.15
30	40	1.62	$0.9474 - 0.7422 = 0.2052$	20.52
≥ 40			$1 - 0.9474 = 0.0526$	5.26

No pooling needed as *all* the E's are 5 or more.

Class	≤ 10	10–	20–	30–	40 or more
O	10	24	45	14	7
E	9.51	27.56	37.15	20.52	5.26

Test statistic $(\chi^2) = \frac{(10-9.51)^2}{9.51} + \frac{(24-27.56)^2}{27.56} + \cdots + \frac{(7-5.26)^2}{5.26}$

t.s. $(\chi^2) = 4.79$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

χ^2 values between 4.79 and 4.94 accepted.

Number of classes.

Degrees of freedom $= 5 - 3$
$= 2$

Number of restrictions.

critical value $= \chi^2_2(1\%) = 9.210$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $N(23.4, 10.222^2)$.

b The shopkeeper could use it to work out the maximum number of cans of the drink they need to keep in stock in order to cope with demand.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 5

Question:

An outfitter sells boys' raincoats and these are stocked in four sizes.
Size 1 fits boys up to 1.25 m in height. Size 2 fits boy from 1.26 to 1.31 m. Size 3 fits boys from 1.32 to 1.37 m. Size 4 fits those over 1.38 m. To assist the outfitter in deciding the stock levels he should order each year, the heights of 120 boys in the right age range were measured with the following results.

Height, h(m)	1.20–1.22	1.23–1.25	1.26–1.28	1.29–1.31	1.32–1.34
Frequency of h	9	9	18	23	20
Height, h(m)	1.35–1.37	1.38–1.40	1.41–1.43		
Frequency of h	19	17	5		

It is suggested that a suitable model for these data is N(1.32, 0.0016).

a Conduct a goodness of fit test using a $2\frac{1}{2}\%$ significance level.

b Estimate values of μ and σ using the observed values and using these conduct a goodness of fit test using a $2\frac{1}{2}\%$ significance level.

c Select the best model and use it to tell the outfitter how many of each size should be ordered per year if the normal annual sales are 1200.

Solution:

a H_0: The data can be modelled by N(1.32, 0.0016).
H_1: The data cannot be modelled by N(1.32, 0.0016).
Here the variance $= 0.0016$ so the standard deviation $= \sqrt{0.0016} = 0.04$
i.e. $\sigma = 0.04$

Class A	b	$z = \frac{b-\mu}{\sigma}$ (rounded values)	$P(a \le X \le b)$	E
	≤ 1.225	-2.38	0.0009	0.108
1.225	1.255	-1.63	$0.0516 - 0.0009 = 0.0507$	6.084
1.255	1.285	-0.88	$0.1894 - 0.0516 = 0.1378$	16.536
1.285	1.315	-0.13	$0.4483 - 0.1894 = 0.2589$	31.068
1.315	1.345	0.63	$0.7357 - 0.4483 = 0.2874$	31.488
1.345	1.375	1.38	$0.9162 - 0.7357 = 0.1805$	21.660
1.375	1.405	2.13	$0.9834 - 0.9162 = 0.0672$	8.064
≥ 1.405			$1 - 0.9834 = 0.0166$	1.992

Pool to get E to be 5 or more.

Same here.

Total frequency $= 120$

This is 1 restriction.

Class	1.25 or less	1.26–1.28	1.29–1.31	1.32–1.34	1.35–1.37	1.38 or more
O	18	18	23	20	19	22
E	6.192	16.536	31.068	31.488	21.660	10.056

$$\text{Test statistic } (\chi^2) = \frac{(18-6.192)^2}{6.192} + \frac{(18-16.536)^2}{16.536} + \cdots + \frac{(22-10.056)^2}{10.056}$$

$$\text{t.s. } (\chi^2) = 43.4$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

χ^2 values in the range 43.35 to 44.4 accepted.

Number of classes after pooling.

Degrees of freedom $= 6 - 1$
$= 5$

Number of restrictions.

critical value $= \chi^2_5 (2\frac{1}{2}\%) = 12.833$

$\chi^2 >$ c.v. so reject H_0.
Conclude the data cannot be modelled by N(1.32, 0.0016).

b

Mid-interval value	1.21	1.24	1.27	1.30	1.33	1.36	1.39	1.42
Frequency	9	9	18	23	20	19	17	5

Mean $= 1.3165 = \mu$
standard derivation $= s = 0.0569 \simeq \sigma$
Total frequency $= 120$

This gives 3 restrictions in total.

H_0: The data can be modelled by $N(1.3165, 0.0569^2)$.

H_1: The data cannot be modelled by $N(1.3165, 0.0569^2)$.

Class a	Class b	$z = \frac{b-\mu}{\sigma}$ (rounded values)	$P(a \le X \le b)$	E
	≤ 1.225	-1.61	0.0537	6.444
1.225	1.255	-1.08	$0.1401 - 0.0537 = 0.0864$	10.368
1.255	1.285	-0.55	$0.2912 - 0.1401 = 0.1511$	18.132
1.285	1.315	-0.03	$0.4880 - 0.2912 = 0.1968$	23.616
1.315	1.345	0.50	$0.6915 - 0.4880 = 0.2035$	24.420
1.345	1.375	1.03	$0.8485 - 0.6915 = 0.1570$	18.840
1.375	1.405	1.56	$0.9406 - 0.8485 = 0.0921$	11.052
≥ 1.405			$1 - 0.9406 = 0.0594$	7.128

Class	1.22 or less	1.23–1.25	1.26–1.28	1.29–1.31	1.32–1.34	1.35–1.37	1.38–1.40	1.41 or more
O	9	9	18	23	20	19	17	5
E	6.444	10.368	18.132	23.616	24.420	18.840	11.052	7.128

$$\text{Test statistic } \left(\chi^2\right) = \frac{(9-6.444)^2}{6.444} + \frac{(9-10.368)^2}{10.368} + \cdots + \frac{(5-7.128)^2}{7.128}$$

$$\text{t.s. } \left(\chi^2\right) = 5.85$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

χ^2 can be in range 5.78 to 5.85.

Degrees of freedom $= 8 - 3$
$= 5$

critical value $= \chi_5^2 \left(2\frac{1}{2}\%\right) = 12.832$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $N(1.3165, 0.0569^2)$.

c The outfitter should use the model in part **b** because it is based on the experimental evidence.

If normal annual sales are 1200 (this is 10 times what is in the model so multiply the E's by 10) s/he should order

Size 1 $(6.444 + 10.368) \times 10$ i.e. 168
Size 2 $(18.132 + 23.616) \times 10$ i.e. 417
Size 3 $(24.420 + 18.840) \times 10$ i.e. 433
Size 4 $(11.052 + 7.128) \times 10$ i.e. 182

Those will vary according to the answers in part **b**.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 6

Question:

A hamster breeder is studying the weight of adult hamsters. Each hamster from a random sample of 50 hamsters is weighed and the results, given to the nearest g, are recorded in the following table.

Weight (g)	85–94	95–99	100–104	105–109	110–119
Frequency	6	9	17	14	4

a Show that an estimate of the mean weight of the hamsters is 102 g.
The breeder proposes that the weight of an adult hamster, in g, should be modelled by the random variable W having a normal distribution with standard deviation 6. The breeder fits a normal distribution and obtains the following expected frequencies.

W	$W \le 94.5$	$94.5 < W \le 99.5$	$99.5 < W \le 104.5$	$104.5 < W \le 109.5$	$W > 109.5$
Expected frequency	r	11.64	s	11.64	t

b Find the values of r, s and t.

c Stating your hypotheses clearly, test at the 10% level of significance whether or not a normal distribution with a standard deviation of 6 is a suitable model for W.

Solution:

a

Mid-interval value	89.5	97	102	107	114.5
Frequency	6	9	17	14	4

$$\text{mean} = \frac{[6\times89.5+9\times97+17\times102+14\times107+4\times114.5]}{[6+9+17+14+4]}$$

$\text{mean} = 102$ This is 1 restriction in part **c**.

b Total frequency $= 50$ This is 1 restriction in part **c**.

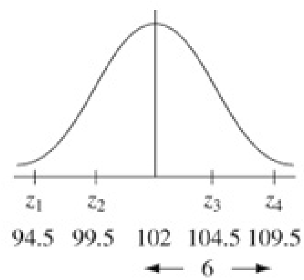
$\sigma = 6$ given This is *not* a restriction since an estimate of it isn't calculated.

$$z_1 = \frac{(94.5-102)}{6} = -1.25$$

$$z_2 = \frac{(99.5-102)}{6} = -0.42$$

$$z_3 = \frac{(104.5-102)}{6} = 0.42$$

$$z_4 = \frac{(109.5-102)}{6} = 1.25$$



$t = \mathrm{E}(W > 109.5) = 50\times[1-\Phi(1.25)] = 50\times[1-0.8944] = 5.28$
so $t = 5.28$
and from symmetry (because the Z-scores are symmetrical)
$r = 5.28$
and $s = 50-[5.28+11.64+11.64+5.28]$
so $s = 16.16$

c H_0: The data can be modelled by $\mathrm{N}(102, 6^2)$.
H_1: The data cannot be modelled by $\mathrm{N}(102, 6^2)$.

Class	$W \le 94.5$	$94.5 < W \le 99.5$	$99.5 < W < 104.5$	$104.5 < W \le 109.5$	$W > 109.5$
O	6	9	17	14	4
E	5.28	11.64	16.16	11.64	5.28

$$\text{Test statistic } (\chi^2) = \frac{(6-5.28)^2}{5.28} + \frac{(9-11.64)^2}{11.64} + \frac{(17-16.16)^2}{16.16} + \frac{(14-11.64)^2}{11.64} + \frac{(4-5.28)^2}{5.28}$$

t.s. $(\chi^2) = 1.53$

Number of classes.

Degrees of freedom $= 5-2 = 3$

Number of restrictions.

Critical value $= \chi^2_3(10\%) = 6.251$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by $\mathrm{N}(102, 6^2)$.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise C, Question 7

Question:

A hamster breeder is studying the weight of adult hamsters. Each hamster from a random sample of 50 hamsters is weighed and the results, given to the nearest g, are recorded in the following table.

Weight (g)	85–94	95–99	100–104	105–109	110–119
Frequency	6	9	17	14	4

a Show that an estimate of the mean weight of the hamsters is 102 g.
The breeder proposes that the weight of an adult hamster, in g, should be modelled by the random variable W having a normal distribution with standard deviation 6. The breeder fits a normal distribution and obtains the following expected frequencies.

W	$W \le 94.5$	$94.5 < W \le 99.5$	$99.5 < W \le 104.5$	$104.5 < W \le 109.5$	$W > 109.5$
Expected frequency	r	11.64	s	11.64	t

b Find the values of r, s and t.

c Stating your hypotheses clearly, test at the 10% level of significance whether or not a normal distribution with a standard deviation of 6 is a suitable model for W.

Solution:

a

Mid-interval value	89.5	97	102	107	114.5
Frequency	6	9	17	14	4

$$\text{mean} = \frac{[6\times 89.5 + 9\times 97 + 17\times 102 + 14\times 107 + 4\times 114.5]}{[6+9+17+14+4]}$$

$\text{mean} = 102$ ← This is 1 restriction in part c.

b Total frequency $= 50$ ← This is 1 restriction in part c.

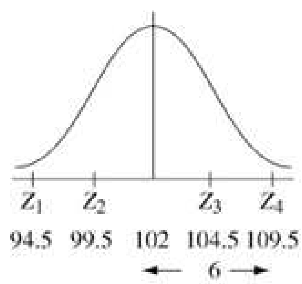
$\sigma = 6$ given ← This is *not* a restriction since an estimate of it isn't calculated.

$$Z_1 = \frac{(94.5-102)}{6} = -1.25$$

$$Z_2 = \frac{(99.5-102)}{6} = -0.42$$

$$Z_3 = \frac{(104.5-102)}{6} = 0.42$$

$$Z_4 = \frac{(109.5-102)}{6} = 1.25$$



$t = \mathrm{E}(W > 109.5) = 50\times[1-\Phi(1.25)] = 50\times[1-0.8944] = 5.28$

so $t = 5.28$

and from symmetry (because the Z-scores are symmetrical)

$r = 5.28$

and $s = 50 - [5.28 + 11.64 + 11.64 + 5.28]$

so $s = 16.16$

c H_0: The data are from $\mathrm{N}(102, 6^2)$.

H_1: The data aren't from $\mathrm{N}(102, 6^2)$.

Class	$W \le 94.5$	$94.5 < W \le 99.5$	$99.5 < W < 104.5$	$104.5 < W \le 109.5$	$W > 109.5$
O	6	9	17	14	4
E	5.28	11.64	16.16	11.64	5.28

Test statistic $= \frac{(6-5.28)^2}{5.28} + \frac{(9-11.64)^2}{11.64} + \frac{(17-16.16)^2}{16.16} + \frac{(14-11.64)^2}{11.64} + \frac{(4-5.28)^2}{5.28}$

t.s. $= 1.53$

Number of classes.

Degrees of freedom $= 5 - 2 = 3$ ← Number of restrictions.

Critical value $= \chi_3^2\,(10\%) = 6.251$

t.s. < c.v. so accept H_0.

Conclude no reason to doubt the data could be from $\mathrm{N}(102, 6^2)$.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 1

Question:

When analysing the results of a 3×2 contingency table it was found that

$$\sum_{i=1}^{6} \frac{\left(O_i - E_i\right)^2}{E_i} = 2.38$$

Write down the number of degrees of freedom and the critical value appropriate to these data in order to carry out a χ^2 test of significance at the 5% level. ***E***

Solution:

$\nu = 2, \quad \chi_2^2\,(5\%) = 5.99$

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 2

Question:

Three different types of locality were studied to see if the ownership, or non-ownership, of a television was or was not related to the locality. $\sum \frac{(O_i - E_i)^2}{E_i}$ was evaluated and found to be 13.1. Using a 5% level of significance, carry out a suitable test and state your conclusion.

Solution:

H_0 : Ownership is not related to locality.
H_1: Ownership is related to locality.
Test statistic $\chi^2 \quad = 13.1$
Degrees of freedom $= (3-1) \times (2-1) = 2$
critical value $= \chi^2_2\ (5\%) = 5.991$
$\chi^2 >$ c.v. so reject H_0.
Conclude there is evidence that ownership of a television is related to the locality.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 3

Question:

In a college, three different groups of students sit the same examination. The results of the examination are classified as Credit, Pass or Fail. In order to test whether or not there is a difference between the groups with respect to the proportions of students in the three grades, the statistic $\sum \frac{(O_i - E_i)^2}{E_i}$ is evaluated and found to be equal to 10.28.

a Explain why there are 4 degrees of freedom in this situation.
b Using a 5% level of significance, carry out the test and state your conclusions.

Solution:

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

b H_0: There is no association between the group and the results.
H_1: There is an association between the group and the results.
Test statistic $= 10.28$
critical value $= \chi_4^2 (5\%) = 9.488$
$\chi^2 >$ c.v. so reject H_0.
Conclude there is evidence of an association between the groups and the exam results they achieve.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 4

Question:

The grades of 200 students in both Mathematics and English were studied with the following results.

		English grades		
		A	*B*	*C*
Maths grades	*A*	17	28	18
	B	38	45	16
	C	12	12	14

Using a 0.05 significance level, test these results to see if there is an association between English and Mathematics results. State your conclusions.

Solution:

H_0 : There is no association between Mathematics and English results.

H_1: There is an association between Mathematics and English results.

Observed frequencies

		English grades *A*	*B*	*C*	Total
Maths grades	*A*	17	28	18	63
	B	38	45	16	99
	C	12	12	14	38
	Total	67	85	48	200

Expected frequency Maths *A* and English *A*

$$= \frac{63 \times 67}{200} = 21.105$$

Expected frequencies

		English Grades *A*	*B*	*C*
Maths grades	*A*	21.105	26.775	15.120
	B	33.165	42.075	23.760
	C	12.730	16.150	9.120

$$\text{Test statistic } (\chi^2) = \frac{(17-21.105)^2}{21.105} + \frac{(28-26.775)^2}{26.775} + \cdots + \frac{(14-9.120)^2}{9.120}$$

$$\text{t.s. } (\chi^2) = 8.57 \text{ (or 8.56)}$$

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

> Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value = χ^2_4 (5%) $= 9.488$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no evidence of an association between Mathematics and English results.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 5

Question:

The number of trains on time, and the number of trains that were late was observed at three different London stations. The results were:

		Observed frequency	
		On time	**Late**
Station	***A***	26	14
	B	30	10
	C	44	26

Using the χ^2 statistic and a significance test at the 5% level, decide if there is any association between station and lateness.

Solution:

H_0: There is no association between station and lateness.
H_1: There is an association between station and lateness.

Observed frequencies

		On Time	*Late*	*Total*
Station	*A*	26	14	40
	B	30	10	40
	C	44	26	70
Total		100	50	150

Expected frequency 'On time' and 'station *A*'

$$= \frac{40 \times 100}{150} = 26.666\ldots \text{ or } 26\frac{2}{3}$$

Expected frequencies

		On Time	*Late*
Station	*A*	26.666	13.333
	B	26.666	13.333
	C	46.666	23.333

$$\text{Test statistic } (\chi^2) = \frac{(26-26.666)^2}{26.666} + \frac{(14-13.333)^2}{13.333} + \cdots + \frac{(26-23.333)^2}{23.333}$$

$$\text{t.s.}(\chi^2) = 1.76$$

$$\text{Degrees of freedom} = (3-1)\times(2-1)$$
$$= 2$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi^2_2(5\%) = 5.991$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no evidence of an association between station and lateness.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 6

Question:

In addition to being classed into grades A, B, C, D and E, 200 students are classified as male or female and their results summarised in a contingency table. Assuming all expected values are 5 or more, the statistic $\sum \frac{(O_i - E_i)^2}{E_i}$ was 14.27.

Stating your hypotheses and using a 1% significance level, investigate whether or not gender and grade are associated.

Solution:

H_0: There is no association between gender and grade.

H_1: There is an association between gender and grade.

Test statistic $\chi^2 = 14.27$

Degrees of freedom $= (5-1) \times (2-1) = 4$

critical value $= \chi_4^2\ (1\%) = 13.277$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between the gender of the student and the grade achieved.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 7

Question:

In a random sample of 60 articles made in factory A, 8 were defective. In factory B, 6 out of 40 similar articles were defective.

a Draw up a contingency table.

b Test at the 0.05 significance level the hypothesis that quality was independent of the factory involved.

Solution:

a

Observed frequencies →

	Factory		
	A	*B*	Total
OK	52	34	86
Defective	8	6	14
Total	60	40	100

b H_0 : quality is independent of factory.

H_1: quality isn't independent of factory.

$$\text{Expected frequency 'OK' and '}A\text{'} = \frac{86 \times 60}{100} = 51.6$$

Expected frequencies →

	A	*B*
OK	51.6	34.4
Defective	8.4	5.6

$$\text{Test statistic } (\chi^2) = \frac{(52-51.6)^2}{51.6} + \frac{(34-34.4)^2}{34.4} + \cdots + \frac{(6-5.6)^2}{5.6}$$
$$= 0.0554$$

$$\text{Degrees of freedom} = (2-1)\times(2-1) = 1$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_1^2(5\%) = 3.841$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt that quality is independent of factory.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 8

Question:

During an influenza epidemic, 15 boys and 8 girls became ill out of a class of 22 boys and 28 girls. Assuming that this group may be treated as a random sample of the age group, test at the 5% significance level the hypothesis that there is no connection between gender and susceptibility to influenza.

Solution:

Observed frequencies →

	Boys	*Girls*	Total
Flu	15	8	23
No Flu	7	20	27
Total	22	28	50

H_0 : There is no association between gender and susceptibility to influenza.

H_1: There is an association between gender and susceptibility to influenza.

Expected frequency 'Boy' and 'Flu' $= \dfrac{23 \times 22}{50} = 10.12$

Expected frequencies →

	Boys	*Girls*
Flu	10.12	12.88
No Flu	11.88	15.12

Test statistic $\left(\chi^2\right) = \dfrac{(15-10.12)^2}{10.12} + \dfrac{(8-12.88)^2}{12.88} + \cdots + \dfrac{(20-15.12)^2}{15.12}$

t.s. $\left(\chi^2\right) = 7.78$

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

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_1^2\ (5\%) = 5.991$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between gender and susceptibility to influenza.

More boys were observed to catch flu than would be expected and fewer girls were observed to catch flu than would be expected so it appears boys are more susceptible to flu.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 9

Question:

In a study of marine organisms, a biologist collected specimens from three beaches and counted the number of males and females in each sample, with the following results.

		Beach		
		A	***B***	***C***
Gender	*Male*	46	80	40
	Female	54	120	160

Using a significance level of 5%, test these results to see if there is any difference between the beaches with regard to the numbers of male and female organisms.

Solution:

H_0: There is no association between gender and beach.
H_1: There is an association between gender and beach.

Observed frequencies →

		Beach			
		A	*B*	*C*	Total
Gender	*Male*	46	80	40	166
	Female	54	120	160	334
	Total	100	200	200	500

Expected frequency 'male' and '*A*' $= \dfrac{166 \times 100}{500} = 33.2$

Expected frequencies →

	A	*B*	*C*
Male	33.2	66.4	66.4
Female	66.8	133.6	133.6

Test statistic $\left(\chi^2\right) = \dfrac{(46-33.2)^2}{33.2} + \dfrac{(80-66.4)^2}{66.4} + \cdots + \dfrac{(160-133.6)^2}{133.6}$

t.s. $\left(\chi^2\right) = 27.3$

Degrees of freedom $= (2-1)\times(3-1) = 2$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_2^2\,(5\%) = 5.991$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between the gender of the organism and the beach on which it is found.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise D, Question 10

Question:

A research worker studying the ages of adults and the number of credit cards they possess obtained the results shown below;

		Number of cards	
		≤ 3	> 3
Age	< 30	74	20
	≥ 30	50	35

Use the χ^2 statistic and a significance test at the 5% level to decide whether or not there is an association between age and number of credit cards possessed. ***E***

Solution:

H_0: There is no association between age and number of credit cards.

H_1: There is an association between age and number of credit cards.

Observed Frequencies

		Number of cards		
		≤ 3	> 3	Total
Age	< 30	74	20	94
	≥ 30	50	35	85
	Total	124	55	179

Expected frequency '< 30' and '≤ 3' $= \dfrac{94 \times 124}{179} = 65.117$

Expected Frequencies

	≤ 3	> 3
< 30	65.117	28.883
≥ 30	58.883	26.117

$$\text{Test statistic } \left(\chi^2\right) = \frac{(74-65.117)^2}{65.117} + \frac{(20-28.883)^2}{28.883} + \ldots + \frac{(35-26.117)^2}{26.117}$$

$$= 8.31 (\text{or } 8.30 \text{ direct from calculator})$$

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

> Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_1^2 (5\%) = 3.841$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between age and number of credit cards possessed.

More people over 30 were observed to carry 3 or more cards than would be expected and more people under 30 were observed to carry fewer than 3 credit cards than would be expected.

Older people are more likely to possess more credit cards than are younger people.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 1

Question:

The random variable Y has a χ^2 distribution with 10 degrees of freedom. Find y such that $P(Y < y) = 0.99$. ***E***

Solution:

23.209

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 2

Question:

The random variable X has a chi-squared distribution with 8 degrees of freedom. Find x such that $\mathrm{P}(X > x) = 0.05$. ***E***

Solution:

15.507

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 3

Question:

As part of an investigation into visits to a Health Centre a 5×3 contingency table was constructed. A χ^2 test of significance at the 5% level is to be carried out on the table. Write down the number of degrees of freedom and the critical region appropriate to this test. ***E***

Solution:

$$\text{Degrees of freedom} = (5-1) \times (3-1)$$
$$= 8$$

Critical region is $\chi^2 > 15.507$

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 4

Question:

Data are collected in the form of a 4×4 contingency table.
To carry out a χ^2 test of significance one of the rows was amalgamated with another row and the resulting value of $\sum\frac{(O-E)^2}{E}$ calculated.

Write down the number of degrees of freedom and the critical value of χ^2 appropriate to this test assuming a 5% significance level. ***E***

Solution:

Amalgamation gives a 3×4 contingency table

$$\begin{aligned}\text{Degrees of freedom} &= (3-1)\times(4-1)\\ &= 6\\ \text{critical value} &= \chi_6^2\ (5\%)\\ &= 12.592\end{aligned}$$

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 5

Question:

A new drug to treat the common cold was used with a randomly selected group of 100 volunteers. Each was given the drug and their health was monitored to see if they caught a cold. A randomly selected control group of 100 volunteers was treated with a dummy pill. The results are shown in the table below.

	Cold	No cold
Drug	34	66
Dummy Pill	45	55

Using a 5% significant level, test whether or not the chance of catching a cold is affected by taking the new drug. State your hypotheses carefully. ***E***

Solution:

H_0: There is no association between catching a cold and taking the new drug.
H_1: There is an association between catching a cold and taking the new drug.

Observed frequencies →

	Cold	*No cold*	Total
Drug	34	66	100
Dummy	45	55	100
Total	79	121	200

Expected frequency 'Drug' and 'cold' $= \frac{100 \times 79}{200} = 39.5$

Expected Frequencies →

	Cold	*No cold*
Drug	39.5	60.5
Dummy	39.5	60.5

Test statistic $(\chi^2) = \frac{(34-39.5)^2}{39.5} + \frac{(66-60.5)^2}{60.5} + \frac{(45-39.5)^2}{39.5} + \frac{(55-60.5)^2}{60.5}$

t.s. $(\chi^2) = 2.53$

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

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

critical value $= \chi_1^2\,(5\%) = 3.841$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no evidence that there is an association between catching a cold and taking the new drug. It appears taking the new drug doesn't affect the chance of a person catching a cold.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 6

Question:

Breakdowns on a certain stretch of motorway were recorded each day for 80 consecutive days. The results are summarised in the table below.

Number of breakdowns	0	1	2	> 2
Frequency	38	32	10	0

It is suggested that the number of breakdowns per day can be modelled by a Poisson distribution.
Using a 5% significant level, test whether or not the Poisson distribution is a suitable model for these data. State your hypotheses clearly. ***E***

Solution:

H_0: The data can be modelled by a Poisson distribution.
H_1: The data cannot be modelled by a Poisson distribution.

Total frequency = 80 — This is 1 restriction (total frequencies agree.)

mean $= \bar{x} = 0.65$ — This is 1 restriction (calculation of estimate of parameter.) Get this directly from your calculator.

$$E(0) = 80 \times P(0) = 80 \times \left[e^{-0.65} \times \frac{0.65^0}{0!} \right] = 41.764$$

$$E(1) = 80 \times P(1) = 80 \times \left[e^{-0.65} \times \frac{0.65^1}{1!} \right] = 27.146$$

$$E(2) = 80 \times P(2) = 80 \times \left[e^{-0.65} \times \frac{0.65^2}{2!} \right] = 8.823$$

Pool these classes to get all the E's to be 5 or more.

$$E(3 \text{ or more}) = 80 - [41.764 + 27.146 + 8.823] = 2.267$$

Class	0	1	2 or more
O	38	32	10
E	41.764	27.146	11.09

$$\text{Test statistic } (\chi^2) = \frac{(38-41.764)^2}{41.764} + \frac{(32-27.146)^2}{27.146} + \frac{(10-11.09)^2}{11.09}$$

$$\text{t.s. } (\chi^2) = 1.31$$

Degrees of freedom $= 3 - 2 = 1$

Number of classes after pooling.

Number of restrictions.

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Critical value $= \chi_1^2 (5\%) = 3.841$

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt that the data can be modelled by a Poisson distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 7

Question:

A survey in a college was commissioned to investigate whether ot not there was any association between gender and passing a driving test. A group of 50 males and 50 females were asked whether they passed or failed their driving test at the first attempt. All the students asked had taken the test. The results were as follows.

	Pass	Fail
Male	23	27
Female	32	18

Stating your hypotheses clearly test, at the 10% level, whether or not there is any evidence of an association between gender and passing a driving test at the first attempt. ***E***

Solution:

H_0 : There is no association between gender and passing a driving test 'first time'.
H_1: There is an association between gender and passing a driving test 'first time'.

Observed frequencies →

	Pass	*Fail*	Total
Male	23	27	50
Female	32	18	50
Total	55	45	100

Expected frequency 'Male' and 'pass' $= \dfrac{50 \times 55}{100} = 27.5$

Expected frequencies →

	Pass	*Fail*
Male	27.5	22.5
Female	27.5	22.5

Test statistic $(\chi^2) = \dfrac{(23-27.5)^2}{27.5} + \dfrac{(27-22.5)^2}{22.5} + \dfrac{(32-27.5)^2}{27.5} + \dfrac{(18-22.5)^2}{22.5}$

t.s. $(\chi^2) = 3.27$

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

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Critical value $= \chi_1^2(10\%) = 2.706$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between gender and passing a driving test 'first time'.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 8

Question:

Successful contestants in a TV game show were allowed to select from one of five boxes, four of which contained prizes, and one of which contained nothing. The boxes were numbered 1 to 5, and, when the show had run for 100 weeks, the choices made by the contestants were analysed with the following results:

Box number	1	2	3	4	5
Frequency	20	16	25	18	21

a Explain why these data could possibly be modelled by a discrete uniform distribution.

b Using a significance level of 5%, test to see if the discrete uniform distribution is a good model in this particular case.

Solution:

a We would expect each box to have an equal chance of being opened. The box numbers are discrete values.

b H_0: The data can be modelled by a discrete uniform distribution.

H_1: The data cannot be modelled by a discrete uniform distribution.

Total frequency $= 100$

This is 1 restriction (total frequencies agree.)

There are 5 boxes so expected frequency for each $= \frac{100}{5} = 20$

Box	1	2	3	4	5
O	20	16	25	18	21
E	20	20	20	20	20

$$\text{Test statistic } \left(\chi^2\right) = \frac{(20-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(21-20)^2}{20}$$

$$\text{t.s. } \left(\chi^2\right) = 2.3$$

Degrees of freedom $= 5 - 1 = 4$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

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

$\chi^2 <$ c.v. so accept H_0.

Conclude there is no reason to doubt that the data could be modelled by a discrete uniform distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 9

Question:

A pesticide was tested by applying it in the form of a spray to 50 samples of 5 flies. The numbers of dead flies after 1 hour were then counted with the following results:

Number of dead flies	0	1	2	3	4	5
Frequency	1	1	5	11	24	8

a Calculate the probability that a fly dies when sprayed.
b Using a significance level of 5%, test to see if these data could be modelled by a binomial distribution.

Solution:

a Total number of dead flies $= 0\times1+1\times1+2\times5+3\times11+4\times24+5\times8$
$= 180$

Total number of flies sprayed $= 50\times5$
$= 250$

$\therefore$ P (fly dies when sprayed) $= \frac{180}{250} = 0.72$

This is 1 restriction in part **b** (calculation of estimate of parameter).

b H_0: The data can be modelled by B(5, 0.72).
H_1: The data cannot be modelled by B(5, 0.72).

This is 1 restriction (total frequencies agree).

Total frequency = 50

$$\left.\begin{aligned}E(0) &= 50\times P(0) = 50\times\left[\binom{5}{0}\times0.72^0\times0.28^5\right] = 0.0861\\ E(1) &= 50\times P(1) = 50\times\left[\binom{5}{1}\times0.72^1\times0.28^4\right] = 1.1064\\ E(2) &= 50\times P(2) = 50\times\left[\binom{5}{2}\times0.72^2\times0.28^3\right] = 5.6900\end{aligned}\right\}$$

Pool these classes to get all E's to be 5 or more.

Similarly $E(3) = 14.6313$, $E(4) = 18.8117$, $E(5) = 9.6746$

Number of dead flies	2 or fewer	3	4	5
O	7	11	24	8
E	6.8825	14.6313	18.8117	9.6746

Test statistic $(\chi^2) = \frac{(7-6.8825)^2}{6.8825} + \frac{(11-14.6313)^2}{14.6313} + \frac{(24-18.8117)^2}{18.8117} + \frac{(8-9.6746)^2}{9.6746}$

t.s. $(\chi^2) = 2.62$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Number of classes after pooling.

Degrees of freedom $= 4 - 2 = 2$

Number of restrictions.

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

$\chi^2 <$ c.v. so accept H_0.

Conclude we have no reason to doubt that B(5, 0.72) is a good model for these data.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 10

Question:

The number of accidents per week at a certain road junction was monitored for four years. The results obtained are summarised in the table.

Number of accidents	0	1	2	> 2
Number of weeks	112	56	40	0

Using a 5% level of significance, carry out a χ^2 test of the hypothesis that the number of accidents per week has a Poisson distribution. ***E***

Solution:

H_0 : A Poisson distribution is a good model.
H_1: A Poisson distribution is not a good model.

Total frequency $= 112+56+40 = 208$

This is 1 restriction (total frequencies agree).

mean $= 0.654$

This is 1 restriction (calculation of estimate of parameter).

$$E(0) = 208\times P(0) = 208\times\left[e^{-0.654}\times\frac{0.654^0}{0!}\right] = 108.15$$

$$E(1) = 208\times P(1) = 208\times\left[e^{-0.654}\times\frac{0.654^1}{1!}\right] = 70.73$$

$$E(2) = 208\times P(2) = 208\times\left[e^{-0.654}\times\frac{0.654^2}{2!}\right] = 23.13$$

$$E(\text{more than } 2) = 208-[108.15+70.73+23.13] = 5.99$$

Number of accidents	0	1	2	2 or more
O	112	56	40	0
E	108.15	70.73	23.13	5.99

$$\text{Test statistic } (\chi^2) = \frac{(112-108.15)^2}{108.15}+\frac{(56-70.73)^2}{70.73}+\frac{(40-23.13)^2}{23.13}+\frac{(0-5.99)^2}{5.99}$$

$$\text{t.s. } (\chi^2) = 21.5$$

Number of classes.

Degrees of freedom $= 4-2 = 2$

Number of restrictions.

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

Critical value $= \chi_2^2(5\%) = 5.991$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence the data cannot be modelled by a Poisson distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 11

Question:

A tensile test is carried out on 100 steel bars which are uniform in section. The distances from the mid-points of the bars at which they fracture is recorded with the following results.

Distance	$0 < d \le 10$	$10 < d \le 20$	$20 < d \le 30$	$30 < d \le 40$	$40 < d \le 50$	$50 < d \le 60$
Frequency	15	17	18	20	12	18

Test at the 0.05 significance level if these data can be modelled by a continuous uniform distribution.

Solution:

H_0: The data can be modelled by a continuous uniform distribution.

H_1: The data cannot be modelled by a continuous uniform distribution.

$$\text{Total frequency} = 100$$

This is 1 restriction (total frequencies agree).

$$\text{Expected frequency for each class} = \frac{100}{6} = 16.667 \quad (\text{or } 16\tfrac{2}{3})$$

$$\therefore \text{ Test statistic } \left(\chi^2\right) = \frac{(15-16.667)^2}{16.667} + \frac{(17-16.667)^2}{16.667} + \cdots + \frac{(18-16.667)^2}{16.667}$$

Values found for χ^2 will vary according to the rounding you do. If possible work to the degree of accuracy of your calculation. Never work to any less than the degree of accuracy in the question (at least one more place is a good idea). A range of answers that encompass sensible degrees of accuracy would be accepted in an examination.

$$\text{t.s. } \left(\chi^2\right) = 2.36$$

$$\text{Degrees of freedom} = 6 - 1 = 5$$

$$\text{Critical value} = \chi_5^2\,(5\%) = 11.070$$

$\chi^2 <$ c.v. so accept H_0.

Conclude no reason to doubt the data could be modelled by a continuous uniform distribution.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 12

Question:

Samples of stones were taken at two points on a beach which were 1 mile apart. The rock types of the stones were found and classified as igneous, sedimentary or other types, with the following results.

		Site	
		A	***B***
Rock type	***Igneous***	30	10
	Sedimentary	55	35
	Other	15	15

Use a 5% significance level to see if the rocks at both sites come from the same population.

Solution:

H_0: There is no association between site and type of rock found.
H_1: There is an association between site and type of rock found.

Observed frequencies

		Site		
		A	*B*	Total
Rock Type	*Igneous*	30	10	40
	Sedimentary	55	35	90
	Other	15	15	30
	Total	100	60	160

Expected frequency '*Igneous*' and '*A*' $= \dfrac{40 \times 100}{160} = 25$

Expected frequencies

	A	*B*
Igneous	25	15
Sedimentary	56.25	33.75
Other	18.75	11.25

Test statistic $\left(\chi^2\right) = \dfrac{(30-25)^2}{25} + \dfrac{(10-15)^2}{15} + \cdots + \dfrac{(15-11.25)^2}{11.25}$

t.s. $\left(\chi^2\right) = 4.74$

Degrees of freedom $= (3-1)\times(2-1)$
$= 2$

Critical value $= \chi_2^2\,(5\%) = 5.991$

$\chi^2 <$ c.v. so accept H_0.

Conclude no evidence of an association between site and type of rock found so the rocks at both sites could come from the same population.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 13

Question:

A small shop sells a particular item at a fairly steady yearly rate. When looking at the weekly sales it was found that the number sold varied. The results for the 50 weeks the shop was open were as shown in the table.

Weekly sales	0	1	2	3	4	5	6	7	8	> 8
Frequency	0	4	7	8	10	6	7	4	4	0

a Find the mean number of sales per week.
b Using a significance level of 5%, test to see if these can be modelled by a Poisson distribution.

Solution:

a mean $= 4.28$

Get this directly from your calculator.

b H_0 : The data can be modelled by Po (4.28).
H_1: The data cannot be modelled by Po (4.28).

This is 1 restriction (total frequencies agree).

Total frequency $= 50$
$\lambda = 4.28$

This is 1 restriction (calculation of estimate of parameter).

$$E(0) = 50 \times P(0) = 50 \times \left[e^{-4.28} \times \frac{4.28^0}{0!} \right] = 0.6921$$

$$E(1) = 50 \times P(1) = 50 \times \left[e^{-4.28} \times \frac{4.28^1}{1!} \right] = 2.9623$$

$$E(2) = 50 \times P(2) = 50 \times \left[e^{-4.28} \times \frac{4.28^2}{2!} \right] = 6.3394$$

Pool these classes to get all the E's to be 5 or more.

Similarly, $E(3) = 9.0442, E(4) = 9.6773, E(5) = 8.2838$
$E(6) = 5.9091, E(7) = 3.6130, E(8) = 1.9329,$
$E(8 \text{ or more}) = 50 - [0.6921 + 2.9623 + \ldots + 1.9329] = 1.5459$

The last 3 classes must be pooled to get all the E's to be 5 or more.

Weekly sales	2 or fewer	3	4	5	6	7 or more
O	11	8	10	6	7	8
E	9.9938	9.0442	9.6773	8.2838	5.9091	7.0918

$$\text{Test statistic } (\chi^2) = \frac{(11-9.9938)^2}{9.9938} + \frac{(8-9.0442)^2}{9.0442} + \cdots + \frac{(8-7.0918)^2}{7.0918}$$

$$\text{t.s. } (\chi^2) = 1.18$$

Number of classes after pooling.

Degrees of freedom $= 6 - 2$
$= 4$

Number of restrictions.

Critical value $= \chi_4^2 \, (5\%) = 9.488$

$\chi^2 <$ c.v. so accept H_0.
Conclude no reason to doubt that Po (4.28) could be a good model for the data.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Goodness of fit and contingency tables
Exercise E, Question 14

Question:

A study was done of how many students in a college were left-handed and how many were right-handed. As well as left- or right-handedness the gender of each person was also recorded with the following results

	Left handed	Right handed
Male	100	600
Female	80	800

Use a significance test at the 0.05 level to see if there is an association between gender and left- and right-handedness.

Solution:

H_0 : There is no association between gender and left- and right-handedness.
H_1: There is an association between gender and left- and right-handedness.

Observed frequencies →

	Left-handed	*Right-handed*	Total
Male	100	600	700
Female	80	800	880
Total	180	1400	1580

$$\text{Expected frequency 'Male' and 'Left-handed'} = \frac{700\times 180}{1580} = 79.747$$

Expected frequencies →

	Left-handed	*Right-handed*
Male	79.747	620.253
Female	100.253	779.747

$$\text{Test statistic } (\chi^2) = \frac{(100-79.747)^2}{79.747}+\frac{(600-620.253)^2}{620.253}+\cdots+\frac{(800-779.747)^2}{779.747}$$
$$= 10.4 \quad (3\text{ s.f.})$$

$$\text{Degrees of freedom} = (2-1)\times(2-1) = 1$$

$$\text{Critical value} = \chi_1^2\,(5\%) = 3.841$$

$\chi^2 >$ c.v. so reject H_0.

Conclude there is evidence of an association between gender and left- and right-handedness in this population.