Qualification Accredited

Oxford Cambridge and RSA

A LEVEL

Examiners' report

FURTHER MATHEMATICS B (MEI)

H645

For first teaching in 2017

Y422/01 Summer 2023 series

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 Y422/01 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:**
• identified the correct hypothesis test to carry out and then used correct terminology and gave non-assertive conclusions • identified appropriate probability distributions to use • used calculators effectively to calculate probabilities, find the product moment correlation coefficient and to find means and standard deviations • applied their knowledge and understanding to new and unfamiliar contexts • chose appropriate levels of accuracy to give answers to and in particular gave probabilities as exact answers or to four decimal places.	• correctly carried out calculations involving the geometric distribution • given summary statistics, correctly found a confidence interval • given summary statistics, correctly found the value of the product moment correlation coefficient • correctly carried out a chi-squared test on a contingency table.

Section A overview

This section consists of more straightforward questions. Most questions in this section were very well answered apart from Question 1 (b) which required an explanation and Questions 1 (d) and 3 (d) which were more involved probability questions.

Question 1 (a)

1 A website simulates the outcome of throwing four fair dice. Ten thousand people take part in a challenge using the website in which they have one attempt at getting four sixes in the four throws of the dice. The number of people who succeed in getting four sixes is denoted by the random variable X.

(a) Show that, for each person, the probability that the person gets four sixes is equal to $\frac{1}{1296}$. **[1]**

This was correctly answered by almost all candidates.

Question 1 (b)

(b) Explain why you could use either a binomial distribution or a Poisson distribution to model the distribution of X. **[3]**

Most candidates scored 1 or 2 marks out of 3 in this part. For the binomial distribution, most mentioned independence and often constant probability but rather less included fixed number of trials, only 2 outcomes and very few stated that the random variable is the number of successes. For the Poisson distribution the correct response of large number of trials and small probability was quite common, but many candidates instead wrongly stated that this distribution was suitable either because the mean and variance were similar or that there was a constant rate of occurrence.

Exemplar 1

1(b)	Each throw is random and independent, the number of people who get 4 sixes (which is a 'success') is being counted, and the probability of getting 4 sixes is constant, $\left(\frac{1}{1296}\right)$, so binomial is appropriate. ~~The~~ A Poisson distribution is appropriate as n is large (10000) and p is small $\left(\frac{1}{1296}\right)$.

This exemplar shows a typical response, where the candidate has a suitable explanation for why a Poisson distribution is suitable, but has only 3 of the 5 possible statements for why a binomial distribution is suitable. It should be noted that only 4 of the 5 statements were required to gain both marks for the binomial distribution.

Question 1 (c)

(c) Use a Poisson distribution to calculate each of the following probabilities.

- $\mathrm{P}(X = 10)$
- $\mathrm{P}(X > 10)$ **[3]**

This was very well answered with most candidates using the correct distribution, although some thought that $\mathrm{P}(X > 10) = 1 - \mathrm{P}(X \leq 9)$.

Question 1 (d)

(d) In another challenge on the website, 50 people are each given 20 independent attempts to try to get four sixes as often as they can.

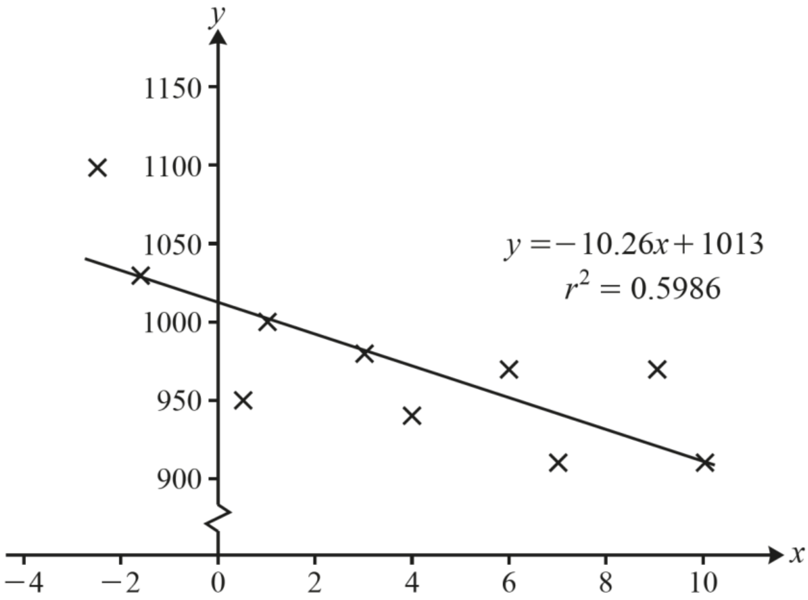
Determine the probability that no more than 2 people succeed in getting four sixes at least once in their 20 attempts. **[3]**

Candidates found this part more demanding, with most either producing an entirely correct response, or something that gained no credit at all. Of the latter candidates, some had little idea of how to start the question and others used a wrong distribution, usually B(50, 20/1296)

Question 2 (a)

2 A student is investigating the link between temperature and electricity consumption in the winter months. The student finds the average minimum temperature, x°C, from across the country on a day. The student then finds the total electricity consumption for that day, y GWh.

The scatter diagram below shows the values of x and y obtained from a random sample of 10 winter days. It also shows the equation of the regression line of y on x and the value of r^2, where r is the product moment correlation coefficient.



(a) Use the regression line to estimate the electricity consumption at each of the following average minimum temperatures.

- 5°C
- −4°C **[2]**

This was almost always correctly answered, except that many candidates gave their answers to too great a degree of precision, usually giving the answer for the estimate for 5°C as 1054.04. Candidates should realise that estimates should not be too precise.

Question 2 (b)

(b) Comment on the reliability of your estimates. **[3]**

A score of 2 marks out of 3 was by far the most common here. Candidates usually mentioned interpolation and extrapolation, but relatively few mentioned the value of r^2 or the closeness of the points to the line. In future such candidates may only gain 1 mark out of 3 rather than the 2 marks that they gained on this occasion.

Question 3 (a)

3 A tennis player is practising her serve. Each time she serves, she has a 55% chance of being successful (getting the serve in the required area without hitting the net). You should assume that whether she is successful on any serve is independent of whether she is successful on any other serve.

(a) Find the probability that the player is not successful in any of her first three serves. **[1]**

This was usually answered correctly.

Question 3 (b)

(b) Determine the probability that the player is successful at least 10 times in her first 20 serves. **[2]**

This was again usually answered correctly.

Question 3 (c)

(c) Determine the probability that the player is successful for the first time on her fifth serve. **[2]**

Almost all candidates answered this correctly.

Question 3 (d)

(d) Determine the probability that the player is successful for the fifth time on her tenth serve. **[2]**

Candidates found this part more challenging, since they had to first use a binomial distribution to find the probability of 4 successes in 9 serves and then multiply by the probability of success in one serve. Many simply had $0.45^5 \times 0.55^5$, although sometimes there was a wrong coefficient added.

Question 3 (e)

Another player is also practising his serve. Each time he serves, he has a probability p of being successful. You should assume that whether he is successful on any serve is independent of whether he is successful on any other serve.

The probability that he is successful for the first time on his second serve is 0.2496 and the probability that he is successful on both of his first two serves is less than 0.25.

(e) Determine the value of p. **[3]**

Despite the more involved nature of this question, it was usually answered correctly. A few candidates had a correct equation for *p* but made an error in solving it.

Section B overview

This section consists of less of the straightforward questions, with more problem solving and interpretation. Questions 4, 5, 6, 9 and 10 (apart from 10 (f)) were generally very well answered, whereas Questions 7, 8 and 11 were found to be the more challenging.

Question 4 (a)

4 A machine manufactures batches of 100 titanium sheets. The thickness of every sheet in a batch is Normally distributed with mean μ mm and standard deviation 0.03 mm. You should assume that each sheet is of uniform thickness and that the thicknesses of different sheets are independent of each other.

The values of μ for three different batches, A, B and C, are 3.125, 3.117 and 3.109 respectively.

(a) Determine the probability that the total thickness of 10 sheets from Batch A is less than 31.0 mm. **[3]**

This question on the Normal distribution was usually answered correctly. The main error was to use a variance of $10^2 \times 0.03^2$ rather than 10×0.03^2.

Question 4 (b)

(b) Determine the probability that, if a single sheet from Batch A is cut into pieces and 10 of the pieces are stacked together, the total thickness of the stack is less than 31.0 mm. **[2]**

This question on the Normal distribution was also usually answered correctly. The main error was the reverse of that in part (a) using a variance of 10×0.03^2 rather than $10^2 \times 0.03^2$.

Question 4 (c)

(c) Determine the probability that, if one sheet from each of Batches A, B and C are stacked together, the total thickness of the stack is at least 9.4 mm. **[3]**

Candidates on the whole did better in this part than in the first two parts of the question, with most gaining full marks. Almost all found the mean correctly and the main error was in finding the variance, sometimes simply using 0.03^2 or alternatively $3^2 \times 0.03^2$ rather than the correct 3×0.03^2.

Question 4 (d)

(d) Determine the probability that the total thickness of 10 sheets from Batch A is less than the total thickness of 10 sheets from Batch B. **[3]**

This part was slightly more difficult than the previous parts since it involved the difference of two Normal distributions. Most candidates coped very well with this part also. Almost all found the correct mean of ±0.08, but rather more had problems in finding the variance, with a variety of errors seen. In this part candidates who tried to subtract the two variances rather than add them should have realised that this was wrong since the result would have been zero.

Question 5 (a)

5 Amari is investigating how accurately people can estimate a short time period. He asks each of a random sample of 40 people to estimate a period of 20 seconds. For each person, he starts a stopwatch and then stops it when they tell him that they think that 20 s has elapsed. The times which he records are denoted by x s. You are given that

$$\sum x = 765, \quad \sum x^2 = 15\,065.$$

(a) Determine a 95% confidence interval for the mean estimated time. **[6]**

This part was very well answered. The main errors seen were to make a mistake in finding the variance, often using a denominator of 40 rather than 39, or to use 1.645 rather than 1.96. A few candidates used the wrong formula, usually simply using the standard deviation rather than the standard error.

Question 5 (b)

(b) Amari says that the confidence interval supports the suggestion that people can estimate 20 s accurately.

Make **two** comments about Amari's statement. **[2]**

Most candidates stated that 20 was in the interval although some made a statement that was too assertive. However, the second statement was often either incorrect or not clear enough.

Question 5 (c)

(c) Discuss whether you could have constructed the confidence interval if there had only been 10 people involved in the experiment. **[2]**

There were not many fully correct answers to this part. Many candidates realised that the confidence interval could not be formed in the same way as that with 40 people. Some simply said that it was not possible to construct a confidence interval. Others said that the *t* distribution should be used, but some of these made no mention of the requirement that the parent population should be Normally distributed. Few made a correct statement about using the *t* distribution and also stated that otherwise a confidence interval could not be formed.

Question 5 (d)

Amari thinks that people would be able to estimate more accurately if he gave them a second attempt. He repeats the experiment with each person and again records the times. Software is used to produce a 95% confidence interval for the mean estimated time. The output from the software is shown below.

Z Estimate of a Mean ▼

Confidence level 0.95

Sample

Mean	19.68
s	1.38
N	40

Result

Z Estimate of a Mean

Mean	19.68
s	1.38
SE	0.2182
N	40
Interval	19.68 ± 0.4277

(d) State the confidence interval in the form $a < \mu < b$. **[1]**

This very straightforward part was almost always answered correctly.

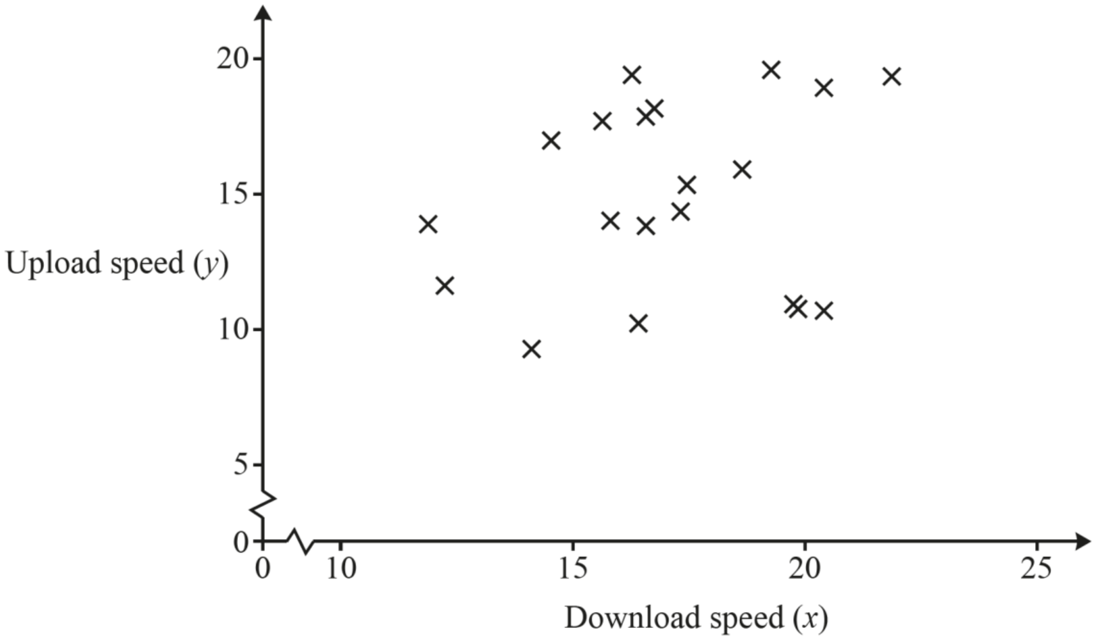
Question 5 (e)

(e) Make **two** comments based on this confidence interval about Amari's opinion that second attempts result in more accurate estimates. **[2]**

This part produced a very similar response to part (b), although rather less stated that 20 was still in the interval than in the previous part. However rather more commented correctly about the lower variance. Any comment suggesting less spread was acceptable.

Question 6 (a)

6 A student wonders if there is any correlation between download and upload speeds of data to and from the internet. The student decides to carry out a hypothesis test to investigate this and so measures the download speed x and upload speed y in suitable units on 20 randomly chosen occasions. The scatter diagram below illustrates the data which the student collected.



(a) Explain why the student decides to carry out a test based on the product moment correlation coefficient. **[2]**

This is a standard question and it was well done. Candidates should be advised however that a statement such as 'The data is bivariate Normal' is not correct. A statement such as 'The underlying population can be assumed to have a bivariate Normal distribution', is what is required.

Question 6 (b)

Summary statistics for the 20 occasions are as follows.

$$\sum x = 342.10 \quad \sum y = 273.65 \quad \sum x^2 = 5989.53 \quad \sum y^2 = 3919.53 \quad \sum xy = 4713.62$$

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

Calculate the product moment correlation coefficient. **[4]**

This was very well answered.

Question 6 (c)

(c) Carry out a hypothesis test at the 5% significance level to investigate whether there is any correlation between download speed and upload speed. **[5]**

The majority of candidates gained at least 4 marks out of 5, with the most common error being the omission of the definition of the correlation coefficient ρ, not mentioning 'population' in the definition or the definition not being in context. Candidates who gave the hypotheses in words rarely scored any marks as again 'population' was not usually mentioned. A few candidates had the wrong critical value, but almost all knew whether or not the null hypothesis should be rejected, whether from correct values or on follow through from their incorrect values. Most gave a non-assertive conclusion.

Assessment for learning

In questions involving hypothesis testing, the hypotheses should always mention the population. If a parametric test is being carried out, the parameter (in this case ρ) should be defined as the population parameter. Thus the definition of ρ should be 'ρ is the population product moment correlation coefficient between download speed and upload speed'. In the case of non-parametric tests, the hypotheses are usually given in words, but they must again include the word 'population'.

The conclusions in such tests should always be given in context. In this case 'There is insufficient evidence to suggest that there is correlation between download speed and upload speed in the population', rather than just 'There is insufficient evidence to suggest that there is correlation.

The conclusions should have an element of doubt. A statement such as 'There is insufficient evidence to suggest that there is correlation, and so download speed and upload speed are not correlated' is too assertive. Such a statement would not be credited with the mark for the conclusion due to the final part of the sentence.

Question 6 (d)

(d) Both of the variables, download speed and upload speed, are random.

Explain why, if download speed had been a non-random variable, the student could not have carried out the hypothesis test to investigate whether there was any correlation between download speed and upload speed. **[1]**

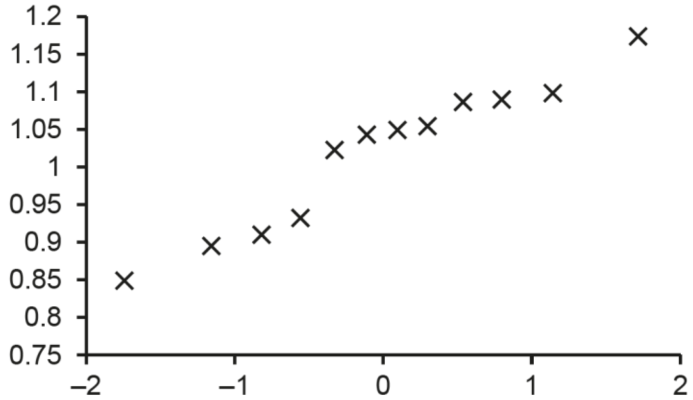
Less than half of candidates gave a correct response with many simply stating that the conditions for a test based on the product moment correlation coefficient would not hold.

Question 7 (a)

7 An analyst routinely examines bottles of hair shampoo in order to check that the average percentage of a particular chemical which the shampoo contains does not exceed the value of 1.0% specified by the manufacturer. The percentages of the chemical in a random sample of 12 bottles of the shampoo are as follows.

1.087 1.171 1.047 0.846 0.909 1.052 1.042 0.893 1.021 1.085 1.096 0.931

The analyst uses software to draw a Normal probability plot for these data, and to carry out a Normality test as shown below.



Kolmogorov-Smirnov Test for Normality

p-value = 0.258

(a) The analyst is going to carry out a hypothesis test to check whether the average percentage exceeds 1.0%.

Explain which test the analyst should use, referring to each of the following.

- The Normal probability plot
- The *p*-value of the Kolmogorov-Smirnov test **[3]**

All that was required in this part was to state that the *p*-value was not very low and the plot was approximately linear which both suggested that the parent population is Normally distributed. Thus the *t* distribution should be used. It is surprising therefore that under half of the candidature scored full marks. Many either mentioned the *p*-value or the straightness but not both. A number of candidates thought that Wilcoxon test should be used, sometimes thinking the distribution was Normal and sometimes not. Some candidates stated that as the *p*-value was more than 0.05 a Wilcoxon test should be used. There was no credit available in part (b) for the use of a Wilcoxon test. This is a standard question and there have been similar questions in a practice paper and in June 2019 and June 2022. The examiners' report made clear on both occasions that no marks could be given for the test itself if the wrong test was chosen.

Question 7 (b)

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

Carry out the test at the 5% significance level. **[10]**

This question was well done (apart from those candidates who thought that a Wilcoxon test should be used). As with Question 6 (c), some candidates made similar errors in stating the hypotheses. Almost all found the sample mean and sample standard deviation correctly, and most then correctly found the test statistic. Some candidates used the wrong critical value, often correctly based on a *t* distribution with 11 degrees of freedom, although sometimes 10 or 12. Most then compared their test statistic to their critical value and came to the correct conclusion.

Question 8 (a)

8 The random variable X has a continuous uniform distribution over [0, 10].

(a) Find the probability that, if two independent values of X are taken, one is less than 3 and the other is greater than 3. **[2]**

This question, which was expected to be very straightforward, proved rather difficult with under a quarter of candidates gaining both marks. Many did not read the question properly and used a discrete uniform distribution, for which no marks were available. Those who did use the correct distribution often omitted the 2 ways in which the events could take place.

Question 8 (b)

The random variable T denotes the sum of 5 independent values of X.

(b) State the value of $P(T \leqslant 25)$. **[1]**

The majority of candidates correctly answered this apparently simple question, with $\frac{5}{11}$ and $\frac{6}{11}$ being the most common wrong responses.

Question 8 (c)

The spreadsheet below shows the heading row and the first 20 data rows from a total of 100 data rows of a simulation of the distribution of X. Each of the 100 rows shows a simulation of 5 independent values of X, together with T, the sum of the 5 values. All of the values have been rounded to 2 decimal places.

In column I the spreadsheet shows the number of values of T that are less than or equal to the corresponding values in column H. For example, there are 75 simulated values of T that are less than or equal to 30.

	A	B	C	D	E	F	G	H	I
1	X_1	X_2	X_3	X_4	X_5	T		t	Number $\leqslant t$
2	3.73	6.65	4.93	0.41	9.33	25.06		0	0
3	4.95	6.58	4.48	2.51	7.26	25.79		5	0
4	8.10	4.87	4.26	3.83	0.79	21.85		10	1
5	6.70	4.10	5.10	1.82	6.76	24.48		15	4
6	3.73	8.38	8.49	9.87	1.31	31.79		20	23
7	3.22	4.36	0.12	1.34	9.49	18.53		25	48
8	9.17	7.13	5.47	4.35	2.44	28.55		30	75
9	3.42	1.93	6.04	2.99	8.85	23.24		35	93
10	0.98	0.68	9.82	9.83	7.28	28.58		40	99
11	5.86	1.67	7.77	4.08	7.14	26.52		45	100
12	9.20	0.31	5.82	5.31	6.45	27.10		50	100
13	7.04	4.30	2.06	0.06	4.16	17.62			
14	0.31	5.02	1.48	5.37	1.77	13.94			
15	3.77	6.04	1.21	7.67	5.01	23.69			
16	1.21	5.54	1.90	1.43	6.91	17.00			
17	9.27	1.98	5.80	9.37	9.34	35.76			
18	4.30	5.66	2.80	1.56	1.19	15.51			
19	7.15	3.19	6.89	5.41	2.18	24.82			
20	6.18	6.32	3.01	6.49	9.12	31.13			
21	5.03	5.99	5.19	6.97	3.55	26.73			

(c) Use the spreadsheet output to estimate each of the following.

- $P(T \leqslant 25)$
- $P(T > 35)$ **[2]**

Once again the majority of candidates had both answers correct. The most common wrong method was to use the uniform distribution to try to answer the question, usually getting answers of 0.5 and 0.05. As mentioned in the Paper Y422/01 series overview candidates should have had experience of using a spreadsheet to simulate a random variable. The responses to this question part would suggest that some have not had sufficient experience.

Question 8 (d)

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

The random variable Y is the mean of 100 independent values of T.

Determine an estimate of $\mathrm{P}(Y > 26)$. **[7]**

Although a minority of the candidature gained no credit, almost half gained at least 6 marks out of the 7 available. Of these latter, the main error of those who did not get full marks, was to simply state that Var(X)=$\frac{25}{3}$ without any explanation. This was a question where candidates were told to show detailed reasoning and so an explanation of Var(X) = $\frac{1}{12}(10\text{-}0)^2$ was required. For those who gained less than 6 marks, there was a wide variety of errors, but many realised that they had to use Var(Y) = Var(T)/100 and also to use the central limit theorem. Some tried to include a continuity correction but of course this was wrong since the original uniform distribution was already continuous. A few candidates tried to use the distribution of the total of 100 independent values of T, and find P(Total >2600) using N(2500, $\frac{12500}{3}$), but most of these made errors, usually in finding the variance..

Question 9 (a)

9 A cyclist who lives on an island suspects that car drivers with locally registered number plates allow more space when passing her than those with non-locally registered number plates. She decides to carry out a hypothesis test and so over a period of time selects a random sample of 250 cars which pass her. For each car she estimates whether the car driver allows at least the recommended 1.5 metres when passing her. The table shows the data which she collected.

		Where registered	
		Local	Non-local
Passing distance	Under 1.5 m	12	11
	At least 1.5 m	157	70

(a) **In this question you must show detailed reasoning.**

Carry out the test at the 5% significance level to examine whether there is any association between where the car is registered and passing distance. **[9]**

This question was generally well done, with most candidates getting at least 8 marks out of 10. For those who did not gain full credit there was a wide variety of errors including: having an over-assertive final conclusion, using the wrong critical value (although almost always with the correct number of degrees of freedom), making a mistake in calculating the test statistic, not showing working when finding the test statistic, giving the hypotheses the wrong way around and very occasionally stating them in terms of correlation instead of association.

Question 9 (b)

(b) A friend of the cyclist suggests that there may be a problem with the data, since the cyclist may have introduced some bias in estimating whether cars were allowing the recommended distance.

Explain how any bias might have arisen. **[1]**

Approximately half of candidates gained the 1 mark available here for a comment such as 'The cyclist may have been more predisposed to estimating locally registered cars passing with at least 1.5 m distance'. Those who did not gain the mark usually made a comment about estimates not being accurate, but this comment would not have necessarily led to bias since it would have affected all cars equally.

Question 10 (a)

10 The continuous random variable X has probability density function given by

$$\mathrm{f}(x) = \begin{cases} \dfrac{4}{15}\left(\dfrac{a}{x^2} + 3x^2 - \dfrac{7}{2}\right) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a positive constant.

(a) Find the cumulative distribution function of X in terms of a. **[4]**

Almost all candidates integrated correctly to find an expression for F(x), but many then some simply stopped there, without any mention of a constant of integration, or simply had '+ c' at the end of their expression. Of those who did try to find the constant of integration, most were successful although some did not give the final full answer. Most successful candidates integrated between 1 and x, but some correctly found the constant by using F(2) = 1.

Exemplar 2

10(a)

$$\mathrm{F}(x) \begin{cases} 0 & x \leqslant 1 \\ \dfrac{4}{15}\left(-\dfrac{a}{x} + x^3 - \dfrac{7x}{2}\right) & 1 \leqslant x \leqslant 2 \\ 1 & x \geqslant 2 \end{cases}$$

$$\frac{4}{15}\int_{-\infty}^{x} \frac{a}{x^2} + 3x^2 - \frac{7}{2}$$

$$\frac{4}{15}\left[-\frac{a}{x} + x^3 - \frac{7x}{2}\right]$$

In this exemplar the candidate has correctly integrated f(x) but then has not included a constant, nor have they integrated with limits. The candidate did know that the F(x) = 0 for $x < 1$ and F(x) = 1 for $x > 2$ but there was no credit for this unless a constant of integration was included.

Question 10 (b)

(b) Hence or otherwise determine the value of a. **[2]**

This was generally very well done, with many candidates using F(2) = 1 to find *a*. An alternative method, which could get full marks even if their F(*x*) had no constant, was to use F(2) - F(1) = 1.

Question 10 (c)

(c) Show that the median value m of X satisfies the equation

$$8m^4 - 28m^2 + 9m - 4 = 0.$$ **[2]**

This question was generally well answered, and candidates who had an incorrect expression for F(*x*) could still get a method mark. Candidates almost universally showed adequate working.

Question 10 (d)

(d) Verify that the median value of X is 1.74, correct to **2** decimal places. **[2]**

Many candidates unfortunately simply used a calculator to solve the equation to several decimal places to try to verify the statement. This approach was not satisfactory. Those who realised what they had to do were almost always successful, with most evaluating the given expression at 1.735 and 1.745 to show a change of sign, with only a few using the main method in the mark scheme of evaluating F(1.735) and F(1.745).

Question 10 (e)

(e) Find $\mathrm{E}(X)$. **[2]**

Only half of the candidature scored full marks here, although many who only gained the first mark simply made an error in integrating. This was a 'Find' question, so candidates could simply use a calculator to carry out the integration, but good practice is to state the integration to be done, especially since many of the numerical values required can only be FT if seen in the working.

Question 10 (f)

(f) Determine the mode of X. **[3]**

This question was not successfully answered with most candidates getting either 0 or 1 mark. The most common method to find the mode was to first differentiate and solve f'(x) = 0. The solutions to this equation, $x = \pm\ 0.639$, were not in the domain of f(x) and so it should have been realised that neither of these was the mode, but many stated that the mode was 0.639. Some candidates did realise that 0.639 was not correct, but then simply stated that the mode was 2 without any further explanation. Since the question requires the mode to be 'determined' rather than 'found', more working was required and this was rarely seen. Only a very few gave a full explanation of why the mode is equal to 2.

Question 11 (a)

11 The random variable X takes the value 1 with probability p and the value 0 with probability $1-p$.

(a) Find each of the following.

- E(X)
- Var(X) **[3]**

The majority of candidates gained full credit on this part. Most of the rest scored 0, either by omitting the question altogether or by including n and so giving answers E(X) = np and Var(X) = np(1-p).

Question 11 (b)

(b) The random variable $Y \sim \mathrm{B}(50, 0.2)$ has mean μ and variance σ^2.

Use the results of part **(a)** to **prove** that

- $\mu = 10$
- $\sigma^2 = 8$. **[6]**

This question was the most difficult on the paper, with only a small number of candidates getting full credit and around half gaining no marks at all. Of those who had partial marks, many correctly used the given value of $p = 0.2$ to find the expectation and variance with the formulae in part (a). Some of these then clearly stated that E(Y) = 50 × E(X) and Var(Y) = 50 × Var(X) and so gained SC marks. Those candidates who simply used the formulae E(X) = np and Var(X) = npq did not get any credit. Only a few stated that $Y = X_1 + X_2 + \ldots + X_{50}$, as in the mark scheme.

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

For the June 2023 series, Exams Officers will be able to download copies of your candidates' completed papers or 'scripts' for all of our General Qualifications including Entry Level, GCSE and AS/A Level. 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 CPD 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 the question builder platform for a range of our GCSE, A Level, Cambridge Nationals and Cambridge Technicals qualifications. Find out more.

ExamBuilder is **free for all OCR centres** with an Interchange account and gives you unlimited users per centre. We need an Interchange username to validate the identity of your centre's first user account for ExamBuilder.

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.

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.

Find out more.

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.