



GuideMyGrades

— YOUR GOALS. OUR EXPERTISE. —

EDEXCEL INTERNATIONAL A LEVEL MATHS: QUESTIONS BY TOPIC

Goodness of Fit & Contingency Tables (S3.4)

International A Level Mathematics · S3: Statistics 3 (WST03)

20 questions · 219 marks · Mark schemes included

Past papers October 2020 - June 2025 · 1 sub-topic

 guidemygrades.com

 @Guidemygrades

 @Guidemygrades

 @Guidemygrades

 info@guidemygrades.com

 Piccadilly Business Centre, Unit C, Aldow Enterprise Park
Blackett Street, Manchester, M12 6AE

HOW TO USE THIS BOOKLET

Questions are grouped by sub-topic and run from the lowest to the highest mark tariff, then oldest paper first.

Click any row below to jump straight to that sub-topic or question.

On every question page, click ANSWER in the top-right corner to jump to its mark scheme; < QUESTION brings you back.

A gold bar in the left margin marks exactly where the question sits on the original page.

Diagrams, graphs, grids and tables are printed exactly as they appear in the original papers.

Each question is filed under its main topic; longer questions may also test skills from other topics.

QUESTION	SOURCE	MARKS	PAGE
S3.4a GOODNESS OF FIT & CONTINGENCY TABLES (-2)		20 Qs · 219 marks	2
1. Q2	IAL S3 (WST03) January 2025	7	3
2. Q2	IAL S3 (WST03) October 2021	8	7
3. Q1	IAL S3 (WST03) January 2024	8	9
4. Q2	IAL S3 (WST03) October 2020	9	13
5. Q3	IAL S3 (WST03) January 2023	9	17
6. Q3	IAL S3 (WST03) June 2025	9	21
7. Q3	IAL S3 (WST03) January 2021	10	25
8. Q4	IAL S3 (WST03) (unused paper) January 2022	10	29
9. Q2	IAL S3 (WST03) June 2023	10	33
10. Q4	IAL S3 (WST03) January 2024	10	37
11. Q4	IAL S3 (WST03) October 2021	11	41
12. Q7	IAL S3 (WST03) June 2022	11	45
13. Q4	IAL S3 (WST03) June 2023	11	50
14. Q4	IAL S3 (WST03) June 2024	11	54
15. Q5	IAL S3 (WST03) June 2025	11	58
16. Q6	IAL S3 (WST03) (unused paper) January 2022	12	62
17. Q6	IAL S3 (WST03) January 2022	14	66
18. Q4	IAL S3 (WST03) January 2023	14	69
19. Q5	IAL S3 (WST03) June 2021	16	73
20. Q5	IAL S3 (WST03) January 2021	18	77



QUESTIONS

20 questions · 219 marks · 1 sub-topic

SUB-TOPIC S3.4a

GOODNESS OF FIT & CONTINGENCY TABLES (-2)

Sub-topic S3.4a · 20 questions · 219 marks · lowest tariff first



2. The manager of a shop investigates the number of customers served by staff that work in the shop.

The manager records the number of customers being served in 100 consecutive five-minute time intervals.

The results are shown in the table below.

Number of customers	0	1	2	3	4	5	6 or more
Frequency	5	38	32	17	7	1	0

The manager plans to test whether or not a Poisson model is suitable for these data.

- (a) Show that the sample mean number of customers served is 1.86

(1)

The manager uses the sample mean to calculate the following expected frequencies, correct to 3 decimal places.

Number of customers	0	1	2	3	4	5	6 or more
Expected frequency	15.567	28.955	26.928	16.696	7.763	2.888	r

- (b) Explain why the value of r is 1.203

(1)

The manager combines cells 4, 5 and '6 or more', before continuing the test.

- (c) Explain why it is necessary for the manager to do this.

(1)

After combining these cells, the manager calculates that $\sum \frac{(O - E)^2}{E} = 12.2$ correct to 3 significant figures.

- (d) Carry out the manager's test at the 1% significance level.

You should clearly state the hypotheses, the degrees of freedom and the critical value used.

(4)







Question 2 continued

Lined area for writing the answer to Question 2.

(Total for Question 2 is 7 marks)

Question 2 continued

Leave
blank

Lined area for writing the answer to Question 2.

(Total 8 marks)

Q2

Mark box for Question 2.



Question 1 continued

Lined area for writing the answer to Question 1 continued.



Question 1 continued

Lined area for writing the answer to Question 1 continued.



Question 1 continued

Lined area for writing the answer to Question 1.

(Total for Question 1 is 8 marks)



Leave
blank

Question 2 continued

Leave
blank



Question 2 continued

Leave
blank

(Total 9 marks)

Q2



This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Question 3 continued

Lined area for writing the answer to Question 3 continued.



(Total for Question 3 is 9 marks)

Question 3 continued

Lined area for writing the answer to Question 3 continued.





Question 3 continued

(Total for Question 3 is 9 marks)



Leave
blank

3. The students in a group of schools can choose a club to join. There are 4 clubs available: Music, Art, Sports and Computers. The director collected information about the number of students in each club, using a random sample of 88 students from across the schools. The results are given in Table 1 below.

	Music	Art	Sports	Computers
No. of students	14	28	27	19

Table 1

The director uses a chi-squared test to determine whether or not the students are uniformly distributed across the 4 clubs.

- (a) (i) Find the expected frequencies he should use.

Given that the test statistic he calculated was 6.09 (to 3 significant figures)

- (ii) use a 5% level of significance to complete the test. You should state the degrees of freedom and the critical value used.

(4)

The director wishes to examine the situation in more detail and takes a second random sample of 88 students. The director assumes that within each school, students select their clubs independently. The students come from 3 schools and the distribution of the students from each school amongst the clubs is given in Table 2 below.

School \ Club	Music	Art	Sports	Computers
School A	3	10	9	8
School B	1	11	13	5
School C	11	6	7	4

Table 2

The director wishes to test for an association between a student's school and the club they choose.

- (b) State hypotheses suitable for such a test.

(1)

- (c) Calculate the expected frequency for School C and the Computers club.

(1)

The director calculates the test statistic to be 7.29 (to 3 significant figures) with 4 degrees of freedom.

- (d) Explain clearly why his test has 4 degrees of freedom.

(2)

- (e) Complete the test using a 5% level of significance and stating clearly your critical value.

(2)



Question 3 continued

Leave
blank

Question 3 continued

Leave
blank



Question 3 continued

Leave
blank

(Total 10 marks)

Q3

-
-
-
4. A manager at a large estate agency believes that the type of property affects the time taken to sell it.

A random sample of 125 properties sold is shown in the table.

	Type of property			Total
	Bungalow	Flat	House	
Sold within three months	7	29	46	82
Sold in more than three months	9	19	15	43
Total	16	48	61	125

Test, at the 5% level of significance, whether there is evidence for an association between the type of property and the time taken to sell it.

You should state your hypotheses, expected frequencies, test statistic and the critical value used for this test.

(10)

Question 4 continued

Leave
blank



Question 4 continued

Leave
blank

Question 4 continued

Leave
blank

(Total 10 marks)

Q4



2. A business accepts cash, bank cards or mobile apps as payment methods.

The manager wishes to test whether or not there is an association between the payment amount and the payment method used.

The manager takes a random sample of 240 payments and records the payment amount and the payment method used.

The manager's results are shown in the table.

		Payment amount		
		Under £50	£50 to £150	Over £150
Payment method	Cash	23	19	18
	Bank card	21	32	31
	Mobile app	16	39	41

Using these results,

- (a) calculate the expected frequencies for the payment amount under £50 that

- (i) use cash
- (ii) use a bank card
- (iii) use a mobile app

(3)

Given that for the other 6 classes $\sum \frac{(O - E)^2}{E} = 2.4048$ to 4 decimal places,

- (b) test, at the 5% level of significance, whether or not there is evidence for an association between the payment amount and the payment method used.
You should state the hypotheses, the test statistic, the degrees of freedom and the critical value used for this test.

(7)





Question 2 continued

Lined area for writing the answer to Question 2 continued.



(Total for Question 2 is 10 marks)



(Total for Question 4 is 10 marks)

Leave
blank

4. A local village radio station, *LSB*, decides to survey adults in its broadcasting area about the programmes it produces.

LSB broadcasts to 4 villages A, B, C and D.

The number of households in each of the villages is given below.

Village	Number of households
A	41
B	164
C	123
D	82

LSB decides to take a stratified sample of 200 households.

- (a) Explain how to select the households for this stratified sample.

(3)

One of the questions in the survey related to the age group of each member of the household and whether they listen to *LSB*. The data received are shown below.

	Age group		
	18–49	50–69	Older than 69
Listen to <i>LSB</i>	130	162	65
Do not listen to <i>LSB</i>	78	98	62

The data are to be used to determine whether or not there is an association between the age group and whether they listen to *LSB*.

- (b) Calculate the expected frequencies for the age group 50–69 that

(i) listen to *LSB*

(ii) do not listen to *LSB*

(2)

Given that for the **other 4** classes $\sum \frac{(O - E)^2}{E} = 4.657$ to 3 decimal places,

- (c) test at the 5% level of significance, whether or not there is evidence of an association between age and listening to *LSB*. Show your working clearly, stating the degrees of freedom and the critical value.

(6)



Question 4 continued

Leave
blank



Question 4 continued

Leave
blank

Question 4 continued

Leave
blank

(Total 11 marks)

Q4







Question 7 continued

(Total for Question 7 is 11 marks)

TOTAL FOR PAPER: 75 MARKS

Question 4 continued

Lined area for writing the answer to Question 4 continued.



(Total for Question 4 is 11 marks)





(Total for Question 4 is 11 marks)





(Total for Question 3 is 11 marks)



Leave
blank

6. The number of emails per hour received by a helpdesk were recorded. The results for a random sample of 80 one-hour periods are shown in the table.

Number of emails per hour	0	1	2	3	4	5	6
Frequencies	1	10	23	15	19	9	3

- (a) Show that the mean number of emails per hour in the sample is 3

(1)

The manager believes that the number of emails per hour received could be modelled by a Poisson distribution.

The following table shows some of the expected frequencies.

Number of emails per hour	Expected Frequencies
0	r
1	11.949
2	17.923
3	17.923
4	13.443
5	s
≥ 6	t

- (b) Find the values of r , s and t , giving your answers to 3 decimal places.

(4)

- (c) Using a 10% significance level, test whether or not a Poisson model is reasonable. You should clearly state your hypotheses, test statistic and the critical value used.

(7)



Question 6 continued

Leave
blank



Question 6 continued

Leave
blank



Question 6 continued

Leave
blank

(Total 12 marks)

Q6



6. A farmer sells strawberries in baskets. The contents of each of 100 randomly selected baskets were weighed and the results, given to the nearest gram, are shown below.

Weight of strawberries (grams)	Number of baskets
302 – 303	5
304 – 305	13
306 – 307	10
308 – 309	18
310 – 311	25
312 – 313	20
314 – 315	5
316 – 317	4

The farmer proposes that the weight of strawberries per basket, in grams, should be modelled by a normal distribution with a mean of 310 g and standard deviation 4 g.

Using his model, the farmer obtains the following expected frequencies.

Weight of strawberries (s , grams)	Expected frequency
$s \leq 303.5$	a
$303.5 < s \leq 305.5$	7.8
$305.5 < s \leq 307.5$	13.6
$307.5 < s \leq 309.5$	18.4
$309.5 < s \leq 311.5$	19.6
$311.5 < s \leq 313.5$	16.3
$313.5 < s \leq 315.5$	10.6
$s > 315.5$	b

- (a) Find the value of a and the value of b . Give your answers correct to one decimal place.

(5)

Question 6 continues on page 23



TOTAL FOR PAPER: 75 MARKS

END



This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.



This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

(Total for Question 4 is 14 marks)



Question 5 continued

Question 5 continued

(Total for Question 5 is 16 marks)



Leave
blank

Question 5 continued

Leave
blank

Question 5 continued

Leave
blank

(Total 18 marks)

Q5



MARK SCHEMES

official mark schemes, in the same order as the questions



Question Number	Scheme		Marks
2 (a)	$\frac{[0 \times 5] + 1 \times 38 + 2 \times 32 + 3 \times 17 + 4 \times 7 + 5 \times 1}{100} [= 1.86]^*$		B1*
			(1)
(b)	$[r = 1.203]$ because total expected frequency must equal 100		B1
			(1)
(c)	[The manager needed to do this] to ensure that [all] expected frequencies were greater than 5		B1
			(1)
(d)	H_0 : Poisson (distribution) is [a] suitable/ sensible (model)		B1
	H_1 : Poisson (distribution) is not [a] suitable/ sensible (model)		
	$\nu = [5 - 1 - 1] = 3$		B1
	$\chi^2_3(0.01) = 11.345 \Rightarrow \text{CR: } X^2 \dots 11.345$		M1
	[Lies in the CR/Reject H_0] Sufficient evidence to say that Poisson is not a suitable model		A1ft
			(4)
	Notes		Total 7
(a)	B1*	For a correct method to show the mean is 1.86 (Ignore use of 6×0) Allow $\frac{[0] + 38 + 64 + 51 + 28 + 5}{100}$	
(b)	B1	A correct explanation referring to the fact that total/sum expected frequency/ E_i must equal total observed frequency e.g. $100 - (15.567 + 28.955 + 26.928 + 16.696 + 7.763 + 2.888) = r$	
(c)	B1	A correct explanation referring to the fact that [all] E_i / expected frequencies/values need to be greater than 5 e.g because expected 5 customers and [expected] 6 or more customers are both less than 5 Allow $2.88 < 5$ and $1.203/r < 5$ or $4.091 < 5$	
(d)	B1	Both hypotheses correct. Must mention Poisson/Po at least once.	
	B1	$\nu = 3$ This mark can be implied by a correct critical value of 11.345 if no DoF given	
	M1	For 11.345 or ft their degrees of freedom $\left[\chi^2_4(0.01) = 13.277 \right]$	
	A1ft	A correct conclusion based on their χ^2 critical value. Must mention Poisson	

Question Number	Scheme						Marks
2	H ₀ : Potassium has no effect on the quality of apple H ₁ : Potassium has an effect on the quality of apple						B1
	Grade	A	B	C	D	E	
	Expected values	9.6	67.2	124.8	24.0	14.4	M1A1
	$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(9-9.6)^2}{9.6} + \dots + \frac{(3-14.4)^2}{14.4} \text{ or}$ $\chi^2 = \sum \frac{O^2}{E} - N = \frac{9^2}{9.6} + \dots + \frac{3^2}{14.4} - 240$						M1
	= 10.657...						awrt 10.7 A1
	Degrees of freedom = 4						B1
	$\chi^2_{4,0.05} = 9.488$						B1ft
	[Reject H ₀] Data suggests that potassium may affect the distribution of the grades of apples or there is evidence that Andy's belief is incorrect						A1ft
							(8)
	Notes						Total 8
	B1	Both hypotheses in context. May use other wording eg The grading of apples remains the same.					
	M1	A correct method to calculate expected values eg 0.04×240					
	A1	At least 3 expected values correct					
	M1	A correct method using their expected values to calculate χ^2 At least one correct, ft their expected values with an intention to add					
	A1	awrt 10.7					
	B1	Degrees of freedom = 4 (may be implied by 9.488)					
	B1ft	9.488 ft their DoF. If no DoF stated then this must be correct for their working.					
	A1ft	ft their χ^2 value provided the 2 nd M1 is awarded and CV. If no hypotheses or hypotheses wrong way round do not award. Must include the word 'apples' or 'belief' oe					

Question Number	Scheme	Marks																																
1.	H_0 : There is no association between treatment and presence of fungus H_1 : There is association between treatment and presence of fungus	B1																																
	<table><tr><th>Expected</th><th>No treatment</th><th>Sulphur</th><th>Copper sulphate</th></tr><tr><td>No Fungus</td><td>$\frac{123 \times 30}{150} [=24.6]$</td><td>$\frac{123 \times 63}{150} [=51.66]$</td><td>$\frac{123 \times 57}{150} [=46.74]$</td></tr><tr><td>Fungus</td><td>$\frac{27 \times 30}{150} [=5.4]$</td><td>$\frac{27 \times 63}{150} [=11.34]$</td><td>$\frac{27 \times 57}{150} [=10.26]$</td></tr></table>	Expected	No treatment	Sulphur	Copper sulphate	No Fungus	$\frac{123 \times 30}{150} [=24.6]$	$\frac{123 \times 63}{150} [=51.66]$	$\frac{123 \times 57}{150} [=46.74]$	Fungus	$\frac{27 \times 30}{150} [=5.4]$	$\frac{27 \times 63}{150} [=11.34]$	$\frac{27 \times 57}{150} [=10.26]$	M1																				
	Expected	No treatment	Sulphur	Copper sulphate																														
	No Fungus	$\frac{123 \times 30}{150} [=24.6]$	$\frac{123 \times 63}{150} [=51.66]$	$\frac{123 \times 57}{150} [=46.74]$																														
	Fungus	$\frac{27 \times 30}{150} [=5.4]$	$\frac{27 \times 63}{150} [=11.34]$	$\frac{27 \times 57}{150} [=10.26]$																														
	<table><tr><th>Observed</th><th>Expected</th><th>$\frac{(O - E)^2}{E}$</th><th>$\frac{O^2}{E}$</th></tr><tr><td>20</td><td>24.6</td><td>0.86016...</td><td>16.2601...</td></tr><tr><td>55</td><td>51.66</td><td>0.21594...</td><td>58.5559...</td></tr><tr><td>48</td><td>46.74</td><td>0.03396...</td><td>49.2939...</td></tr><tr><td>10</td><td>5.4</td><td>3.91851...</td><td>18.5185...</td></tr><tr><td>8</td><td>11.34</td><td>0.98373...</td><td>5.6437...</td></tr><tr><td>9</td><td>10.26</td><td>0.15473...</td><td>7.8947...</td></tr><tr><td colspan="2">Totals:</td><td>6.167...</td><td>156.167...</td></tr></table>	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	20	24.6	0.86016...	16.2601...	55	51.66	0.21594...	58.5559...	48	46.74	0.03396...	49.2939...	10	5.4	3.91851...	18.5185...	8	11.34	0.98373...	5.6437...	9	10.26	0.15473...	7.8947...	Totals:		6.167...	156.167...	dM1
	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$																														
	20	24.6	0.86016...	16.2601...																														
	55	51.66	0.21594...	58.5559...																														
	48	46.74	0.03396...	49.2939...																														
10	5.4	3.91851...	18.5185...																															
8	11.34	0.98373...	5.6437...																															
9	10.26	0.15473...	7.8947...																															
Totals:		6.167...	156.167...																															
$X^2 = \sum \frac{(O - E)^2}{E} \quad \text{or} \quad \sum \frac{O^2}{E} - 150$ = awrt 6.17 $\nu = (3 - 1)(2 - 1) = 2$ $\chi^2_2(0.05) = 5.991$ [Reject H_0 /significant/in the CR] There is sufficient evidence to suggest there is an association between <u>treatment</u> and presence of <u>fungus</u> .	dM1 A1 B1 B1ft A1ft																																	
Notes		[8]																																
1 st B1 both hypotheses correct with treatment (oe) and fungus (oe) (treatment and fungus need to only appear in either H_0 or H_1). May be written in terms of independence. 1 st M1 attempt at $\frac{\text{row total} \times \text{column total}}{\text{total}}$ (can be implied by at least one correct E_i to 1dp) 2 nd M1 (dep on 1 st M1) at least 2 correct terms for $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i (allow 2sf accuracy) (May be implied by awrt 6.17 or awrt 156.17) 3 rd M1 (dep on 2 nd M1) for using $\sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 150$ (May be implied by awrt 6.17) 1 st A1 awrt 6.17 2 nd B1 DoF/ $\nu = 2$ (May be implied by 5.991) 3 rd B1ft 5.991 (or better) allow ft from their stated degrees of freedom) 2 nd A1ft (dep on 3 rd M1 and 3 rd B1) for a correct ft contextualised conclusion. Must include “treatment” and “fungus”. Ignore any non-contextual statements. If hypotheses are the wrong way round then A0.																																		

Question Number	Scheme	Marks
1.		
(a)	$E(2\bar{X}) = 2E(\bar{X}) = \frac{2(1+\alpha)}{2}$	M1
	$= 1 + \alpha \neq \alpha$ (therefore $2\bar{X}$ is a biased estimator of α)	A1
		(2)
(b)	$\bar{x} = 6 \quad 2\bar{x} - 1 =$	M1
	<u>11</u>	A1
		(2)
		Total 4
	Notes	
(a)	M1 for use of $2 \times \frac{1+\alpha}{2}$ Also correct comparison with α and no incorrect working seen. $\frac{1+\alpha}{2} \neq \alpha$ is M0A0	
(b)	M1 for attempt at $\bar{x}$ and use of $2\bar{x} - 1$ or ft use of their $E(\bar{X})$ from part (a) to find an estimate for α A1 11 cao	

Question Number	Scheme	Marks
2. (a)	Expected values: 13.75, 19.25, 17	B1
	$\frac{(18-13.75)^2}{13.75}$ $\frac{(15-19.25)^2}{19.25}$ $\frac{(17-17)^2}{17}$	M1
	Pass 1.31 0.94 0 2.25	A2
	9.43	
	(4)	
(b)	H_0 : Degree/category is independent of department (no association)	B1
	H_1 : Degree/category is not independent of department (association)	
	$\nu = (3-1)(3-1) = 4$	B1
	$\chi^2_4(5\%) = 9.488$	B1ft
	'9.43' < 9.488, so do not reject H_0 / not significant	M1
	Not enough evidence to show that the Degree/category is not independent of department (no association) (at the 5% level of significance)	A1ft
	(5)	
	Total 9	
	Notes	
(a)	B1 All correct expected values (may be implied by A2) M1 Attempting $\frac{(O-E)^2}{E}$ for their E (may be implied by one correct value) A2 all 5 values correct (allow awrt) (A1 any 2 values correct)	
(b)	B1 Both hypotheses required must mention degree/category and department at least once. Use of "relationship" or "correlation" or "connection" or "link" award B0. B1 4 can be implied by 9.488 seen B1ft 9.488 or better. Follow through their ν so may see $\chi^2_{3,0.05} = 7.815$ $\chi^2_{2,0.05} = 5.991$ M1 Correct non-contextual conclusion for their (a) and c.v. Can be implied by correct conclusion in context ft their (a) and c.v. A1ft A correct comment in context. Condone "relationship" or "connection" here but not "correlation". Follow through from their test statistic and cv, but hypotheses must be correct.	

Question	Scheme			Marks
3 (a)	$\frac{86 \times 300}{1200}$ or $\frac{1114 \times 300}{1200}$			M1
	21.5 and 278.5			A1
				(2)
(b)	H ₀ : Making a claim and age are independent (not associated) H ₁ : Making a claim and age are not independent (associated)			B1
	Observed	Expected	$\frac{(O - E)^2}{E}$	M1
	14	"21.5"	$\frac{(14 - "21.5")^2}{"21.5"} = 2.6162...$	
	286	"278.5"	$\frac{(286 - "278.5")^2}{"278.5"} = 0.20197...$	
	$\sum \frac{(O - E)^2}{E} = 7.123 + "2.616..." + "0.2019..."$			M1
	$= 9.941...$ awrt 9.94			A1
	$\nu = (2 - 1)(3 - 1) = 2$			B1
	$\chi^2_2(0.01) = 9.210 \Rightarrow \text{CR: } X^2 \geq 9.21[0]$			B1ft
	[in the CR/significant/Reject H ₀] There is sufficient evidence to suggest that making a claim is not independent of age .			dA1ft
				(7)
	Notes			Total 9
(a)	M1	A correct method for finding one expected value. Implied by one correct value.		
	A1	Correct answer for both 21.5 and 278.5		
(b)	B1	For both hypotheses correct. Must mention claim and age at least once. Use of “relationship” or “correlation” or “connection” is B0		
	M1	A correct method for finding both contributions to the χ^2 value or awrt 2.62 or awrt 0.202 Allow truncated answers of 2.61 and 0.201 May be implied by awrt 9.94		
	M1	Adding their two values to 7.123 (may be implied by a full χ^2 calculation, with at least 3 correct expressions or values. Do not ISW)		
	A1	awrt 9.94		
	B1	$\nu = 2$ This mark can be implied by a correct critical value of 9.21 or better		
	B1ft	9.21[0] or better ft their Degrees of freedom common ones $\nu = 3$ is 11.345		
	dA1ft	Independent of hypotheses but dependent on both M marks being awarded. We will ft their test statistic and CV only. A correct contextual conclusion compatible with their values, which has the words claim and age. eg if they have 11.345 and 9.94 they should say it is independent/ not associated. Do not allow contradicting statements.		
Full calculations for(b)				
eg $\frac{(24 - 14.33)^2}{14.33} + \frac{(176 - 185.67)^2}{185.67} + \frac{(48 - 50.17)^2}{50.17} + \frac{(652 - 649.83)^2}{649.83} + \frac{(14 - "21.5")^2}{21.5} + \frac{(286 - "278.5")^2}{278.5}$				
or awrt 6.52 + awrt 0.5 + awrt 0.09 + awrt 0.01 + awrt 2.62 + 0.20				
or $\frac{24^2}{14.33} + \frac{176^2}{185.67} + \frac{48^2}{50.17} + \frac{652^2}{649.83} + \frac{14^2}{"21.50"} + \frac{286^2}{"278.5"} - 1200$				
or awrt 40.19 + awrt 166.83 + awrt 45.92 + awrt 654.17 + awrt 9.116 + awrt 293.702 – 1200				

Question Number	Scheme			Marks
3 (a)	$\frac{36 \times 35}{250}$			M1
	5.04			A1
				(2)
(b)	H ₀ : There is no association between when the car first experienced an engine problem and the type of engine H ₁ : There is an association between when the car first experienced an engine problem and the type of engine			B1
	Observed	Expected	$\frac{(O - E)^2}{E}$	M1
	10	5.04	$\frac{(10 - '5.04')^2}{'5.04'} (= 4.881)$	
	X ² = 7.444 + '4.881'			dM1
	= 12.325 awrt 12.3			A1
	ν = (3 – 1)(3 – 1) = 4 χ ₄ ² (0.05) = 9.488 ⇒ CR: X ² ...9.488			B1 B1ft
	[In the CR/Significant/Reject H ₀] There is evidence of an association between when the car first experienced an engine problem and the type of engine			A1
				(7)
	Notes			Total 9
(a)	M1	For a correct method for finding the expected value (ignore other values if found)		
	A1	5.04 oe isw e.g. if they attempt to find other expected frequencies as well. 5.04 with no incorrect working seen is 2/2		
(b)	B1	Both hypotheses correct. Engine problems and type of engine mentioned at least once. (may be written in terms of independence (or association) e.g. H ₀ : When the car first experienced an engine problem and the type of engine are independent H ₁ : When the car first experienced an engine problem and the type of engine are not independent		
	M1	For a correct method for finding the contribution to the X ² value (seen or implied). May be seen in (a) or by the table		
	dM1	Dependent on the previous method mark. For finding the test statistic X ² . Usually by adding their value to 7.444 (seen or implied). If they attempt to find the individual values using the table then the method must be correct.		
	A1	awrt 12.3 (sight of awrt scores M1dM1A1)		
	B1	ν = 4 If not seen then may be implied by 9.488 or 11.143		
	B1ft	9.488 or better or ft their DoF (e.g. ν = 2 , 5.991 or better or ν = 8, 15.507 or ν = 9, 16.919 or better)		
	A1	Dependent on all previous method marks and their X ² must lie in the critical region i.e. X ² > "9.488" (where X ² > 7.444) A correct contextualised conclusion which is rejecting H ₀ Must mention engine problem and type of engine oe If hypotheses are the wrong way round, then A0 here. Do not withhold this mark if their hypotheses are in terms of ρ . Contradictory statements score A0. Condone “relationship” or “connection” here but not “correlation”.		

Question Number	Scheme	Marks
3. (a)	All expected frequencies are $(88 \div 4) = 22$ Degrees of freedom = 3, so critical value $\chi_3^2(5\%) = 7.815$ (Not significant so) insufficient evidence to suggest <u>not</u> uniformly distributed	B1 B1, B1ft B1 (4)
(b)	e.g. H_0 : School is independent of club chosen H_1 : Club chosen depends on which school a student is from	B1 (1)
(c)	$\frac{28 \times 17}{88} = 5.409...$ awrt <u>5.41</u>	B1 (1)
(d)	Expected frequency for Music and School C = $4.77 < 5$ (Allow $\frac{105}{22}$ for 4.77) So combine Music column with another column giving 3x3 table so 4 df	B1 B1 (2)
(e)	Critical value $\chi_4^2(5\%) = 9.488$ [Not significant so] insufficient evidence of an association between school and choice of club	B1 B1 (2)
[10 marks]		
Notes		
(a)	Ignore values of any test statistics calculated in (a) or (e) 1 st B1 for 22 2 nd B1 for degrees of freedom = 3 (can be implied by sight of 7.815 as cv) 3 rd B1ft for 7.815 (or better - cal: 7.814727910... or correct 5% cv for their d.f.) 4 th B1 for comment suggesting uniform distribution is a suitable model. Must follow from comparing 6.09 with their cv. Do not allow contradictory statements e.g. "significant" so uniform dist' is suitable	
(b)	B1 for both hypotheses with some context ("club" and "school" mentioned at least once) Use of "independence" or "association"	
(c)	B1 for a correct expression or awrt 5.41 (allow $\frac{119}{22}$)	
(d)	1 st B1 for identifying that Music & School C has E_i that is < 5 (a value to 2 sf should be seen, may be in (c), but must state <u>this</u> $E_i < 5$ as well) 2 nd B1 for pooling <u>music</u> with another <u>column</u> leading to 3x3 table and 4 degrees of freedom Must clearly state the pooling and evidence for 4 df e.g. allow $(3-1) \times (4-1-1)$ [NB pooling with Art gives 4.3987..., with Sports 4.3247..., with Computers 7.2879...]	
(e)	1 st B1 for 9.488 (or awrt 9.488) 2 nd B1 for a correct, not significant, conclusion mentioning <u>school</u> and <u>clubs</u>	

Question Number	Scheme					Marks
4	H ₀ : There is no association between type of property and the time taken to sell it H ₁ : There is an association between type of property and the time taken to sell it					B1
	Expected	Bungalow	Flat	House	Total	M1 A1
	Within 3 months	10.496	31.488	40.016	(82)	
	More than 3 months	5.504	16.512	20.984	(43)	
	Total	(16)	(48)	(61)	(125)	
	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$		dM1 A1
	7	10.496	1.1644...	4.6684...		
	29	31.488	0.1965...	26.7085...		
	46	40.016	0.8948...	52.8788...		
	9	5.504	2.2205...	14.7165...		
	19	16.512	0.3748...	21.8628...		
	15	20.984	1.7064...	10.7224...		
	Totals		6.557...	131.557...		
	$[X^2 =] \sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 125$					dM1
= 6.557... awrt 6.56					A1	
$\nu = (2 - 1)(3 - 1) = 2$					B1	
$\text{c}_2^2(0.05) = 5.991 \Rightarrow \text{CR: } X^2 \dots 5.991$					B1	
[in the CR/significant/Reject H ₀] There is sufficient evidence to suggest that there is an association between type of property and the time taken to sell it.					A1	
					(10)	
Notes					Total 10	
4	B1	Both hypotheses correct. Must mention “type of property” and “time taken” at least once. (may be written in terms of independence)				
	M1	Some attempt at $\frac{(\text{Row Total})(\text{Column Total})}{(\text{Grand Total})}$ Can be implied by at least one correct E_i to 1dp				
	A1	All expected frequencies correct				
	dM1	Dependent on 1 st M1 for at least 2 correct terms for $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i . Accept 2 sf accuracy.				
	A1	At least 3 correct $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ terms to 2dp or better. Allow truncated answers.				
	dM1	Dependent on 2 nd M1 For applying either $\sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 125$				
	A1	awrt 6.56				
	B1	$\nu = 2$ This mark can be implied by a correct critical value of 5.991				
	B1	5.991				
	A1	Dependent on the 3 rd M1 and 3 rd B1. A correct contextualised conclusion which is rejecting H ₀ . Must mention type and time . Contradictory statements score A0. e.g. “significant, do not reject H ₀ ”.Condone “relationship” or “connection” here but not “correlation”.				

Question Number	Scheme			Marks
2 (a)	$\frac{60 \times 60}{240}$ or $\frac{60 \times 84}{240}$ or $\frac{60 \times 96}{240}$			M1
	15 and 21 and 24			A2
				(3)
(b)	H_0 : There is no association between the payment amount and payment method used H_1 : There is an association between the payment amount and payment method used			B1
	Observed	Expected	$\frac{(O - E)^2}{E}$	M1
	23	15	$\frac{(23 - '15')}{'15'} = 4.2667$	
	21	21	$\frac{(21 - '21')}{'21'} = 0$	
	16	24	$\frac{(16 - '24')}{'24'} = 2.6667$	
	$\chi^2 = 2.4048 + '4.2667' + '0' + '2.6667'$			M1
	$= 9.3381 \dots$ awrt 9.34			A1
	$\nu = (3 - 1)(3 - 1) = 4 \quad \chi^2_4(0.05) = 9.488 \Rightarrow \text{CR: } X^2 \geq 9.488$			B1 B1ft
	[Not in the CR/Not significant/Do not reject H_0] There is no evidence of an association between the payment amount and payment method used			dA1
			(7)	
	Notes			Total 10
(a)	M1	For a correct method for finding one expected value		
	A2	For all 3 answers correct (A1 for 2 correct answers or 1 correct and 3 values that sum to 60)		
(b)	B1	Both hypotheses correct. Must mention method and amount with payment at least once. (may be written in terms of independence)		
	M1	For a correct method for finding all three contributions to the χ^2 value ft their part a May be implied by 3 correct values If expected values are incorrect then working must be shown		
	M1	For adding their values to 2.4048 (If all 9 values are calculated the 6 values not found in part (a) must have working shown or the correct values seen or awrt 9.34)		
	A1	awrt 9.34		
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488		
	B1ft	9.488 or better ft their DoF		
	dA1	Dependent on both M marks. A correct contextualised conclusion which is not rejecting H_0 Must mention method and amount . If no hypotheses or they are the wrong way round, then A0 here. Contradictory statements score A0. e.g. “Significant, do not reject H_0 ” ”.Condone “relationship” or “connection” here but not “correlation”.		

Question Number	Scheme	Marks
4. (a)	$\frac{[0 \times 24] + 1 \times 34 + 2 \times 28 + 3 \times 21 + 4 \times 8 + 5 \times 5}{120} [= 1.75]^*$	B1*cs0 (1)
(b)	$[s =] 120 \times \frac{e^{-1.75} 1.75^4}{4!} [= 8.15]^*$ or $[s =] 120 - \left(20.85 + 36.49 + 31.93 + 120 \times \frac{e^{-1.75} 1.75^3}{3!} + 3.95 \right) [= 8.15]^*$	B1*cs0 (1)
(c)	$[r =] 18.63$	B1 (1)
(d)	H_0 : Poisson distribution is a good fit. H_1 : Poisson distribution is not a good fit $\sum \frac{(O_i - E_i)^2}{E_i} = 1.43 + \frac{(8 + 5 - (8.15 + 3.95))^2}{8.15 + 3.95}$ $= 1.49694... \text{ awrt } 1.5(0)$ $\nu = 5 - 1 - 1 = 3$ $\chi^2_3(0.05) = 7.815$ [Do not reject H_0/not significant] There is insufficient evidence to reject the office manager's belief or the number of jobs sent to the printer are consistent with a Poisson distribution.	B1 M1 M1 A1 B1 B1ft A1 (7) [10]
Notes		
(a)	B1cs0 correct calculation, minimum working $\frac{34 + 56 + 63 + 32 + 25}{120} = 1.75^*$	
(b)	B1cs0 fully correct calculation (may be seen in stages) leading to 8.15*	
(c)	For 18.63 (This may be seen in part (b) if labelled as r)	
(d)	1 st B1 both hypotheses correct (mention of 1.75 is B0) 1 st M1 evidence of combining last 2 cells e.g. 8 + 5 and 8.15 + 3.95 2 nd M1 use of $1.43 + \sum \frac{(O_i - E_i)^2}{E_i}$ for remaining cells (Condone cells not combined. May be implied by $1.43 + 0.00276... + 0.279... \text{ or awrt } 1.71$) 1 st A1 awrt 1.50 (allow 1.5 from correct working) 2 nd B1 Dof/ $\nu = 3$ implied by a correct critical value of 7.815 3 rd B1ft 7.815 (allow ft on the ν so may see 9.488 or 11.070 etc) 2 nd A1 (dep on 2 nd M1) a correct conclusion which states that the office manager's belief is correct/the data are consistent with a Poisson distribution which must be consistent with the test statistic and CV. Condone Po(1.75) is a suitable model. This mark is independent of the hypotheses	

Question Number	Scheme		Marks									
4(a)	Label the houses in area A 1- 41, area B 1 – 164, area C 1 – 123 and area D 1 - 82		M1									
	<u>Use random numbers</u> to select a ...		M1									
	Simple random sample of <u>20</u> area <u>A</u> , <u>80</u> area <u>B</u> , <u>60</u> area <u>C</u> and <u>40</u> area <u>D</u>		A1									
			(3)									
(b)	$\frac{357 \times 260}{595}$ or $\frac{238 \times 260}{595}$		M1									
	156 and 104		A1									
			(2)									
(c)	<table><tr><td>Observed</td><td>Expected</td><td>$\frac{(O-E)^2}{E}$</td></tr><tr><td>162</td><td>"156"</td><td>$\frac{(162 - "156")^2}{"156"} = \frac{3}{13} = 0.2307...$</td></tr><tr><td>98</td><td>"104"</td><td>$\frac{(98 - "104")^2}{"104"} = \frac{9}{26} = 0.3461...$</td></tr></table>		Observed	Expected	$\frac{(O-E)^2}{E}$	162	"156"	$\frac{(162 - "156")^2}{"156"} = \frac{3}{13} = 0.2307...$	98	"104"	$\frac{(98 - "104")^2}{"104"} = \frac{9}{26} = 0.3461...$	M1
Observed	Expected	$\frac{(O-E)^2}{E}$										
162	"156"	$\frac{(162 - "156")^2}{"156"} = \frac{3}{13} = 0.2307...$										
98	"104"	$\frac{(98 - "104")^2}{"104"} = \frac{9}{26} = 0.3461...$										
	$\chi^2 = 4.657 + "0.2307..." + "0.346..."$		M1									
	$= 5.234....$	awrt 5.23	A1									
	$\nu = (2 - 1)(3 - 1) = 2$		B1									
	$\chi^2_2(0.05) = 5.991 \Rightarrow \text{CR: } \chi^2 > 5.991$		B1ft									
	There is no evidence to suggest that there is an association between <u>age</u> and <u>listening to LSB</u>		dA1									
			(6)									
	Notes		Total 11									
(a)	M1	For suitable labelling of all four areas. E.g. for area A: 1 – 41 or 0 - 40										
	M1	For use of random numbers to select houses in each area.										
	A1	For 20 A, 80B, 60C and 40 D (dependent on 2 nd M1 only)										
		NB A simple random sample of 20 A, 80B, 60C and 40 D scores M0M1A1.										
		Allow M1 : allocate random numbers to each house										
		M1 : arrange the numbers in order										
		A1 : select the 1 st 20 for area <u>A</u> , <u>80</u> for area <u>B</u> , <u>60</u> for area <u>C</u> and <u>40</u> for area <u>D</u>										
		SC If M0M0 scored then award B1 for <u>20</u> area <u>A</u> , <u>80</u> area <u>B</u> , <u>60</u> area <u>C</u> and <u>40</u> area <u>D</u>										
(b)	M1	A correct method for finding one expected value.										
	A1	Correct answer for both values										
(c)	M1	A correct method for finding both contributions to the χ^2 value										
	M1	Adding the two values to 4.657 (may be implied by a full χ^2 calculation, do not ISW)										
	A1	awrt 5.23										
	B1	2										
	B1ft	5.991 or better ft their DoF										
	dA1	A correct contextual conclusion, which has the words age and listening dependent on both M marks being awarded.										
		NB if they give a p value of 0.0730... rather than the CV they can get M1M1B1B0A1										

Question Number	Scheme				Marks
7 (a)	H_0 : The observed distribution can be modelled by a discrete uniform distribution H_1 : The observed distribution cannot be modelled by a discrete uniform distribution				B1 (1)
(bi)	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	B1 M1
	$x + 6$	x	$\frac{36}{x}$	$\frac{(x + 6)^2}{x}$	
	$x - 8$	x	$\frac{64}{x}$	$\frac{(x - 8)^2}{x}$	
	$x + 8$	x	$\frac{64}{x}$	$\frac{(x + 8)^2}{x}$	
	$x - 5$	x	$\frac{25}{x}$	$\frac{(x - 5)^2}{x}$	
	$x + 4$	x	$\frac{16}{x}$	$\frac{(x + 4)^2}{x}$	
	$x - 5$	x	$\frac{25}{x}$	$\frac{(x - 5)^2}{x}$	
	Total = 6x	Total = 6x	Total = $\frac{230}{x}$	Total = $\frac{6x^2 + 230}{x}$	
(bii)	$X^2 = \sum \frac{(O - E)^2}{E}$ or $\sum \frac{O^2}{E} - 6x$; $\frac{230}{x}$ or $\frac{6x^2 + 230}{x} - 6x$				M1 ; A1
	$\nu = 6 - 1 = 5$; $c^2_5(0.05) = 11.070 \Rightarrow \text{CR: } X^2 \dots 11.070$				B1 ; B1
	Do not reject H_0 if $\frac{230}{x}$ „ '11.070' or $\frac{6x^2 + 230}{x} - 6x$ „ '11.070'				M1
	$x \dots 20.7768\dots$ So $x = 21$				A1 (8)
	Hence the die was rolled “21” $\times 6 = 126$ times				M1 A1 (2)
Notes					Total 11
(a)	B1	for both hypotheses correct Allow H_0 : the die is not biased H_1 : the die is biased			
(bi)	B1	for expected frequency = x			
	M1	for one correct $\frac{(O - E)^2}{E}$ or $\frac{O^2}{E}$ ft their expected frequency			
	M1	for an attempt at X^2 ft their values (At least 4 of these need to be seen and added)			
	A1	for either $\frac{230}{x}$ or $\frac{6x^2 + 230}{x} - 6x$			
	B1	for $\nu = 6 - 1 = 5$ May be implied by a correct critical value			
	B1	for a correct critical value ft their DOF (NB common error is $\nu = 4$ so $c^2_4(0.05) = 9.488$)			
	M1	for either $\frac{230}{x}$ „ their CV or $\frac{6x^2 + 230}{x} - 6x$ „ their CV Allow < rather than „			
	A1	for $x = 21$ provided the previous M mark has been awarded			
(bii)	M1	for their 21×6 Allow $6 \times x$ or the answer to $6 \times$ their value for x			
	A1	cao			

Question Number	Scheme		Mark
4 (a)	$\left[\int_a^{a+1} \frac{2}{25} t \, dt \right] = \frac{2}{25} \left[\frac{t^2}{2} \right]_a^{a+1} \text{ or } F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{25} t^2 & 0 \leq t < 5 \text{ or} \\ 1 & t > 5 \end{cases}$		M1
	$\frac{1}{2} \left(\frac{2}{25} (a+1) + \frac{2}{25} a \right) (a+1-a)$		M1
	$\frac{1}{25} ((a+1)^2 - a^2) \text{ or } \frac{1}{25} (a+1)^2 - \frac{1}{25} a^2 \text{ or } \left(\frac{1}{25} a + \frac{1}{25} + \frac{1}{25} a \right)$		M1
	$\frac{1}{25} (a^2 + 2a + 1 - a^2) \text{ oe } \left[= \frac{1}{25} (2a + 1) \right]^*$		A1*
			(3)
(b)	H_0 : The data could be modelled by the p.d.f H_1 : The data could not be modelled by the p.d.f		B1
	Expected frequencies: 6, 18, 30, 42, 54		M1 A1
	$\sum \frac{(O - E)^2}{E} = \frac{(10 - '6')^2}{'6'} + \dots + \frac{(68 - '54')^2}{'54'}$		M1
	or $\sum \frac{O^2}{E} - N = \frac{10^2}{'6'} + \dots + \frac{68^2}{'54'} - 150 \text{ or } 2.666\dots + 1.388\dots + 1.2 + 1.166\dots + 3.629$		
	= 10.05... awrt 10.1		A1
	$\nu = 4$		B1
	$\chi^2_{4}(0.05) = 9.488 \Rightarrow CR \geq 9.488$		B1ft
	[In the CR so there is sufficient evidence to reject H_0]		
	Sufficient evidence to say that data does not fit the given p.d.f		dA1
			(8)
	Notes		Total 11
(a)	M1	For correct integration, ignore limits or finding the area of a trapezium	
	M1	For substitution of the limits. May be implied by $\frac{1}{25} (a^2 + 2a + 1 - a^2)$ or simplifying the expression for the area of the trapezium	
	A1*	Answer is given so no incorrect working should be seen. At least one correct line of working from the method mark to the final answer should be seen	
(b)	B1	Both hypotheses correct. Allow H_0 : The p.d.f/f(t) is a suitable model H_1 : The p.d.f/f(t) is not a suitable model	
	M1	For a correct method to find at least one expected frequency e.g. $\frac{1}{25} \times 150$ Ignore any reference to limits	
	A1	For all 5 expected frequencies correct	
	M1	For an attempt at the test statistic, at least 2 correct expressions/values ft their expected frequencies	
	A1	awrt 10.1	
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488	
	B1ft	9.488 or better ft their DoF	
	dA1	Dependent on 2 nd M1. A correct conclusion based on their χ^2 critical value If no hypotheses or they are the wrong way round, then A0 here.	

Question Number	Scheme	Marks																									
4(a)	H ₀ : favourite flavours occur in the ratio 10 : 5 : 2 : 3	B1																									
	H ₁ : favourite flavours do not occur in the ratio 10 : 5 : 2 : 3																										
	<table><tr><td></td><td>Chocolate</td><td>Vanilla</td><td>Strawberry</td><td>Other</td></tr><tr><td>observed</td><td>188</td><td>95</td><td>40</td><td>77</td></tr><tr><td>expected</td><td>200</td><td>100</td><td>40</td><td>60</td></tr><tr><td>$\frac{(O_i - E_i)^2}{E_i}$</td><td>$\frac{(188 - "200")^2}{"200"}$ (= 0.72)</td><td>$\frac{(95 - "100")^2}{"100"}$ (= 0.25)</td><td>$\frac{(40 - "40")^2}{"40"}$ (= 0)</td><td>$\frac{(77 - "60")^2}{"60"}$ (= 4.816...)</td></tr><tr><td>$\frac{O_i^2}{E_i}$</td><td>$\frac{188^2}{"200"}$ (=176.72)</td><td>$\frac{95^2}{"100"}$ (=90.25)</td><td>$\frac{40^2}{"40"}$ (=40)</td><td>$\frac{77^2}{"60"}$ (=98.816...)</td></tr></table>		Chocolate	Vanilla	Strawberry	Other	observed	188	95	40	77	expected	200	100	40	60	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(188 - "200")^2}{"200"}$ (= 0.72)	$\frac{(95 - "100")^2}{"100"}$ (= 0.25)	$\frac{(40 - "40")^2}{"40"}$ (= 0)	$\frac{(77 - "60")^2}{"60"}$ (= 4.816...)	$\frac{O_i^2}{E_i}$	$\frac{188^2}{"200"}$ (=176.72)	$\frac{95^2}{"100"}$ (=90.25)	$\frac{40^2}{"40"}$ (=40)	$\frac{77^2}{"60"}$ (=98.816...)	M1 M1
	Chocolate	Vanilla	Strawberry	Other																							
observed	188	95	40	77																							
expected	200	100	40	60																							
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(188 - "200")^2}{"200"}$ (= 0.72)	$\frac{(95 - "100")^2}{"100"}$ (= 0.25)	$\frac{(40 - "40")^2}{"40"}$ (= 0)	$\frac{(77 - "60")^2}{"60"}$ (= 4.816...)																							
$\frac{O_i^2}{E_i}$	$\frac{188^2}{"200"}$ (=176.72)	$\frac{95^2}{"100"}$ (=90.25)	$\frac{40^2}{"40"}$ (=40)	$\frac{77^2}{"60"}$ (=98.816...)																							
	$\sum \frac{(O_i - E_i)^2}{E_i} = 5.786... \text{ or } \sum \frac{O_i^2}{E_i} - 400 = 405.7867 - 400 \text{ (awrt 5.79)}$	A1																									
	$\nu = 3$	B1																									
	CV is 7.815	B1ft																									
	[5.79 < 7.815] so insufficient evidence to reject H ₀ <u>or</u> not significant	M1																									
	There is no evidence to suggest that people’s favourite flavours of ice-cream do not occur in the given ratio . oe	A1ft																									
		(8)																									
(b)(i)(ii)	$\frac{188 \times 130}{400} \text{ or } \frac{112 \times 95}{400}$	M1																									
	61.1 and 26.6	A1																									
		(2)																									
(c)	6	B1																									
		(1)																									
	Notes	Total 11																									
(a)	B1 Both hypotheses correct. Must state the ratio or refer to the “given ratio” oe. Does not need to state the actual flavours. Accept statements e.g. H ₀ : the proportion of people who like different ice creams is 10:5:2:3 Accept H ₀ : the ratio 10:5:2:3 is correct. Do not accept e.g. H ₀ : the manager’s belief is correct (without reference to the ratio)																										
	M1 At least 2 expected values correct. Check if each cell has been calculated separately and the totals add up to 200, 100, 40 and 60 (seen)																										
	M1 Correct method seen or may be implied by values for at least 2 flavours. Implied by awrt 5.79																										
	A1 awrt 5.79																										
	B1 correct degrees of freedom stated or implied by any awrt 0.072 0.115 0.216 0.352 0.584 6.251 7.815 9.348 11.345 12.838																										
	B1ft awrt 7.815 follow through on 6 degrees only = 12.592																										
	M1 Independent of hypotheses, ft their χ^2 value and their CV. A correct comment. Allow																										

	“Accept H_0 ” or e.g. “do not reject H_0 ” Do not award if contradicting non contextual comments given. If their χ^2 value < their CV e.g. “do not reject H_0 ” If their χ^2 value > their CV e.g. “reject H_0 ” (May be implied by a correct ft contextual comment)
	A1ft dependent on all the previous method marks. A correct contextual comment for their χ^2 value and their CV that mentions flavour and ratio oe (does not need to state the actual ratio) Accept sufficient evidence to support the manager’s belief oe condone “the manager is correct”
(b)	M1 A correct method to find one of the values. Implied by one correct value
	A1 Both answers correct.
(c)	B1 cao

Question Number	Scheme		Marks
5 (a)	$\hat{p} = \frac{0 \times 2 + 1 \times 40 + 2 \times 90 + 3 \times 85 + 4 \times 30 + 5 \times 3}{250 \times 5}$ or		M1
	$\hat{p} = \frac{40 + 180 + 255 + 120 + 15}{250 \times 5} \left(= \frac{610}{1250} = \frac{122}{250} \right)$		
	$= 0.488$		A1
			(2)
(b)	$250 - (41.92 + \dots + 6.92)$		M1
	$= 8.79$		A1
			(2)
(c)	H_0 : Binomial distribution is suitable/sensible (model) H_1 : Binomial distribution is not suitable/sensible (model)		B1
	$\frac{(2 - '8.8')^2}{'8.8'} [= 5.25 \dots]$		M1
	$\left[\sum \frac{(O - E)^2}{E} = \right] 5.70 + '5.25' = 10.95$		dM1 A1
	$\nu = 6 - 1 - 1 = 4$		B1
	$\chi^2_4(0.05) = 9.488 \Rightarrow \text{CR} \dots 9.488$		B1ft
	[In CR/Significant/Reject H_0] Significant evidence to suggest that a binomial distribution is not a suitable model		A1
			(7)
Notes			Total 11
(a)	M1	For a correct method to find $\hat{p}$ May be implied by 0.488. Allow one error or omission for one of their products.	
	A1	0.488 oe e.g. $\frac{61}{125}$	
(b)	M1	For a correct method to find r (Allow $0.512^5 \times 250$) May be implied by 8.79 or 8.8(0)	
	A1	Allow 8.8(0)	
(c)	B1	Both Hypotheses correct. Must be attached to H_0 and H_1 Do not allow B(5, 0.488) is a suitable model	
	M1	For use of $\frac{(O - E)^2}{E}$ ft their part (b). If no value is found in (b) or they use 5.7 then M0. May be implied by awrt 5.25 or awrt 10.95 or 10.9 or 11.0	
	dM1	Dependent on the first method mark. For $5.70 + '5.25'$ ft their 5.25 May correctly attempt $\sum \frac{(O - E)^2}{E}$ or $\sum \left(\frac{O^2}{E} \right) - 250$. May be implied by awrt 10.95 or allow 10.9 or 11.0	
	A1	awrt 10.95 (allow 10.9 or 11.0)	
	B1	4 [degrees of freedom] If not seen then may be implied by 9.488 or 11.143 (or better)	
	B1ft	9.488 or better or ft on their dof e.g. 2: 5.991, 3: 7.815, 5: 11.07(0), 6: 12.592	
	A1	Dependent on all previous method marks and a critical value > 5 where $X^2 > CV$ Requires a correct conclusion rejecting the null hypothesis. Must mention the (binomial) distribution/model	

Question Number	Scheme					Marks
6 (a)	$\frac{0 \times 1 + 1 \times 10 + 2 \times 23 + 3 \times 15 + 4 \times 19 + 5 \times 9 + 6 \times 3}{80} = 3 *$					B1
						(1)
(b)	$r = e^{-3} \times 80 = 3.983 \quad s = \frac{e^{-3} \times 3^5}{5!} \times 80 ; = 8.066$					M1 ; A1
	$t = 80 - (r + 11.949 + 17.923 + 17.923 + 13.443 + s) ; = 6.713$					M1 ; A1
						(4)
(c)	H_0 : Poisson (distribution) is a reasonable/suitable/ sensible (model) H_1 : Poisson (distribution) is not a /reasonable/suitable/ sensible (model).					B1
	Number of emails	Combined Observed	Combined Expected	$\frac{(O - E)^2}{E}$	$\frac{O^2}{E}$	M1
	≤ 1	11	15.932	1.5267...	7.5947...	
	2	23	17.923	1.4381...	29.5151...	
	3	15	17.923	0.4767...	12.5537...	
	4	19	13.443	2.2971...	26.8541...	
	5	9	8.065	0.1083...	10.0433...	
	≥ 6	3	6.714	2.0544...	1.3404...	
	Totals			7.901...	87.901...	
	$X^2 = \sum \frac{(O - E)^2}{E} \quad \text{or} \quad \sum \frac{O^2}{E} - 80$					M1
	$= 7.901...$					awrt 7.90
	$v = 6 - 1 - 1 = 4$					B1
	$c^2_4(0.10) = 7.779 \Rightarrow \text{CR: } X^2 \dots 7.779$					B1
	[since $X^2 = 7.90$ does lie in CR, then there is sufficient evidence to reject H_0]					
	Sufficient evidence to say that Poisson is not a reasonable model					A1
					(7)	
Notes					Total 12	
6 (a)	B1	For a correct method to shown that the mean is 3				
(b)	M1	Use of $\frac{e^{-\lambda} \times \lambda^r}{r!} \times 80$ or May be implied by a correct answer for either r or s				
	A1	$r = 3.983$ and $s = 8.066$ (allow $r = 3.984$ and $s = 8.064$ as these come from tables)				
	M1	A correct method that ensures that expected totals = 80				
	A1	$t = 6.713$ (allow $t = 6.714$ if tables used)				
(c)	B1	Both hypotheses correct. Must mention Poisson at least once.				
	M1	Combining 0 emails and 1 email. Must have both observed and expected frequencies				
	M1	An attempt at the test statistic, at least 2 correct expressions/values (to awrt 2dp)				
	A1	awrt 7.90 Accept 7.9 if no incorrect working seen				
	B1	$v = 4$ This mark can be implied by a correct critical value of 7.779				
	B1	7.779				
	A1	A correct conclusion based on their X^2 value and their χ^2 critical value				

Qu. No.	Scheme		Marks
6(a)	$P(S < 303.5) = P\left(Z < \frac{303.5 - 310}{4}\right)$ or $P(S > 315.5) = P\left(Z > \frac{315.5 - 310}{4}\right)$		M1
	$= 0.05208$ or $0.084565\dots$ awrt 0.052 or awrt 0.084/0.085		A1
	So $a = 5.2$ or $b = 8.5$ awrt 5.2 or awrt 8.4/8.5		A1
	e.g. $b = 100 - 10.6 - 16.3 - 19.6 - 18.4 - 13.6 - 7.8 - '5.2'$		M1
	Both $a = 5.2$ and $b = 8.5$ awrt 5.2/5.3 and awrt 8.4/8.5		A1
			(5)
(b)	H_0 : The normal distribution $N(310, 16)$ is a suitable model/The data are consistent with the model. H_1 : The normal distribution $N(310, 16)$ is not a suitable model/The data are not consistent with the model.		B1
	$[X^2 =] \sum \frac{(O - E)^2}{E} = \frac{(5 - '5.2')^2}{'5.2'} + \frac{(4 - '8.5')^2}{'8.5'} + 9.71$		M1 M1
	$= 12.10\dots$ awrt 12.0 to 12.1		A1
	$\nu = 7$		B1
	$\chi^2_7(0.05) = 14.067$		B1ft
	[not in the CR/not significant/Do not reject H_0] There is not sufficient evidence to suggest that $N[(310, 16)]$ is not a suitable model/The model is suitable/The data are consistent with the model		A1
			(7)
(c)	$\nu = 8 - 3 = 5$ / two parameters estimated so additional degrees of freedom subtracted		M1
	Therefore the critical value is reduced/now 11.070		A1
			(2)
Notes			Total 14
(a)	M1	for standardising with 303.5 or 315.5, 310 and 4	
	A1	awrt 0.052 or awrt 0.084/0.085	
	A1	either correct value	
	M1	a complete method to find the second missing value	
	A1	both correct values	
(b)	B1	both hypotheses correct. If mentioning normal, must mention $N(310, 16)$ at least once.	
	M1	for either $\frac{(5 - '5.2')^2}{'5.2'}$ or $\frac{(4 - '8.5')^2}{'8.5'}$	
	M1	for a complete method to find $\sum \frac{(O - E)^2}{E}$ e.g. $9.71 + 2$ additional terms this mark is independent of the 1 st M1	
	A1	allow awrt 12.0 to 12.1	
	B1	$\nu = 7$ This mark can be implied by a correct critical value of 14.067	
	B1ft	14.067 (may see 5.991, 7.815, 9.488, 11.070, 12.592)	
	A1	dependent on 2 nd M1 a correct conclusion which states that the model is suitable and must be consistent with their X^2 value and their χ^2 critical value. If no hypotheses or hypotheses wrong way round do not award.	
(c)	M1	a statement that implies 2 additional degrees of freedom are subtracted	
	A1	a correct conclusion from correct reasoning.	

Question	Scheme	Marks
4 (a)	H_0 : B(4, 0.5) is a suitable model	B1
	H_1 : B(4, 0.5) is not a suitable model	
	Expected frequencies 12.5, 50, 75, 50, 12.5	M1 A1
	$\sum \frac{(O-E)^2}{E} = \frac{(15-12.5)^2}{12.5} + \dots + \frac{(10-12.5)^2}{12.5}$	M1
	or $\sum \frac{O^2}{E} - N = \frac{15^2}{12.5} + \dots + \frac{10^2}{12.5} - 200$	
	$= 10.84 \text{ (or 10.8)}$	A1
	$\nu = 4$	B1
	$\chi^2_4(0.05) = 9.488 \Rightarrow CR \geq 9.488$	B1
	Sufficient evidence to say that the research students claim is not supported	A1ft
		(8)
(b)	$[0 \times 15 + 1 \times 68 + 2 \times 69 + 3 \times 38 + 4 \times 10] = 360$	M1
	$\frac{360}{200 \times 4} = 0.45 *$	A1*
		(2)
(c)	H_0 : Binomial is a suitable model	B1
	H_1 : Binomial is not a suitable model	
	$\nu = 3$	B1
	$\chi^2_3(0.05) = 7.815 \Rightarrow CR \geq 7.815$	B1ft
	No significant evidence to say that the binomial is not a reasonable model	B1ft
		(4)
Notes		Total 14
(a)	B1	Both hypotheses correct. Must mention B(4, 0.5) at least once. (may be in words need Binomial, probability (p) = 0.5 and a reference to 4 children or $n = 4$) Condone B(0.5, 4)
	M1	For a correct method to find at least one expected frequency e.g. $0.5^4 \times 200 [= 12.5]$ or $4 \times 0.5^4 \times 200 [= 50]$ or $6 \times 0.5^4 \times 200 [= 75]$ May be implied by correct answer 10.84 or 10.8
	A1	For all 5 expected frequencies correct. These must be seen and cannot be implied.
	M1	For an attempt at the test statistic, at least 2 correct expressions/ values seen (include $- 200$ if needed)
	M1	$\sum \frac{(O-E)^2}{E} = 0.5 + 6.48 + 0.48 + 2.88 + 0.5$ or $\sum \frac{O^2}{E} - N = 18 + 92.48 + 63.48 + 28.88 + 8 - 200$ May be implied by correct answer 10.84 or 10.8
	A1	10.84 Allow 10.8
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488
	B1	9.488 ft their degrees of freedom if given. For $\nu = 3$ it is 7.815
	A1ft	Dep on the 2 nd M1. independent of hypotheses. Need claim or student or binomial. ft their CV and their test statistic only. A correct conclusion based on their test statistic value and their χ^2 critical value (Allow in terms of Binomial eg does not follow a binomial distribution) If their Test statistic > their CV then must say not supported (not binomial) . If their Test statistic < their CV then must say supported (is binomial)
(b)	M1	A correct method for finding the total number of girls. At least 3 non zero terms correct. useful figures $[0] + 68 + 138 + 114 + 40$. Implied by 360 or 1.8
	A1*	cso allow for $360/800$ or $1.8/4$ or $1.8 = 4p$
(c)	B1	Both hypotheses correct. Must mention binomial at least once. Condone inclusion of B(4,0.45)/B(0.45,4)
	B1	$\nu = 3$ This mark can be implied by a correct critical value of 7.815 Condone (their ν in part(a) – 1)
	B1ft	7.815 ft their degrees of freedom if they have (their ν in part(a) – 1)
	B1ft	Ft their CV only. Independent of hypotheses. A correct conclusion based on awrt 2.47 and their χ^2 critical value only. Ignore any parameter given. Do not allow contradicting statements.

Question Number	Scheme	Notes	Marks																														
5.(a)	Relief of symptoms is either a “success” or a “failure”. The probability the medicine being a success is constant. Samples from different medical practices are independent.	Any 2. Context required in one assumption.	B1 B1																														
			(2)																														
5(b)	Mean = $\frac{0 \times 4 + 1 \times 6 + 2 \times 3 + \dots + 8 \times 2}{50} = 3.54^*$	At least two correct terms on the numerator and 50 on the denominator, fully correct expression or $\frac{177}{50}$ dep on M1 scored cso.	M1, A1cso																														
			(2)																														
5(c)	$p = \frac{3.54}{8} = 0.4425$	Can be implied by at least 1 correct value for f or g .	B1																														
	$f = 50 \times C_4^8 \times 0.4425^4 \times 0.5575^4 = 12.96$ $g = 50 \times 0.4425^8 = 0.07$	Use of Bin(50, p) for M1, Allow awrt 12.96, awrt 0.07	M1A1A1																														
			(4)																														
5(d)	H_0 : Binomial distribution is a suitable model H_1 : Binomial distribution is not a suitable model	Both hypotheses correct If parameters used then B0.	B1																														
	<table><tr><th>No of succe sses</th><th>O</th><th>E</th><th>$\frac{(O-E)^2}{E}$</th><th>$\frac{O^2}{E}$</th></tr><tr><td>0,1,2</td><td>13</td><td>11.66</td><td>0.154</td><td>14.494</td></tr><tr><td>3</td><td>12</td><td>13.07</td><td>0.088</td><td>11.018</td></tr><tr><td>4</td><td>10</td><td>12.96</td><td>0.676/7</td><td>7.716</td></tr><tr><td>5,6,7, 8</td><td>15</td><td>12.31</td><td>0.588/7</td><td>18.278</td></tr><tr><td></td><td>50</td><td>50</td><td>1.506</td><td>51.506</td></tr></table>	No of succe sses	O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$	0,1,2	13	11.66	0.154	14.494	3	12	13.07	0.088	11.018	4	10	12.96	0.676/7	7.716	5,6,7, 8	15	12.31	0.588/7	18.278		50	50	1.506	51.506	Combining 0,1,2 or 5,6,7,8. M1 for attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ with at least 2 correct expressions or 2 correct values to 2sf.	M1
No of succe sses	O	E	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$																													
0,1,2	13	11.66	0.154	14.494																													
3	12	13.07	0.088	11.018																													
4	10	12.96	0.676/7	7.716																													
5,6,7, 8	15	12.31	0.588/7	18.278																													
	50	50	1.506	51.506																													
	$\sum \frac{(O-E)^2}{E} = \sum \frac{O^2}{E} - 50 = 1.50\dots$	awrt 1.5 (calculator: 1.50498...)	A1																														
	$\nu = 4 - 2 = 2, \chi_2^2(10\%) = 4.605$	2 can be implied by 4.605 seen Only f.t. $\nu = r - 2$	B1 B1f.t.																														
	Insufficient evidence to reject H_0	For correct non-contextual statement linking their test statistic and their cv.	M1																														
	Data is consistent with a binomial distribution (oe)	A correct comment suggesting that binomial model is suitable / good fit. Hypotheses wrong way around scores A0 here. Condone parameters here.	A1																														
			(8)																														
			Total 16																														

Question Number	Scheme	Marks																		
5. (a)	H_0 : $N(6, 0.75^2)$ is a suitable model for the length of fallen pine cones H_1 : $N(6, 0.75^2)$ is NOT a suitable model for the lengths of the pine cones e.g. E_i : $5 \leq x < 5.5 = 80 \times P(5 \leq X < 5.5) = 80 \times P(-\frac{4}{3} \leq Z < -\frac{2}{3}) [= 12.77 \sim 12.90]$ or E_i : $6 \leq x < 6.5 = 80 \times P(0 \leq Z < \frac{2}{3}) [= 19.80 \sim 19.89]$ E_i : $5.5 \leq x < 6 = 19.80 \sim 19.89$ or $x \geq 6.5 = 40 - "19.80" = 20.11 \sim 20.20$ <table><tr><td></td><td>$x < 5$</td><td>$5 \leq x < 5.5$</td><td>$5.5 \leq x < 6$</td><td>$6 \leq x < 6.5$</td><td>$x \geq 6.5$</td></tr><tr><td>E_i</td><td>7.30~7.43</td><td>12.77~12.90</td><td>19.80~19.89</td><td>19.80~19.89</td><td>20.11~20.20</td></tr><tr><td>$\frac{(O-E)^2}{E}$</td><td>0.23~0.28</td><td>0.093~0.12</td><td>0.84~0.90</td><td>1.87~1.95</td><td>5.08~5.16</td></tr></table> $\sum \frac{(O_i - E_i)^2}{E_i}$ or $\sum \frac{O_i^2}{E_i} - 80 = 8.308... ;$ answer in [8.15 ~ 8.4] $\nu = 5 - 1 = 4 \Rightarrow \chi_4^2(10\%) = 7.779$ (significant result so) the data do not support Chrystal's belief		$x < 5$	$5 \leq x < 5.5$	$5.5 \leq x < 6$	$6 \leq x < 6.5$	$x \geq 6.5$	E_i	7.30~7.43	12.77~12.90	19.80~19.89	19.80~19.89	20.11~20.20	$\frac{(O-E)^2}{E}$	0.23~0.28	0.093~0.12	0.84~0.90	1.87~1.95	5.08~5.16	B1 M1 A1 M1 A1 dM1; A1 B1; B1ft A1ft (10)
	$x < 5$	$5 \leq x < 5.5$	$5.5 \leq x < 6$	$6 \leq x < 6.5$	$x \geq 6.5$															
E_i	7.30~7.43	12.77~12.90	19.80~19.89	19.80~19.89	20.11~20.20															
$\frac{(O-E)^2}{E}$	0.23~0.28	0.093~0.12	0.84~0.90	1.87~1.95	5.08~5.16															
(b)	$\hat{\mu} = \frac{464}{80} = \underline{5.8}$ (cm) ; $s^2 = \frac{2722.59 - 80 \times "5.8^2"}{79}$ $s^2 = 0.39734... \text{ awrt } \underline{0.397}$ (cm²)	B1; M1 A1 (3)																		
(c)	$\nu = 5 - 3 = 2 ;$ so $\chi_2^2(10\%) = 4.605$ (Not sig') so a normal distribution is a plausible model for length of pine cones	B1; B1ft B1ft (3)																		
(d)	$P(X > 7 \mid \mu = 5.8 \text{ and } s = \sigma = 0.63035...) = P\left(Z > \frac{7 - "5.8"}{\sqrt{0.397..}}\right) = P(Z > 1.90..)$ $= \underline{0.028 \sim 0.029}$	M1 A1 (2) [18m'ks]																		
Notes																				
(a)	1st B1 for both hypotheses. Must include the model and mention “length(s)” and “cones” 1st M1 for correct use of normal to find E_i for one cell 1st A1 for a middle value e.g. awrt 12.77~12.90 inclusive (12.77 is from tables, 12.90 calc) 2nd M1 for use of symmetry to get E_i for $5.5 \leq x < 6$ (same as $6 \leq x < 6.5$) or $x \geq 6.5$ (40 - ...) 2nd A1 for a correct set of expected frequencies (all awrt in given ranges) 3rd dM1 (dep on 1st M1) for a correct attempt to find test statistic...at least one correct term 3rd A1 for answer in the range 8.15-8.4 (inclusive) 2nd B1 for degrees of freedom = 4 3rd B1ft for a correct 10% critical value using their degrees of freedom 4th A1ft dep on M3 and cv = awrt 7.78 for contextual conclusion: length, cones, N (μ, σ not needed) or Chrystal's belief																			
(b)	B1 for 5.8 M1 for a correct expression (ft their mean) A1 for awrt 0.397 (Condone $\frac{3139}{7900}$)																			
(c)	1st B1 for degrees of freedom = 2 2nd B1ft for a correct cv (different from their part (a)) ft their df 3rd B1ft for a correct conclusion in context ft cv (“length” and “cones”) Ignore any μ or σ																			
(d)	M1 for standardising with 7, their 5.8 ($\neq 6$) and their s.d. from (b). Ignore any $\times 80$ A1 for a correct proportion of 0.028 or 0.029. (ISW if correct ans followed by $\times 80$)																			