OCR
Oxford Cambridge and RSA

Friday 16 June 2023 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

9970413141

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for Further Mathematics B (MEI)
- a scientific or graphical calculator

QP

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined pages at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give your final answers to a degree of accuracy that is appropriate to the context.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **120**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **12** pages.

ADVICE
- Read each question carefully before you start your answer.

© OCR 2023 [H/508/5594]
DC (LK/SG) 319813/3

OCR is an exempt Charity

Section A (25 marks)

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]**

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

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

- $P(X = 10)$
- $P(X > 10)$ **[3]**

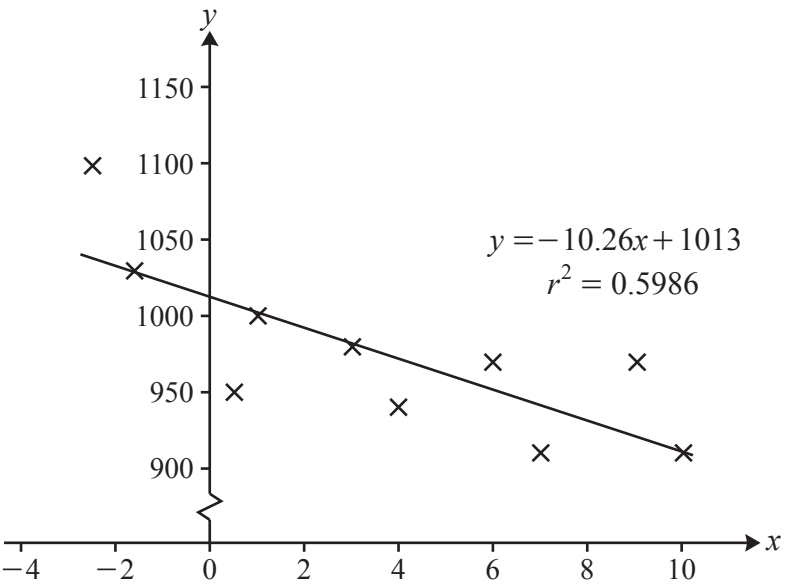
(d) In another challenge on the website, 50 people are each given 20 independent attempts to try to get four sixes as often as they can.

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

© OCR 2023

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]**

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

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]**

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

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

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

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]**

© OCR 2023

Section B (95 marks)

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]**

(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]**

(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]**

(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]**

© OCR 2023

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

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

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

(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]**

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

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

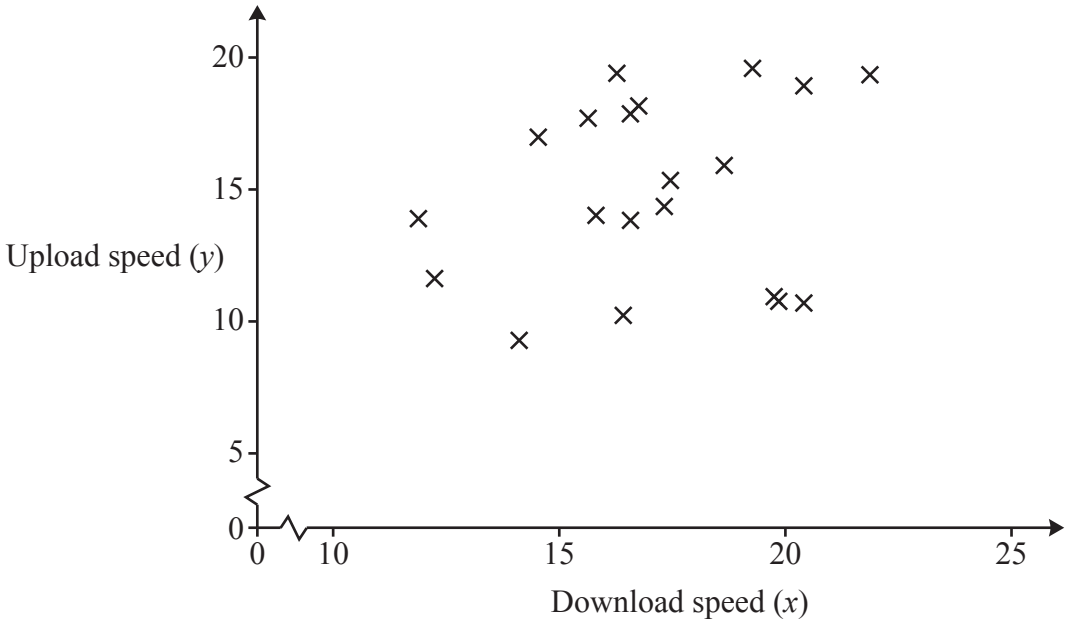
Mean	19.68
s	1.38
SE	0.2182
N	40
Interval	19.68 ± 0.4277

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

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

© OCR 2023

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]**

Summary statistics for the 20 occasions are as follows.

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

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

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

(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]**

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

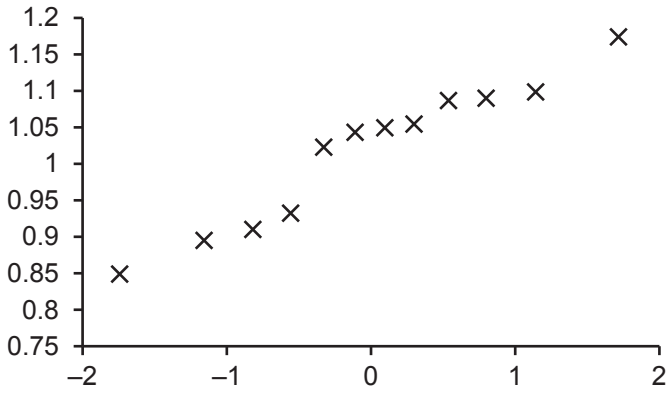
Explain why, if download speed had been a non-random variable, the student could not have carried out the hypothesis test to investigate whether there was any correlation between download speed and upload speed. **[1]**

© OCR 2023

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]**

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

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

© OCR 2023

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]**

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

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

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]

(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 $P(Y > 26)$. **[7]**

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]**

(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]**

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

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

where a is a positive constant.

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

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

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

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

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

(e) Find $E(X)$. **[2]**

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

© OCR 2023

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.

- $\mathrm{E}(X)$
- $\mathrm{Var}(X)$ **[3]**

(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]**

END OF QUESTION PAPER

© OCR 2023

OCR
Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of Cambridge University Press & Assessment, which is itself a department of the University of Cambridge.

© OCR 2023