Qualification Accredited

OCR
Oxford Cambridge and RSA

A LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H645

For first teaching in 2017

Y422/01 Summer 2024 series

Version 1

ocr.org.uk/maths

Contents

Introduction

Our examiners' reports are produced to offer constructive feedback on candidates' performance in the examinations. They provide useful guidance for future candidates.

The reports will include a general commentary on candidates' performance, identify technical aspects examined in the questions and highlight good performance and where performance could be improved. A selection of candidate answers is also provided. The reports will also explain aspects which caused difficulty and why the difficulties arose, whether through a lack of knowledge, poor examination technique, or any other identifiable and explainable reason.

Where overall performance on a question/question part was considered good, with no particular areas to highlight, these questions have not been included in the report.

A full copy of the question paper and the mark scheme can be downloaded from OCR.

Paper Y422 series overview

Y422/01 (Statistics Major) is an optional examined component for GCE Further Mathematics B (MEI). The component provides two thirds of the total mark for the applied part of the course and includes the content of Y432 (Statistics Minor) together with various other topics. The component focuses on:

- Discrete random variables, including the Poisson, geometric and discrete uniform distributions.
- Bivariate data, including Pearson's product moment correlation coefficient, Spearman's rank correlation coefficient and regression analysis.
- Chi-squared tests for contingency tables and for goodness of fit.
- Continuous random variables, including probability density and cumulative distribution functions, the Normal distribution and the continuous uniform distribution.
- Statistical inference including confidence intervals and hypothesis tests, using the Central Limit Theorem if necessary.
- Simulation.

Candidates are expected to know the content of A Level Mathematics and the Core Pure mandatory paper for Further Mathematics (Y420). Candidates should have gained experience during their course of spreadsheets or other software to explore data sets and to conduct hypothesis tests and construct confidence intervals. They should also have had experience of using a spreadsheet to simulate a random variable.

Most of the questions in Y422/01 are in context and many require interpretation in addition to understanding. Questions may also require candidates to comment about the modelling assumptions underlying their answers.

In general candidates did well in questions involving calculations, but rather less well in questions requiring reasons or explanations.

Candidates who did well on this paper generally:	Candidates who did less well on this paper generally:
• used calculators effectively to calculate probabilities, find the equation of a regression line, find expectation and variance and solve simultaneous equations • chose appropriate levels of accuracy for their answers • carried out hypothesis tests giving the hypotheses appropriately in terms of the population, using correct terminology, identifying critical values and giving non-assertive conclusions • applied their knowledge and understanding to new and unfamiliar contexts.	• found the equation of a regression line and used it to predict values • correctly carried out a chi-squared test • did not sufficiently clearly answer questions requiring comments or explanations • made errors in carrying out hypothesis tests such as not mentioning the population, using the wrong critical value or giving a conclusion that was too assertive.

Section A overview

This section consists of more straightforward questions. Most questions in this section were very well answered although some candidates did find Question 3, which required candidates to use information relating to three Normal distributions, challenging .

Question 1 (a)

1 The number of insurance policy sales made per month by a salesperson is modelled by the random variable X, with probability distribution shown in the table.

r	0	1	2	3	4	5	6
$P(X = r)$	0.05	0.1	0.25	0.3	0.15	0.1	0.05

(a) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

This was correctly answered by almost all candidates.

Question 1 (b)

The salesperson is paid a basic salary of £1000 per month plus £500 for each policy that is sold.

(b) Find the mean and standard deviation of the salesperson's monthly salary. **[3]**

Almost all candidates found the correct value of the mean and most also found the standard deviation correctly. Some only found the variance and some multiplied by 500 rather than 500^2 and so only gained credit for the correct mean.

Question 2 (a)

2 The number of cars arriving per minute to queue at a drive-through fast-food restaurant is modelled by the random variable X. The standard deviation of X is 0.6. You should assume that arrivals are random and independent and occur at a constant average rate.

(a) Find the mean of X. **[2]**

This was correctly answered by the vast majority of candidates, although a few thought that the mean was equal to the given standard deviation of 0.6.

Question 2 (b) (i)

(b) (i) Calculate $\mathrm{P}(X = 1)$. **[1]**

This was correctly answered by almost all candidates, with those who has got part (a) wrong almost always following through their answer correctly.

Question 2 (b) (ii)

(ii) Calculate $\mathrm{P}(X > 1)$. **[2]**

This again was correctly answered by most candidates, again with those who had got part (a) wrong almost always following through their answer correctly. A few found $P(X \geq 1)$ rather than $P(X > 1)$.

Question 2 (c)

(c) Find the probability that fewer than 5 cars arrive in a randomly chosen 20-minute period. **[2]**

This was correctly answered by the majority of candidates, following through if necessary. In this part, once a candidate had found the new mean, rather more made a similar error to that in part (b)(ii) and found $P(X \leq 5)$ rather than $P(X < 5)$.

Question 3 (a)

3 At a launderette the process of cleaning a load of clothes consists of three stages: washing, drying and folding. The times in minutes for each process are modelled by independent Normal distributions with means and standard deviations as shown in the table.

	Mean	Standard deviation
Washing	35	2.4
Drying	46	3.1
Folding	12	2.2

(a) Find the probability that drying a randomly chosen load of clothes takes more than 50 minutes. **[1]**

This was correctly answered by almost all candidates.

Question 3 (b)

(b) It is given that for 99% of loads of clothes the washing time is less than k minutes.

Find the value of k. **[1]**

Most candidates answered this correctly although wrong answers were more common than in part (a). Some candidates are presumably rather less familiar with finding an inverse Normal probability.

Question 3 (c)

(c) Determine the probability that the drying time for a randomly chosen load of clothes is less than the total of the washing and folding times. **[3]**

This question on combining Normal distributions was well answered. Those candidates who made errors almost always found the mean correctly but then did not find the correct variance.

Question 3 (d)

(d) Determine the probability that the mean time for cleaning 5 randomly chosen loads of clothes is less than 90 minutes. You should assume that the time for cleaning any load is independent of the time for cleaning any other load. **[3]**

This question was reasonably well answered, and a number of candidates who had the wrong variance in part (c) found the correct variance for the time for one load in this part. This was exactly the same as the variance in part (c). The most common error was to divide the variance for one load by 5^2 instead of just by 5.

Question 4 (a)

4 An archer fires arrows at a circular target of radius 50 cm. The distance in cm that an arrow lands from the centre of the target is modelled by the random variable X, with probability density function given by

$$f(x) = \begin{cases} ax & 0 \leqslant x \leqslant 50, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(a) Determine the value of a. **[2]**

This was correctly answered by almost all candidates.

Question 4 (b)

(b) Determine the probability that an arrow will land within 5 cm of the centre of the target. **[2]**

This was correctly answered by most candidates, although a few integrated from 45 to 50 rather than from 0 to 5.

Question 4 (c)

(c) Determine the median distance from the centre of the target that an arrow will land. **[3]**

Almost all candidates answered this correctly, with some giving their answer in surd form which was acceptable.

Section B overview

This section consists of a mixture of more and less straightforward questions. The only question that proved to be very challenging was Question 11 involving a discrete uniform distribution. Question 9(d) involving comments on the results of a chi-squared test was also fairly challenging. The remainder of the questions were generally well done, apart from some explanation/comment questions each worth at most 2 marks.

Question 5 (a)

5 A researcher is investigating whether doing yoga has any effect on quality of sleep in older people. The researcher selects a random sample of 40 older people, who then complete a yoga course. Before they start the course and again at the end, the 40 people fill in a questionnaire which measures their perceived sleep quality. The higher the score, the better is the perceived quality of sleep.

The researcher uses software to produce a 90% confidence interval for the difference in mean sleep quality (sleep quality after the course minus sleep quality before the course). The output from the software is shown below.

Z Estimate of a Mean ▼

Confidence level 0.9

Sample

Mean 0.586

s 2.14

N 40

Result

Z Estimate of a Mean

Mean	0.586
s	2.14
SE	0.3384
N	40
Lower limit	0.029
Upper limit	1.143
Interval	0.586 ± 0.557

(a) Explain why the confidence interval is based on the Normal distribution even though the distribution of the population of differences is not known. **[2]**

Almost all candidates were given 1 mark for stating that the sample was large. Many of these then mentioned the central limit theorem but only a few mentioned the distribution of the sample mean which was required in order to be given the second mark.

Question 5 (b)

(b) Explain whether the confidence interval suggests that the mean sleep qualities before and after completing a yoga course are different. **[2]**

Many candidates were given 1 mark for stating that zero was not in the confidence interval. Rather fewer gained the second mark, either because there was no uncertainty in their answer or because they mentioned an improvement rather than a difference.

Assessment for learning

Candidates should be advised that when making a conclusion based on a confidence interval, they should very clearly state whether the value of interest (in this case zero) lies within or outside of the interval. This is essential if any marks are to be given.

Question 5 (c) (i)

(c) In the output from the software, SE stands for 'standard error'.

(i) Explain what standard error is. **[1]**

This part was found to be very challenging, with only a few candidates giving a suitable answer, and many simply stating how to calculate standard error.

Question 5 (c) (ii)

(ii) Show how the standard error was calculated in this case. **[1]**

Almost all candidates answered this correctly.

Question 5 (d)

(d) A colleague of the researcher suggests that the confidence level should have been 95% rather than 90%.

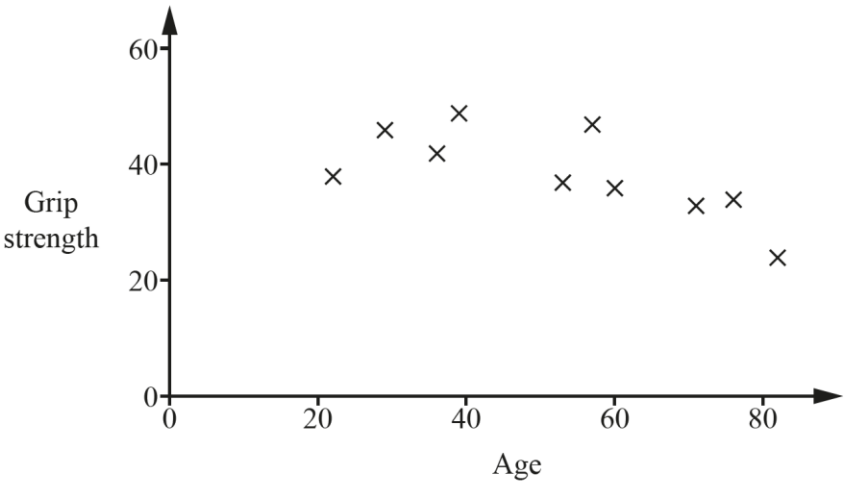
Determine whether this would have made a difference to your answer to part **(b)**. **[4]**

The vast majority of candidates found the new confidence interval correctly and were given 3 marks. The final mark was for stating zero was now in the interval so it had made a difference to the answer. Some did not get this final mark due to comments such as 'it could now have a negative impact', without mentioning that zero was now in the interval.

Question 6 (a)

6 A student is investigating the relationship between age and grip strength in adults. The student selects 10 people and records their ages in years and the grip strengths of their dominant hand, measured in kg. The data are shown in the table below, together with a scatter diagram to illustrate the data.

Age	22	29	36	39	53	57	60	71	76	82
Grip strength	38	46	42	49	37	47	36	33	34	24



The student decides to carry out a hypothesis test to investigate whether there is negative association between age and grip strength.

(a) Explain why the student decides to carry out a test based on Spearman's rank correlation coefficient. **[2]**

Most candidates correctly stated that the scatter diagram did not appear to be elliptical, for which the first mark was available. Fewer candidates mentioned that this suggested that the distribution was bivariate Normally distributed to gain the second mark. Some candidates stated that 'the data is bivariate Normally distributed', which is not correct as data does not have a distribution.

Question 6 (b)

(b) State what property of the sample is required in order for it to be valid to carry out a hypothesis test. **[1]**

Most candidates answered this correctly. A few gave incorrect answers, the most common of which was 'random on random'.

Question 6 (c)

(c) In this question you must show detailed reasoning.

Assuming that the property in part **(b)** holds, carry out the test at the 5% significance level. **[8]**

The modal mark in this question was 7, with almost all of those candidates who were given this mark forgetting to mention population in their hypotheses. Most candidates ranked the data, usually correctly, and then found the value of Spearman's rank correlation coefficient. Most also found the correct critical value, made a comparison and gave a conclusion. Some of these conclusions were rather too assertive and so such candidates were usually given 6 marks in total.

Assessment for learning

In questions involving hypothesis testing, the hypotheses should always mention the population. In this case and in Question 9, 'in the population' should be stated at the end of each hypothesis. If a parametric test is being carried out as in Question 7, the parameter (usually μ) should be defined as the population parameter. Thus in Question 7 (d) the definition of μ should be 'μ is the population mean selenium level in carrots from this field' rather than 'μ is the mean selenium level in carrots from this field'.

The conclusions in such tests should always be given in context. In this case 'There is sufficient evidence to suggest that there is negative association between age and grip strength in the population', rather than just 'There is insufficient evidence to suggest that there is negative association'.

The conclusions should have an element of doubt. A statement such as 'There is insufficient evidence to suggest that there is negative association, and so age and grip strength are negatively associated' is too assertive. Such a statement would not be given the mark for the conclusion due to the final part of the sentence.

Exemplar 1

6(c)

H_0 : There is no association between ~~the relationship age grip~~ age and grip strength in adults.

H_1 : There is a negative association between ~~the~~ age and grip strength in adults.

age: 22 (1) 29 (2) 36 (3) 39 (4) 53 (5) 57 (6) 60 (7) 71 (8) 76 (9) 82 (10)

grip: 38 (6) 46 (8) 42 (7) 49 (10) 37 (5) 47 (9) 36 (4) 33 (2) 34 (3) 24 (1)

$d = 2 - 5$ -6 -4 $= -6$ 0 -3 3 6 6 9

$d^2 = 25$ 36 16 36 0 9 9 36 36 81

$\Sigma d^2 = 284$

$r_s = \dfrac{1 - 6\Sigma d^2}{n(n^2-1)}$ $n = 10$, $\rho = 5\%$

$r_{scrit} = 0.5636$

$= \dfrac{1 - 6 \times 284}{10(10^2-1)}$ -1.7202

-0.5636 0.5636

$= -1.7202$ As $-1.7202 < -0.5636$, Reject H_0,

this is a significant result. There is sufficient evidence to suggest that there is a negative association between age and grip strength in adults.

In this response the candidate does not mention population in stating the hypotheses, and so is given 1 mark out of the possible 2. The data is ranked correctly and $\sum d^2$ is found correctly but then the formula used is wrong. After this error, no more marks can be given, other than for the correct critical value. The candidate should have realised that a value outside the range $[-1, 1]$ is impossible.

Question 7 (a)

7 An environmental investigator wants to check whether the level of selenium in carrots in fields near a mine is different from the usual level in the country, which is 9.4 ng / g (nanograms per gram). She takes a random sample of 10 carrots from fields near the mine and measures the selenium level of each of them in ng / g, with results as follows.

6.20 10.72 11.42 16.32 15.33 10.56 8.83 9.21 7.78 14.32

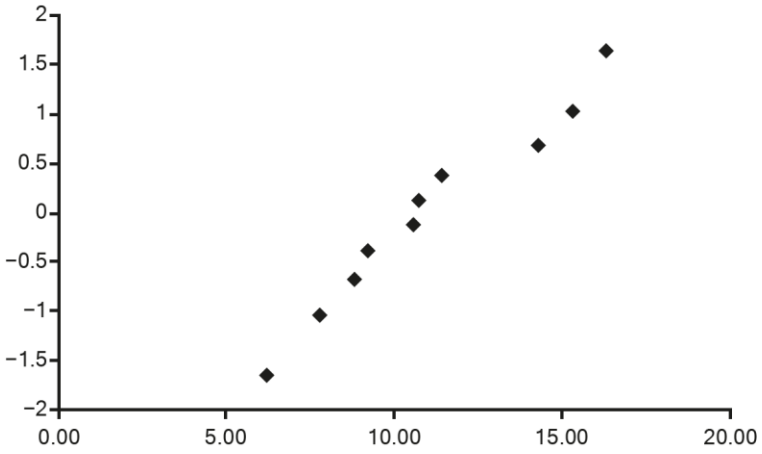
(a) Find estimates of each of the following.

- The population mean
- The population standard deviation **[2]**

Although most candidates gained both marks, a significant minority only got the first mark for the population mean. Most of the latter divided by 10 rather by 9 in their calculation for the standard deviation.

Question 7 (b)

The investigator produces a Normal probability plot and carries out a Kolmogorov-Smirnov test for these data as shown in the diagram.



Kolmogorov-Smirnov test for Normality

p-value = 0.68

(b) Comment on what the Normal probability plot and the p-value of the test suggest about the data. **[3]**

This was very well answered with most candidates given full marks. Some got the first 2 marks only, as they stated 'the data is Normally distributed' or made another incorrect statement.

Question 7 (c)

(c) State the null hypothesis for the Kolmogorov-Smirnov test for Normality. **[1]**

This part caused difficulties for a good number of candidates. The two most commonly seen wrong answers were 'the Normal distribution does not fit the data' and '$\mu = 0$' which is the null hypothesis for the test in part (d).

Question 7 (d)

(d) In this question you must show detailed reasoning.

Carry out a test at the 5% significance level to investigate whether the mean selenium level in carrots from fields near the mine is different from $9.4\,\text{ng/g}$. **[8]**

It was pleasing to see that almost all candidates attempted to perform a *t* test (or at least a Normal test). No marks were available for those very few who attempted a Wilcoxon test. Last year, when a Wilcoxon test was required, a significant number of candidates attempted a *t* test for which again no marks were available.

Many candidates were given at least 7 marks out of 8. Most of those who were given only 7 missed out 'population' in defining μ. Candidates who had got the wrong value of the standard deviation in part (a) could not be given full marks. A significant number of candidates wrongly thought that they could use a Normal distribution, despite the fact that the variance had been estimated.

Some candidates found the probability of finding a result at least as extreme as 11.069 and then compared this to 0.025. If the probability was correct then full marks could be given for this method (provided that everything else was correct and fully explained). However many candidates did not have a correct probability and were often only given the marks for the hypotheses. This method is therefore not recommended. Other candidates found a 95% confidence interval for the population mean (based on $\mu = 9.4$). With this method, even if candidates made an error, their working usually allowed more part marks to be given.

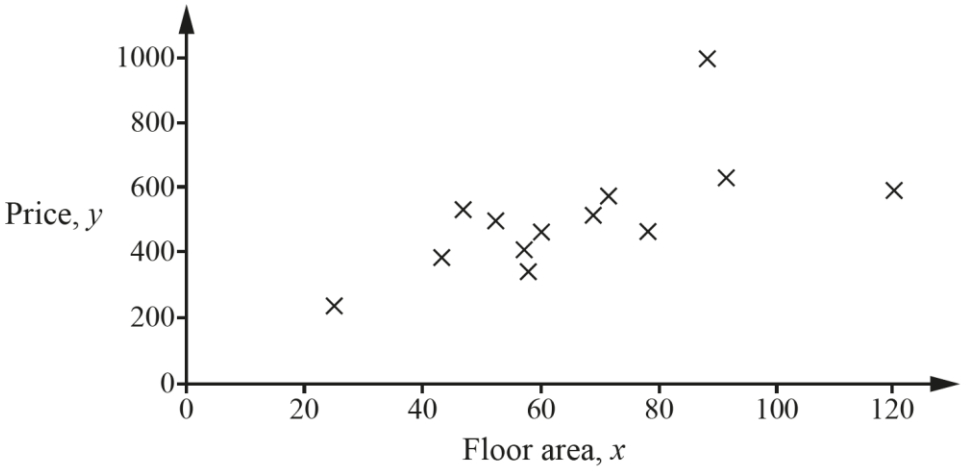
Question 7 (e)

(e) If the *p*-value of the Kolmogorov-Smirnov test for Normality had been 0.007, explain what procedure you could have used to investigate the selenium level in carrots from fields near the mine. **[2]**

Most candidates answered this correctly although a few did not explicitly state that a Normal model would be unlikely to be appropriate and so only gained 1 mark for suggesting a Wilcoxon test..

Question 8 (a)

8 An estate agent collects data for a random selection of 13 flats in order to investigate the link between the floor areas of flats and their price. The scatter diagram shows the floor areas, $x\,\text{m}^2$, and prices, £y thousand, of the 13 flats.



(a) The estate agent notes that two of the data points are outliers. One is Flat A which has a large floor area but is in poor condition. The other is Flat B which has a balcony with a desirable view overlooking the sea.

Label these two data points on the copy of the scatter diagram in the **Printed Answer Booklet**. **[2]**

This was well answered but unfortunately a significant number of candidates simply circled the two points and did not specify which was which.

Question 8 (b)

The estate agent decides to remove these two data points from the analysis. Summary statistics for the remaining 11 flats are as follows.

$\Sigma x = 652.5 \quad \Sigma y = 5067 \quad \Sigma x^2 = 41\,987.35 \quad \Sigma y^2 = 2\,456\,813 \quad \Sigma xy = 315\,928.2$

(b) In this question you must show detailed reasoning.

Calculate the equation of a regression line which is suitable for estimating the price of a flat from its floor area. **[5]**

Most candidates were given full marks. Some used the value $n = 13$ rather than the correct $n = 11$ so were given at most 3 marks out of 5, and a few others made arithmetic errors.

Question 8 (c)

(c) Use the regression line to estimate the price for the following floor areas.

- $40\,\text{m}^2$
- $110\,\text{m}^2$

[2]

Although most candidates knew what to do, the vast majority either gave their answers to too many significant figures or omitted 'thousands', or both. Since the answers were estimates, they should have been given to 2 or at most 3 significant figures.

Assessment for learning

When estimating from a regression line, candidates should never give answers to more than 3 significant figures. They should also give their answers in context so in this case many candidates gave the estimated values of *y* which is of course not the price, but one thousandth of the price.

Question 8 (d)

(d) Given that the value of the product moment correlation coefficient for these 11 data items is 0.765, comment on the reliability of your estimates. **[3]**

Most candidates were given 2 marks, for stating that the first estimate was reasonably reliable due to interpolation, but the second was not due to extrapolation. Very few gained the third mark as most commented only on the value of the product moment correlation coefficient, and not on the closeness of the points to a line of best fit, nor on the value of the coefficient of determination.

Question 8 (e)

(e) The estate agent thinks that he can predict the floor area of a flat from its price, using the equation of the regression line found in part **(b)**.

Comment briefly on the estate agent's idea. **[2]**

Most candidates gained a mark for stating that an *x* on *y* equation was required. Rather fewer put this statement in context to be given the second mark.

Question 9 (a)

9 A cyclist has 3 bicycles, a road bike, a gravel bike and an electric bike. She wishes to know if the bicycle which she is riding makes any difference to whether she reaches a speed of 25 mph or greater on a journey. She selects a random sample of 120 journeys and notes the bicycle and whether or not her maximum speed was 25 mph or greater. She decides to carry out a chi-squared test to investigate whether there is any association between bicycle type and whether her maximum speed is 25 mph or greater. **Tables 9.1** and **9.2** show the data and some of the expected frequencies for the test.

Table 9.1

		Bicycle			
		Road	Gravel	Electric	Total
Maximum speed	Less than 25 mph	2	21	19	42
	25 mph or greater	13	47	18	78
	Total	15	68	37	120

Table 9.2

Expected frequency		Bicycle		
		Road	Gravel	Electric
Maximum speed	Less than 25 mph			12.95
	25 mph or greater			24.05

(a) Complete the table of expected frequencies in the **Printed Answer Booklet**. **[2]**

Almost all candidates were given both marks here.

Question 9 (b)

(b) Determine the contribution to the chi-squared test statistic for the Electric bicycle and maximum speed 25 mph or greater. Give your answer correct to **4** decimal places. **[2]**

This part was again almost always done correctly.

Question 9 (c)

The contributions to the chi-squared test statistic for the remaining categories are shown in **Table 9.3**.

Table 9.3

Contribution to the test statistic		Bicycle		
		Road	Gravel	Electric
Maximum speed	Less than 25 mph	2.0119	0.3294	2.8264
	25 mph or greater	1.0833	0.1774	

(c) In this question you must show detailed reasoning.

Carry out the test at the 5% significance level. **[6]**

This was also very well answered, with the main errors being to give an answer that was too assertive, or occasionally to use the wrong critical value. In this part it was not necessary to mention 'population' but in future questions of this type this may be required.

Question 9 (d)

(d) For each type of bicycle, give a brief interpretation of what the data suggest about maximum speed. **[3]**

Only a few candidates were given full marks in this part. Some simply made statements such as 'more observed than expected for …'. This type of statement does not gain any credit since it is not in context. Candidates must mention bicycles or journeys or equivalent. To get all 3 marks candidates must mention the size of the contributions as well as whether there were more or less than expected with maximum speed $\geq$ or $<$ 25 mph.

Exemplar 2

9(d)

Road bicycle: the bicyle reachd a maxin speed of 25mph or greater on slightly ~~less~~ more journeys then expected and reachd a maxin speed of less then 25mph on slightly ~~more~~ less journeys then expected

Gravel bicycle: the gravel bicyle reached a maxin speed of less thn 25mph on slightly less journeys thn expeted and a maxim speed of ~~slightly~~ 25mph or greater on slightly more journeys thn expected

Electric bicycle: the electric reached maxin speed of less thn 25mph on more journeys than expect and maxin speeds of ~~less~~ than 25mph or greater on less journeys then expected.

This response is given 1 mark out of 3. The comments for 'Electric bicycle' are correct, but those for 'Gravel bicycle' are not, since the contributions for this bicycle are both low, and so the candidate should say that the numbers of journeys are as expected. The comments for 'Road bicycle' are not correct since the candidate states that there a 'slightly' less and 'slightly' more rather than simply less and more. Furthermore the candidate does not mention the size of the contributions, although in this case the candidate would still have only been given 1 mark if contributions had been mentioned as there are errors in both the statements for Road bicycle and also Gravel bicycle. Candidates should **always** refer to the sizes of the contributions when interpreting the results of a chi-squared test. If the contributions are much less than 1 then they should say that the numbers are as expected.

Question 10 (a)

10 Ben takes an underground train to work and back home each day. The waiting time is defined as the time from when he reaches the station platform until he boards the train.

On his way to work the waiting time is X minutes, where X is modelled by a continuous uniform distribution on $[0, 6]$.

On his way back from work, the waiting time is Y minutes, where Y is modelled by a continuous uniform distribution on $[0, 4]$.

Ben's total waiting time for both journeys is Z minutes, where $Z = X + Y$. You should assume that X and Y are independent.

(a) Find $E(Z)$. **[2]**

This was usually answered correctly although a few candidates presumably misread the question and found means of $\frac{6+1}{2}$ and $\frac{4+1}{2}$.

Question 10 (b)

(b) Ben thinks that Z will be well modelled by a continuous uniform distribution on $[0, 10]$.

By considering variances, show that he is not correct. **[3]**

Most candidates were given 2 marks out of 3, correctly finding the variance of Z. Having then found the correct variance of the model, most candidates simply stated that the two variances were not equal. In order to be a good model the two variances must be fairly similar, but not necessarily equal so these candidates could not be given the final mark.

Question 10 (c)

(c) Ben's friend Jamila constructs the spreadsheet below, which shows a simulation of 20 values of X, Y and Z. All of the values have been rounded to 2 decimal places.

	A	B	C
1	***X***	***Y***	***Z***
2	1.17	3.83	5.01
3	2.01	0.81	2.82
4	1.27	1.52	2.78
5	1.41	3.94	5.35
6	4.11	2.94	7.05
7	1.76	0.96	2.72
8	3.29	0.98	4.27
9	0.77	0.22	0.99
10	0.99	1.44	2.43
11	4.79	2.43	7.22
12	3.82	3.93	7.75
13	5.25	2.74	7.99
14	2.64	0.48	3.12
15	1.54	2.18	3.72
16	2.71	1.66	4.36
17	0.04	3.24	3.28
18	5.95	3.12	9.07
19	5.22	1.21	6.42
20	4.16	0.11	4.27
21	1.02	0.99	2.01
22			

Write down an estimate of $P(Z > 6)$. **[1]**

Almost all answers were correct although a very few gave an answer of 0.25, having presumably miscounted the number of values of Z greater than 6.

Question 10 (d)

(d) Use a Normal approximation to determine the probability that Ben's total waiting time when travelling to and from work on 40 days is more than 210 minutes. **[3]**

Most candidates were given all 3 marks and even if they had the wrong values of the mean and/or the median they could still be given 2 marks on follow through.

Question 11 (a)

11 The discrete random variable X has a uniform distribution over the set of all integers between 25 and n inclusive, where n is a positive integer with $n > 25$.

(a) Determine $\mathrm{P}\left(X < \frac{n+25}{2}\right)$ in each of the following cases.

- n is even
- n is odd

[4]

This question proved to be very demanding with many candidates given at most 2 marks. Usually this was for stating that the probability that *n* is even was 0.5 and for using $n - 24$ in their working for *n* is odd.

Question 11 (b)

(b) Determine an expression in terms of n for the variance of the mean of 100 independent values of X. **[2]**

This question was reasonably well answered, with many candidates having a correct expression for the variance of *X*. *C*andidates who had an incorrect expression for the variance of *X* could still get a mark for dividing their variance by 100. However, some multiplied their variance by 100 rather than dividing.

Question 11 (c)

(c) Given that $n = 75$, calculate an estimate of the probability that the mean of 100 independent values of X is less than 48. **[5]**

This question proved to be fairly demanding with many candidates given 3 marks or less. Most found the mean correctly but many had an incorrect value for the variance. However those candidates could still be given a method mark for substituting 75 into their expression from part (b), and a further method mark for stating the distribution (provided that their mean was correct). Those candidates who had the correct distribution often forgot to include a continuity correction; in which case they could still gain a fourth mark for an answer of 0.0871. Some used the wrong continuity correction and so could not be given more than 3 marks.

Question 12 (a)

12 The cumulative distribution function of the continuous random variable X is given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 20, \\ a(x^2 + bx + c) & 20 \leqslant x \leqslant 30, \\ 1 & x > 30, \end{cases}$$

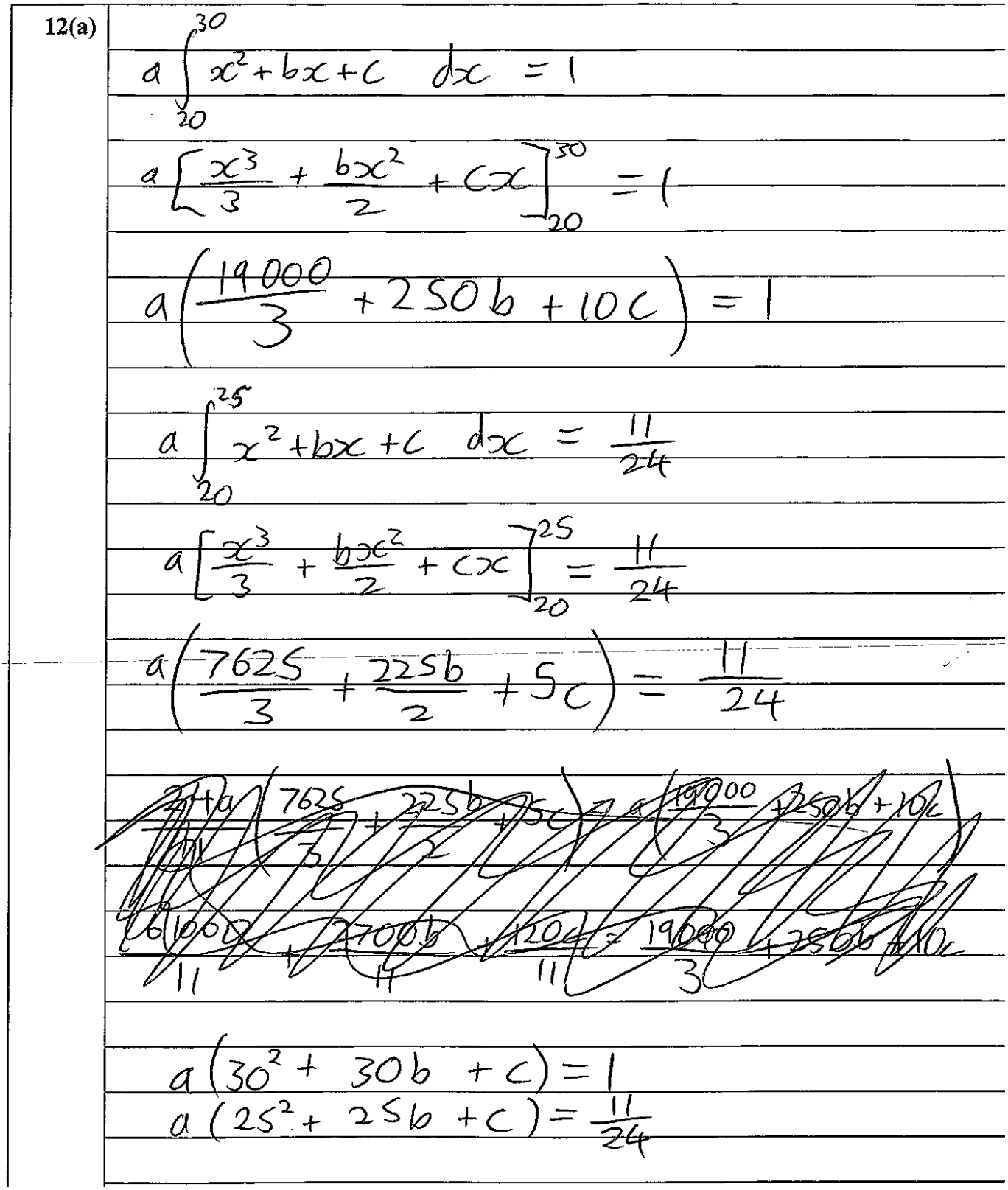
where a, b and c are constants.

You are given that $\mathrm{P}(X < 25) = \frac{11}{24}$.

(a) Find $\mathrm{P}(X > 27)$. **[7]**

Candidates used one of two methods when attempting this question. Candidates could substitute 20, 25 and 30 into the expression for F(x) and equate these to 0, $\frac{11}{24}$ and 1 respectively and then solve to find the values of *a, b* and *c*. Alternatively they could use the fact that $\mathrm{F}(30) - \mathrm{F}(20) = 1$ and $\mathrm{F}(25) - \mathrm{F}(20) = \frac{11}{24}$ and then solve the resulting equations to find the values of *a* and *b*. The two methods were roughly equally popular and many candidates produced a fully correct solution. The main errors were either in trying to solve the equations (despite the fact that calculators can solve simultaneous equations in 3 variables) or in having at least one equation wrong. Those who had the correct values of *a* and *b* (and *c* for the former method), usually found $\mathrm{P}(X < 27)$ correctly and those who did not could still get a method mark for using their values correctly. Some candidates mistakenly tried to integrate the given F(x) as in the exemplar below.

Exemplar 3



In this response the candidate perhaps did not realise the meaning of the cumulative distribution function, or perhaps they gave two different responses in the hope that one was correct. Their first attempt was incorrect since they have integrated the cumulative distribution function. Their second method was correct although it does not get to an answer. Since neither method was crossed out, the final attempt is marked. Luckily for this candidate the final method was correct and they were given 3 marks.

Question 12 (b)

(b) Find the 90th percentile of X. **[2]**

Those candidates who had the correct values in part (a) usually found the 90th percentile correctly. Those who did not do this were often given a method mark for using their values correctly.

Supporting you

Teach Cambridge

Make sure you visit our secure website **Teach Cambridge** to find the full range of resources and support for the subjects you teach. This includes secure materials such as set assignments and exemplars, online and on-demand training.

Don't have access? If your school or college teaches any OCR qualifications, please contact your exams officer. You can forward them this link to help get you started.

Reviews of marking

If any of your students' results are not as expected, you may wish to consider one of our post-results services. For full information about the options available visit the OCR website.

Access to Scripts

We've made it easier for Exams Officers to download copies of your candidates' completed papers or 'scripts'. Your centre can use these scripts to decide whether to request a review of marking and to support teaching and learning.

Our free, on-demand service, Access to Scripts is available via our single sign-on service, My Cambridge. Step-by-step instructions are on our website.

Keep up-to-date

We send a monthly bulletin to tell you about important updates. You can also sign up for your subject specific updates. If you haven't already, sign up here.

OCR Professional Development

Attend one of our popular professional development courses to hear directly from a senior assessor or drop in to a Q&A session. Most of our courses are delivered live via an online platform, so you can attend from any location.

Please find details for all our courses for your subject on **Teach Cambridge**. You'll also find links to our online courses on NEA marking and support.

Signed up for ExamBuilder?

ExamBuilder is a free test-building platform, providing unlimited users exclusively for staff at OCR centres with an Interchange account.

Choose from a large bank of questions to build personalised tests and custom mark schemes, with the option to add custom cover pages to simulate real examinations. You can also edit and download complete past papers.

Find out more.

Active Results

Review students' exam performance with our free online results analysis tool. It is available for all GCSEs, AS and A Levels and Cambridge Nationals (examined units only).

Find out more.

You will need an Interchange account to access our digital products. If you do not have an Interchange account please contact your centre administrator (usually the Exams Officer) to request a username, or nominate an existing Interchange user in your department.

Need to get in touch?

If you ever have any questions about OCR qualifications or services (including administration, logistics and teaching) please feel free to get in touch with our customer support centre.

Call us on
01223 553998

Alternatively, you can email us on
support@ocr.org.uk

For more information visit
ocr.org.uk/qualifications/resource-finder
ocr.org.uk
facebook.com/ocrexams
twitter.com/ocrexams
instagram.com/ocrexaminations
linkedin.com/company/ocr
youtube.com/ocrexams

We really value your feedback

Click to send us an autogenerated email about this resource. Add comments if you want to. Let us know how we can improve this resource or what else you need. Your email address will not be used or shared for any marketing purposes.

I like this **I dislike this**

Please note – web links are correct at date of publication but other websites may change over time. If you have any problems with a link you may want to navigate to that organisation's website for a direct search.

CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT