Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 1

Question:

During a village show, two judges, P and Q, had to award a mark out of 30 to some flower displays. The marks they awarded to a random sample of 8 displays were as follows:

Display	A	B	C	D	E	F	G	H
Judge P	25	19	21	23	28	17	16	20
Judge Q	20	9	21	13	17	14	11	15

a Calculate Spearman's rank correlation coefficient for the marks awarded by the two judges.

After the show, one competitor complained about the judges. She claimed that there was no positive correlation between their marks.

b Stating your hypotheses clearly, test whether or not this sample provides support for the competitor's claim. Use a 5% level of significance. ***E***

Solution:

a

	A	B	C	D	E	F	G	H
P Rank	2	6	4	3	1	7	8	5
Q Rank	2	8	1	6	3	5	7	4
d	0	−2	3	−3	−2	2	1	1
d^2	0	4	9	9	4	4	1	1

Remember to rank the data. It does not matter whether you rank from highest to lowest or vice versa as long as you do the same for both judges.

The sum of your d's should be zero. The d^2 should all be positive.

$\sum d^2 = 32$

$$r_s = 1 - \frac{6 \times 32}{8 \times (8^2 - 1)}$$

Using $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ from the formula book.

$$= \frac{13}{21} \text{ or } 0.619$$

If you give your answer as a decimal it should be given to 3 significant figures.

b $H_0 : \rho = 0 \quad H_1 : \rho > 0$

Make sure your hypotheses are clearly written using the symbol ρ. This is a one-tail test so only interested if positive i.e.

r_s 1-tail 5% critical value is 0.6429

Look up the value under 0.05 in the table for Spearman's. Quote the figure in full.

$0.619 < 0.6429$ so accept H_0 or not significant.
So insufficient evidence of a positive correlation between judges
or
competitor's claim is justified.

Draw a conclusion in the context of the question.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2

Exercise A, Question 2

Question:

The Director of Studies at a large college believed that students' grades in Mathematics were independent of their grades in English. She examined the results of a random group of candidates who had studied both subjects and she recorded the number of candidates in each of the 6 categories shown.

	Mathematics grade A or B	Mathematics grade C or D	Mathematics grade E or U
English grade A or B	25	25	10
English grade C to U	5	30	15

a Stating your hypotheses clearly, test the Director's belief using a 10% level of significance. You must show each step of your working.

The Head of English suggested that the Director was losing accuracy by combining the English grades C to U in one row. He suggested that the Director should split the English grades into two rows, grades C or D and grades E or U as for Mathematics.

b State why this might lead to problems in performing the test. ***E***

Solution:

a H_0 : Mathematics grades are independent of English grades
or
no association between Mathematics grades and English grades.
H_1: Mathematics and English grades are dependent.
or
There is an association between Mathematics grades and English grades.

For a contingency table.
For H_0 you should use the words 'no association' or 'independent'.
For H_1 you should use the words 'is an association' or 'dependent'.

Expected frequencies

	$M_{A,B}$	$M_{C\text{ or }D}$	$M_{E,U}$
$E_{A,B}$	16.364	30	13.636
$E_{C\text{ to }U}$	13.636	25	11.364

Expected frequency 'Maths A or B' and 'English A or B' $\frac{60 \times 30}{110} = 16.364$

Show the working for at least one calculation of an expected value.

Test statistic $= \sum \frac{(O_i - E_i)^2}{E_i}$

$$= \frac{(25-16.364)^2}{16.364} + \frac{(25-30)^2}{25} + \cdots + \frac{(15-11.364)^2}{11.364}$$

t.s. $= 13.994$

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

Critical value $= X_2^2\ (10\%) = 4.605$

t.s. > c.v. since $13.994 > 4.605$
so reject H_0.

Conclude there is evidence of an association between Mathematics and English grades.

b May have some expected frequencies < 5 (and hence need to pool rows/columns).

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 3

Question:

A quality control manager regularly samples 20 items from a production line and records the number of defective items x. The results of 100 such samples are given in Table 1 below.

x	0	1	2	3	4	5	6	7 or more
Frequency	17	31	19	14	9	7	3	0

Table 1

a Estimate the proportion of defective items from the production line.
The manager claimed that the number of defective items in a sample of 20 can be modelled by a binomial distribution. He used the answer in part **a** to calculate the expected frequencies given in Table 2.

x	0	1	2	3	4	5	6	7 or more
Expected frequency	12.2	27.0	r	19.0	s	3.2	0.9	0.2

Table 2

b Find the value of r and the value of s giving your answers to 1 decimal place.
c Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim.
d Explain what the analysis in part **c** tells the manager about the occurrence of defective items from this production line. ***E***

Solution:

a mean number of defectives $= \dfrac{0\times17+1\times31+\cdots+6\times3}{17+31+\cdots+3} = \dfrac{200}{100} = 2$ ← This is 1 restriction in part c.

Total frequency $= 17+31+\cdots+3 = 100$ ← This is 1 restriction in part c.

In binomial distribution with $n = 20$
mean $= np$ so $2 = 20\times p$
$\therefore p = 0.1$

b $r = 100\times\binom{20}{2}(0.1)^2(0.9)^{18}$ ← You could use the tables to work this out $100\times(0.6769-0.3917)$.

$= 28.517$

$= 28.5$ (1 d.p.)

$s = 100-91 = 9.0$ (1 d.p.) ← The total of the expected frequencies is the same as the total of the observed frequencies. Here it is 100.

c H_0: B(20, 0.1) is a good/suitable model/fit
H_1: B(20, 0.1) is *not* a suitable model

x	0	1	2	3	≥ 4
O_i	17	31	19	14	19
E_i	12.2	27.0	28.5	19.0	13.3
$\dfrac{(O-E)^2}{E}$	1.889	0.593	3.167	1.316	2.443

← The classes for 4, 5, 6 and 7 or more have been combined. This is so that all the expected frequencies are greater than 5.

Test statistic $= \sum\dfrac{(O-E)^2}{E} = 9.41$

or

$$\sum\frac{O_i^2}{E_i} - N = \frac{17^2}{12.2}+\frac{31^2}{27}+\ldots+\frac{19^2}{13.3}-100$$

$= 9.41$

← It is often easier to use the formula $\sum\dfrac{O_i^2}{E_i} - N$.

$\nu = 5-2 = 3$ ← Degrees of freedom = (number of cells after pooling) − (restrictions)

critical value $= \chi_3^2\,(5\%) = 7.815$ ← Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

t.s. > c.v. so reject H_0.
(significant result) binomial distribution is not a suitable model

d Defective items do not occur independently **or** not with constant probability. ← Since the binomial does not fit then the laws for a binomial distribution can not be true.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 4

Question:

The table below shows the price of an ice cream and the distance of the shop where it was purchased from a particular tourist attraction.

Shop	Distance from tourist attraction (m)	Price (£)
A	50	1.75
B	175	1.20
C	270	2.00
D	375	1.05
E	425	0.95
F	580	1.25
G	710	0.80
H	790	0.75
I	890	1.00
J	980	0.85

a Find, to 3 decimal places, the Spearman rank correlation coefficient between the distance of the shop from the tourist attraction and the price of an ice cream.

b Stating your hypotheses clearly and using a 5% one-tailed test, interpret your rank correlation coefficient. ***E***

Solution:

Shop	Distance	Price	d	d^2
A	1	9	−8	64
B	2	7	−5	25
C	3	10	−7	49
D	4	6	−2	4
E	5	4	1	1
F	6	8	−2	4
G	7	2	5	25
H	8	1	7	49
I	9	5	4	16
J	10	3	7	49

Remember to rank the data. It does not matter whether you rank from highest to lowest or vice versa as long as you do the same for both distance and price.

The sum of your d's should be zero. The d^2 should all be positive.

$\sum d^2 = 286$

a $r_s = 1 - \frac{6 \times 286}{10(100-1)}$

Using $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ from the formula book.

$= -\frac{11}{15}$ or -0.733

If you give your answer as a decimal it should be given to 3 significant figures.

b $H_0 : \rho = 0$

$H_1 : \rho < 0$

Make sure your hypotheses are clearly written using the parameter ρ.

test statistic $= -0.733$

c.v. $= -0.5636$

t.s. < c.v. since $-0.733 < -0.5636$

Look up the value under 0.05 in the table for Spearman's. Quote the figure in full.

Reject H_0, evidence there is a significant *negative* correlation between the rank of the price of an ice cream and the rank of the distance from a tourist attraction.
i.e. the further from a tourist attraction you travel the less you are likely to pay for an ice cream.

Draw a conclusion in the context of the question.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 5

Question:

Five coins were tossed 100 times and the number of heads recorded. The results are shown in the table below.

Number of heads	0	1	2	3	4	5
Frequency	6	18	29	34	10	3

a Suggest a suitable distribution to model the number of heads when five unbiased coins are tossed.

b Test, at the 10% level of significance, whether or not the five coins are unbiased. State your hypotheses clearly. ***E***

Solution:

a B(5, 0.5)

b H_0 : B(5, 0.5) is a suitable model (good fit)
H_1: B(5, 0.5) is a not a suitable model (not a good fit)

Total frequency $= 6+18+29+34+10+3=100$ ← This is 1 restriction.

Expected value for no heads $= 100\times 0.5^5 = 3.125$

Expected value for 1 head $= 100\times\binom{5}{1}0.5^5 = 15.625$

Expected value for 2 heads $= 100\times\binom{5}{2}0.5^5 = 31.25$

← Show the working for at least one calculation of an expected value.

Expected value for 3 heads $= 100\times\binom{5}{3}0.5^5 = 31.25$

Expected value for 4 heads $= 100\times\binom{5}{4}0.5^5 = 15.625$

← Write down all the expected values.

Expected value for 5 heads $= 100\times\binom{5}{5}0.5^5 = 3.125$

	O	E	$\frac{(O-E)^2}{E}$
0 or 1	24	18.75	1.47
2	29	31.25	0.162
3	34	31.25	0.242
4 or 5	13	18.75	1.76

test statistic $= \sum\frac{(O-E)^2}{E} = 3.64$

or

$$\sum\frac{O_i^2}{E_i} - N = \frac{24^2}{18.75}+\frac{29^2}{31.25}+\frac{34^2}{31.25}+\frac{13^2}{18.75}-100$$
$$= 3.64$$

← It is often easier to use the formula $\sum\frac{O_i^2}{E_i} - N$.

$\nu = 4-1=3$ ← Degrees of freedom = (number of cells after pooling) − 1 since the parameter p is known.

$\chi_3^2(10\%) = 6.251$ ← Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

t.s. < c.v. since $3.64 < 6.251$
Insufficient evidence to reject H_0.
B(5, 0.5) is a suitable model.
No evidence that coins are biased.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 6

Question:

People over the age of 65 are offered an annual flu injection. A health official took a random sample from a list of patients who were over 65. She recorded their gender and whether or not the offer of an annual flu injection was accepted or rejected. The results are summarised below.

Gender	Accepted	Rejected
Male	170	110
Female	280	140

Using a 5% significance level, test whether or not there is an association between gender and acceptance or rejection of an annual flu injection. State your hypotheses clearly. ***E***

Solution:

H_0: No association between gender and acceptance
or
gender is independent of acceptance
H_1: There is an association between gender and acceptance
or
gender is not independent of acceptance

For a contingency table
For H_0 you should use the words 'no association' or 'independent'.
For H_1 you should use the words 'is an association' or 'dependent'.

'male' and 'accepted' expected frequency $= \frac{450 \times 280}{700} = 180$

Show the working for at least one calculation of an expected value.

Expected

Expected (obs)	*Accept*	*Not accept*	Total
Males	180 (170)	100 (110)	280
Females	270 (280)	150 (140)	420
Totals	450	250	700

Write down all the expected values.

O	E	$\frac{(O-E)^2}{E}$
170	180	0.5556
110	100	1.0000
280	270	0.3704
140	150	0.6667

The formula $\sum \frac{(O_i - E_i)^2}{E_i}$ is in the formula book. Write down at least two of the calculations.

$$\sum \frac{(O-E)^2}{E} = 2.59$$

$$\sum \frac{O_i^2}{E_i} - N = \frac{170^2}{180} + \frac{110^2}{100} + \ldots + \frac{140^2}{150} - 700$$

$$= 2.59$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$v = (2-1) \times (2-1) = 1$

Contingency table therefore degrees of freedom $= (c-1)(r-1)$.

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

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

$3.841 > 2.59$. There is insufficient evidence to reject H_0
There is no association between a person's gender and their acceptance of the offer of a flu jab.

Draw a conclusion in the context of the question.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 7

Question:

An area of grass was sampled by placing a $1\text{m} \times 1\text{m}$ square randomly in 100 places. The numbers of daisies in each of the squares were counted. It was decided that the resulting data could be modelled by a Poisson distribution with mean 2. The expected frequencies were calculated using the model.
The following table shows the observed and expected frequencies.

Number of daisies	Observed frequency	Expected frequency
0	8	13.53
1	32	27.07
2	27	r
3	18	s
4	10	9.02
5	3	3.61
6	1	1.20
7	0	0.34
≥ 8	1	t

a Find values for r, s and t.

b Using a 5% significance level, test whether or not this Poisson model is suitable. State your hypotheses clearly.

An alternative test might have been to estimate the population mean by using the data given.

c Explain how this would have affected the test. ***E***

Solution:

a $r = 100 \times (0.6767 - 0.4060) = 27.07$

$s = 100 \times (0.8571 - 0.6767) = 18.04$

$t = 100 - [13.53 + 27.07 + 27.07 + 18.04$
$+ 9.02 + 3.61 + 1.20 + 0.34]$

$t = 0.12$

The total of the observed values is 100.

t is worked out using the total of the expected frequencies is 100.

b H_0: A Poisson model Po(2) is a suitable model.

H_1: A Poisson model Po(2) is *not* a suitable model.

Number of daisies	Observed	Expected
0	8	13.53
1	32	27.07
2	27	27.07
3	18	18.04
4	10	9.02
≥ 5	5	5.27

The classes for 5, 6, 7 and ≥ 8 have been combined. This is so that all the expected frequencies are greater than 5.

$$\sum \frac{(O_i - E_i)^2}{E_i} = \frac{(8 - 13.53)^2}{13.53} + \frac{(32 - 27.07)^2}{27.07} + \ldots + \frac{(5 - 5.27)^2}{5.27}$$

$= 3.28$ (awrt)

The formula $\sum \frac{(O_i - E_i)^2}{E_i}$ is in the formula book. Write down at least two of the calculations.

or

$$\sum \frac{O_i^2}{E_i} - N = \frac{8^2}{13.53} + \frac{32^2}{27.07} + \cdots + \frac{5^2}{5.27} - 100$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$= 3.28$ (awrt)

$\nu = 6 - 1 = 5$

Degrees of freedom = number of cells − 1 since the parameter λ is known.

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

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

$3.28 < 11.070$ There is insufficient evidence to reject H_0.

Po(2) is a suitable model.

c The mean must be calculated and then λ = mean. The expected values, and hence $\sum \frac{(O-E)^2}{E}$ would be different, and the degrees of freedom would be 1 less, also changing the critical value.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 8

Question:

The numbers of deaths from pneumoconiosis and lung cancer in a developing country are given in the table.

Age group (years)	20–29	30–39	40–49	50–59	60–69	70 and over
Deaths from pneumoconiosis (1000s)	12.5	5.9	18.5	19.4	31.2	31.0
Deaths from lung cancer (1000s)	3.7	9.0	10.2	19.0	13.0	18.0

The correlation between the number of deaths in the different age groups for each disease is to be investigated.

a Give **one** reason why Spearman's rank correlation coefficient should be used.

b Calculate Spearman's rank correlation coefficient for these data.

c Use a suitable test, at the 5% significance level, to interpret your result. State your hypotheses clearly. ***E***

Solution:

a The variables cannot be assumed to be normally distributed.

b

	20–29	30–39	40–49	50–59	60–69	70+
Rank x	5	6	4	3	1	2
Rank y	6	5	4	1	3	2
d	−1	1	0	2	−2	0
d^2	1	1	0	4	4	0

Remember to rank the data. It does not matter whether you rank from highest to lowest or vice versa as long as you do the same for both.

The sum of your d's should be zero. The d^2 should all be positive.

$\sum d^2 = 10$

$r_s = 1 - \frac{6\times 10}{6(36-1)}$

Using $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ from the formula book.

$= \frac{5}{7}$ or 0.714

If you give your answer as a decimal it should be given to 3 significant figures.

c $H_0 : \rho = 0$

$H_1 : \rho \neq 0$ (or $\rho_s > 0$)

Make sure your hypotheses are clearly written using the parameter ρ.

$n = 6 \Rightarrow 5\%$ critical value $= 0.8857$ (or 0.8286)

$0.714 < 0.8857$ (or $0.714 < 0.8826$)

Look up the value under 0.05 in the table for Spearman's. Quote the figure in full.

No evidence to reject H_0. No evidence of a positive correlation between the rates of deaths from pneumoconiosis and lung cancer.

Draw a conclusion in the context of the question.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 9

Question:

Students in a mixed sixth form college are classified as taking courses in either arts, science or humanities. A random sample of students from the college gave the following results.

		Course		
		Arts	Science	Humanities
Gender	Boy	30	50	35
	Girl	40	20	42

Showing your working clearly, test, at the 1% level of significance, whether or not there is an association between gender and the type of course taken. State your hypotheses clearly. ***E***

Solution:

H_0: There is no association between course and gender or course is independent of gender
H_1: There is an association between course and gender or course is dependent on gender

For a contingency table
For H_0 you should use the words 'no association' or 'independent'.
For H_1 you should use the words 'is an association' or 'dependent'.

	Arts	*Science*	*Hums*	Total
Boy	30	50	35	115
Girl	40	20	42	102
Total	70	70	77	217

Expected frequency 'boy' and 'arts'
$= \frac{115 \times 70}{217} = 37.0967\ldots$

Show the working for at least one calculation of an expected value.

Expected (Obs)	A	S	H
Boy	37.1 (30)	37.1 (50)	40.8 (35)
Girl	32.9 (40)	32.9 (20)	36.2 (42)

Write down all the expected values.

$$\sum \frac{(O-E)^2}{E} = \frac{(30-37.1)^2}{37.1} + \frac{(40-32.9)^2}{32.9} + \ldots + \frac{(42-36.2)^2}{36.2}$$
$$= 1.358 + 4.485 + 0.824 + 1.532 + 5.058 + 0.929 = 14.18$$

The formula $\sum \frac{(O_i - E_i)^2}{E_i}$ is in the formula book. Write down at least two of the calculations.

$$\left[\text{or } \sum \frac{O^2}{E} - N = \frac{30^2}{37.1} + \frac{40^2}{32.9} + \ldots + \frac{42^2}{36.2} - 217 \right]$$
$$= 14.2 \text{ (3 s.f.)}$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$v = (3-1)(2-1) = 2$

Contingency table therefore degrees of freedom $= (c-1)(r-1)$.

χ^2_2 (1%) critical value is 9.210

Look up the value under 0.01 in the percentage points of the χ^2 distribution. Quote the figure in full.

$14.18 > 9.210$
Significant result or reject null hypothesis.

There is evidence of an association between course taken and gender.

Draw a conclusion in the context of the question.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 10

Question:

The product-moment correlation coefficient is denoted by r and Spearman's rank correlation coefficient is denoted by r_s .

a Sketch separate scatter diagrams, with five points on each diagram, to show
i $r = 1$,
ii $r_s = -1$ but $r > -1$.

Two judges rank seven collie dogs in a competition. The collie dogs are labelled A to G and the rankings are as follows.

Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Judge 1	A	C	D	B	E	F	G
Judge 2	A	B	D	C	E	G	F

b **i** Calculate Spearman's rank correlation coefficient for these data.
ii Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the judges are generally in agreement. ***E***

Solution:

a i

ii

b i

r_1	1	4	2	3	5	6	7
r_2	1	2	4	3	5	7	6
d	0	2	−2	0	0	−1	1
d^2	0	4	4	0	0	1	1

Although the data is ranked it is easiest to rewrite it in a familiar form.

The sum of your d's should be zero. The d^2 should all be positive.

$\sum d^2 = 10$

$r_s = 1 - \frac{6 \times 10}{7(49-1)}$

$= \frac{23}{28}$ or 0.821 (3 s.f.)

Using $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$ from the formula book.

If you give your answer as a decimal it should be given to 3 significant figures.

ii $H_0 : \rho_s = 0 \quad H_1 : \rho_s > 0$

Make sure your hypotheses are clearly written using the parameter ρ_s. This time you are testing if in agreement therefore you are testing if positively correlated.

test statistic $= r_s = 0.821$

critical value is 0.7143

Look up the value under 0.05 in the table for Spearman's. Quote the figure in full.

$0.821 > 0.7143$ so significant result or reject null hypothesis.

There is evidence of a (positive) correlation between the ranks awarded by the judges *or* the judges agree.

Draw a conclusion in the context of the question.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 11

Question:

Ten cuttings were taken from each of 100 randomly selected garden plants. The numbers of cuttings that did not grow were recorded.
The results are as follows.

Number of cuttings which did not grow	0	1	2	3	4	5	6	7	8, 9 or 10
Frequency	11	21	30	20	12	3	2	1	0

a Show that the probability of a randomly selected cutting, from this sample, not growing is 0.223.

A gardener believes that a binomial distribution might provide a good model for the number of cuttings, out of 10, that do not grow.
He uses a binomial distribution, with the probability 0.2 of a cutting not growing. The calculated expected frequencies are as follows.

Number of cuttings which did not grow	0	1	2	3	4	5 or more
Expected frequency	r	26.84	s	20.13	8.81	t

b Find the values of r, s and t.

c State clearly the hypotheses required to test whether or not this binomial distribution is a suitable model for these data.

The test statistic for the test is 4.17 and the number of degrees of freedom used is 4.

d Explain fully why there are 4 degrees of freedom.

e Stating clearly the critical value used, carry out the test using a 5% level of significance. ***E***

Solution:

a $\text{mean} = \dfrac{0\times 11+1\times 21+2\times 30+\ldots+7\times 1}{11+21+30+\ldots+1}$

$\text{mean} = \dfrac{223}{100} = 2.23$

In binomial, $n = 10$, mean $= np$, $2.23 = 10 \times p$

so $p = 0.223$

b $r = (0.8)^{10} \times 100 = 10.7374 = 10.74$ (2 d.p.)

You could use the tables to work these out. But you will need to use $p = 0.2$ so it is easier to do them this way.

$s = \binom{10}{2}(0.8)^8 \times (0.2)^2 \times 100 = 30.198\ldots$

$= 30.20$ (2 d.p.)

$t = 100 - [r + s + 26.84 + 20.13 + 8.81]$

$= 3.28$

The total of the expected frequencies is the same as the total of the observed frequencies. Here it is 100.

c H_0: B(10, 0.2) is a suitable model for these data.

H_1: B(10, 0.2) is *not* a suitable model for these data.

d Since $t < 5$, the last two groups are combined and $\nu = 5 - 1 = 4$

Since there are then 5 cells and the parameter p is given

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

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

$4.17 < 9.488$ so not significant or do not reject null hypothesis.

The binomial distribution with $p = 0.2$ is a suitable model for the number of cuttings that do not grow.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 12

Question:

A researcher carried out a survey of three treatments for a fruit tree disease.

	No action	Remove diseased branches	Spray with chemicals
Tree died within 1 year	10	5	6
Tree survived for 1–4 years	5	9	7
Tree survived beyond 4 years	5	6	7

Test, at the 5% level of significance, whether or not there is any association between the treatment of the trees and their survival. State your hypotheses and conclusion clearly.

Solution:

H_0: There is no association between treatment and survival *or* treatment is independent of survival
H_1: There is association between treatment and survival *or* treatment is dependent on survival

For a contingency table
For H_0 you should use the words 'no association' or 'independent'.
For H_1 you should use the words 'is an association' or 'dependent'.

No action and tree died within 1 year
expected frequency $= \frac{20 \times 21}{60} = 7$

Show the working for at least one calculation of an expected value.

Expected (Obs)	No action	Remove diseased branches	Spray with chemicals	Totals
Tree died within 1 year	7(10)	7(5)	7(6)	21
Survived 1–4 years	7(5)	7(9)	7(7)	21
Survived > 4 years	6(5)	6(6)	6(7)	18
Totals	20	20	20	60

Write down all the expected values

$$\sum \frac{(O-E)^2}{E} = \frac{9}{7} + \frac{4}{7} + \frac{1}{7} + \frac{4}{7} + \frac{4}{7} + 0 + \frac{1}{6} + 0 + \frac{1}{6}$$
$$= 3.4761\ldots$$

The formula $\sum \frac{(O_i - E_i)^2}{E_i}$ is in the formula book. Write down at least two of the calculations.

or

$$\sum \frac{O_i^2}{E_i} - N = \frac{10^2}{7} + \frac{5^2}{7} + \ldots + \frac{6^2}{7} - 60$$
$$= 3.47619\ldots$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$v = (3-1) \times (3-1) = 4$

Contingency table therefore degrees of freedom $= (c-1)(r-1)$.

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

or CR: $\chi^2 > 9.488$

$3.47619 < 9.488$

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

(or since 3.47619... is *not* in the critical region (i.e. < 9.488) there is insufficient evidence to reject H_0.

Draw a conclusion in the context of the question.

There is no evidence of association between treatment and length of survival.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 13

Question:

Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. The results are shown in the table.

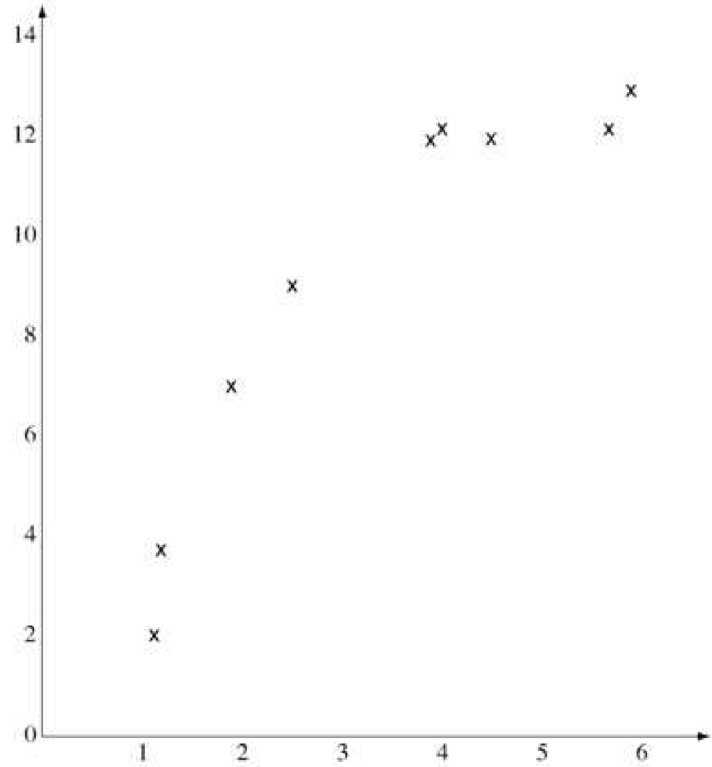
s	1.2	1.9	3.2	3.9	2.5	4.5	5.7	4.0	1.1	5.9
d	3.8	7.0	11.0	12.0	9.0	12.0	13.5	12.2	2.0	13.9

[Use $\sum s^2 = 141.51$, $\sum d^2 = 1081.74$ and $\sum sd = 386.32$]

a Draw a scatter diagram to represent these data.
b State what is measured by the product-moment correlation coefficient.
c Calculate S_{ss}, S_{dd} and S_{sd}.
d Calculate the value of the product-moment correlation coefficient r between s and d.
e Stating your hypotheses clearly, test, at the 1% significance level, whether or not the correlation coefficient is greater than zero.
f With reference to your scatter diagram, comment on your result in part **e**. ***E***

Solution:

a



b The strength of the linear link between two variables.

c $S_{ss} = 141.51 - \frac{33.9^2}{10} = 26.589; S_{dd} = 152.444; S_{sd} = 59.524$

d $r = \frac{59.524}{\sqrt{152.444 \times 26.589}}$
$= 0.93494\ldots$

The formulae for these are under S1 in the formula book.

Make sure your hypotheses are clearly written using the parameter ρ.

e $H_0 : \rho = 0; H_1 : \rho > 0$
test statistic $= r = 0.935$
Critical value at $1\% = 0.7155$
$0.935 > 0.7155$

Look up the value under 0.01 in the table for product-moment coefficient. Quote the figure in full.

so reject H_0 : levels of serum and disease are positively correlated.

Draw a conclusion in the context of the question.

f Linear correlation significant but scatter diagram looks non-linear.

The product-moment correlation coefficient should not be used here since the association/relationship is not linear.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2

Exercise A, Question 14

Question:

The number of times per day a computer fails and has to be restarted is recorded for 200 days. The results are summarised in the table.

Number of restarts	Frequency
0	99
1	65
2	22
3	12
4	2

Test whether or not a Poisson model is suitable to represent the number of restarts per day. Use a 5% level of significance and state your hypothesis clearly. ***E***

Solution:

H_0: Poisson distribution is a suitable model

H_1: Poisson distribution is not a suitable model

$$\hat{\lambda} = \frac{(0\times 99)+(1\times 65)+\ldots+(4\times 2)}{200} = \frac{153}{200} = 0.765$$

As λ is not given you must work it out.

Total frequency = 200

This is 1 restriction.

Expected frequency for $(X=2) = \frac{0.765^2\, e^{-0.765}}{2} \times 200$
$= 27.23250$

Show the working for at least one calculation of an expected value.

Number of restarts gives

X	Observed frequency	Expected frequency	
0	99	93.06678...	
1	65	71.19604...	
2	22	27.23250...	
3	12 } 14	6.94428... }	8.50468
≥ 4	2	1.56040...	

Write down all the expected values.

Combine the classes for 3 and ≥ 4. This is so that all the expected frequencies are greater than 5.

$$\sum \frac{(O_i - E_i)^2}{E_i} = \frac{(99-93.07\ldots)^2}{93.066\ldots} + \ldots + \frac{(14-8.505)^2}{8.505}$$
$$= 5.47$$

or

$$\sum \frac{O_i^2}{E_i} - N = \frac{99^2}{93.066\ldots} + \ldots + \frac{14^2}{8.505} - 200$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$\nu = 4 - 2 = 2$

Degrees of freedom = number of cells − 2 since the parameter λ is unknown. It had to be estimated from the information given.

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

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

test statistic < c.v. so 5.47 is not in the critical region so accept H_0.

Number of computer failures per day can be modelled by a Poisson distribution.

Solutionbank S3

Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 15

Question:

A research worker studying colour preference and the age of a random sample of 50 children obtained the results shown below.

Age in years	Red	Blue	Totals
4	12	6	18
8	10	7	17
12	6	9	15
Totals	28	22	50

Using a 5% significance level, carry out a test to decide whether or not there is an association between age and colour preference. State your hypotheses clearly. ***E***

Solution:

H_0: No association between age and colour preference (they are independent)
H_1: Association between age and colour preference (they are not independent)

For a contingency table
For H_0 you should use the words 'no association' or 'independent'.
For H_1 you should use the words 'is an association' or 'dependent'.

'4' and 'red'

expected frequency $y = \frac{18 \times 28}{50} = 10.08$

Show the working for at least one calculation of an expected value.

O	E	$\frac{(O-E)^2}{E}$
12	10.08	0.3657...
6	7.92	0.4654...
10	9.52	0.0242...
7	7.48	0.0308...
6	8.4	0.6857...
9	6.6	0.8727...

Write down all the expected values.

test statistic $= \sum \frac{(O_i - E_i)^2}{E_i} = 2.4446..$

or

$$\sum \frac{O_i^2}{E_i} - N = \frac{12^2}{10.08} + .. + \frac{9^2}{6.6\ldots} - 500$$
$$= 2.4446\ldots$$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$v = (3-1) \times (2-1) = 2$

Contingency table therefore degrees of freedom $= (c-1)(r-1)$.

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

(or CR: $\chi^2 > 5.991$)

Look up the value under 0.05 in the percentage points of the χ^2 distribution. Quote the figure in full.

$2.4446 < 5.991$
so insufficient evidence to reject H_0.
No association between age and colour preference.

Solutionbank S3
Edexcel AS and A Level Modular Mathematics

Review Exercise 2
Exercise A, Question 16

Question:

A manufacture claims that the batteries used in his mobile phones have a mean lifetime of 360 hours and a standard deviation of 20 hours, when the phone is left on standby. To test this claim 100 phones were left on standby until the batteries ran flat. The lifetime t hours of the batteries was recorded.
The results are as follows.

t	300–	320–	340–	350–	360–	370–	380–	400–
Frequency	1	9	28	20	16	18	7	1

A researcher believes that a normal distribution might provide a good model for the lifetime of the batteries
She calculated the expected frequencies as follows using the distribution $N \sim (360, 20)$.

t	< 320	320–	340–	355–	365–	370–	380–	400–
Expected frequency	2.28	13.59	24.26	r	s	14.98	13.59	2.28

a Find the values of r and s.
b Stating clearly your hypotheses, test, at the 1% level of significance, whether or not this normal distribution is a suitable model for these data.

Solution:

a $P(355 < T < 365) = P\left(z < \frac{365-360}{20}\right) - P\left(z < \frac{355-360}{20}\right)$

$= P(z < 0.25) - P(z < -0.25)$

$= 0.5987 - (1 - 0.5987)$

$= 0.1974$

Using $z = \frac{x-\mu}{\sigma}$.

$r = 0.1974 \times 100$

$= 19.74$

$s = 100 - 2.28 - 13.59 - 24.26 - 19.74 - 14.98 - 13.59 - 2.28$

$= 9.28$

You could use the normal distribution to work out the expected value. This is quicker.

b H_0 : N ~ (360, 20) is a suitable model.

H_1 : N ~ (360, 20) is not a suitable model.

t	< 340	340–	355–	365–	370–	380–
Observed frequency	10	28	20	16	18	8
Expected frequency	15.87	24.26	19.74	9.28	14.98	15.87

test statistic $= \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(10-15.87)^2}{15.87} + \ldots + \frac{(8-15.87)^2}{15.87}$

$= 12.13$

or

$\sum \frac{O_i^2}{E_i} - N = \frac{10^2}{15.87} + .. + \frac{8^2}{15.87} - 100$

It is often easier to use the formula $\sum \frac{O_i^2}{E_i} - N$.

$v = 6 - 1 = 5$

Degrees of freedom = number of cells − 1 since μ and σ are given.

Critical value $\chi_5^2(1\%) = 15.086$

Look up the value under 0.01 in the percentage points of the χ^2 distribution. Quote the figure in full.

$12.13 < 15.086$ so accept H_0.

The distribution can be modelled by a N ~ (360, 20).