Oxford Cambridge and RSA

Friday 14 June 2024 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

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 page 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 2024 [H/508/5594]
DC (PQ/CT) 334980/2
OCR is an exempt Charity

Section A (27 marks)

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
$\mathrm{P}(X = r)$	0.05	0.1	0.25	0.3	0.15	0.1	0.05

(a) Find each of the following.

- $\mathrm{E}(X)$
- $\mathrm{Var}(X)$ **[2]**

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]**

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]**

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

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

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

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]**

(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]**

(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]**

(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]**

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]**

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

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

Section B (93 marks)

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]**

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

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

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

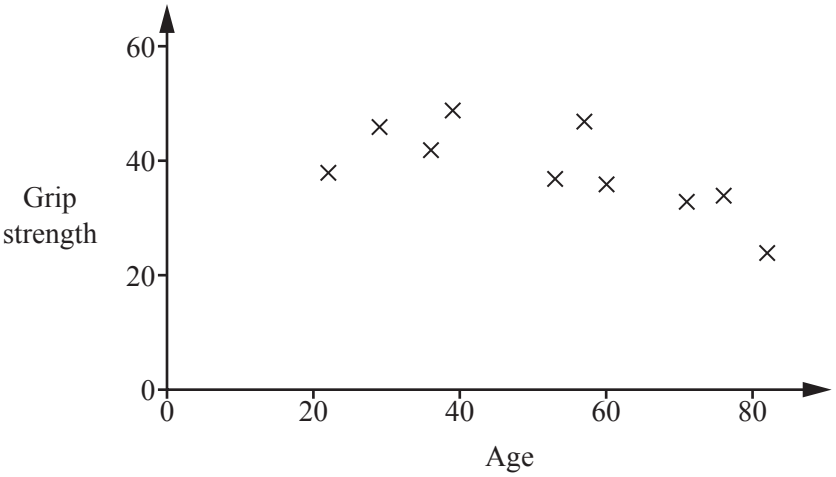
(ii) Show how the standard error was calculated in this case. **[1]**

(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]**

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]**

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

(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]**

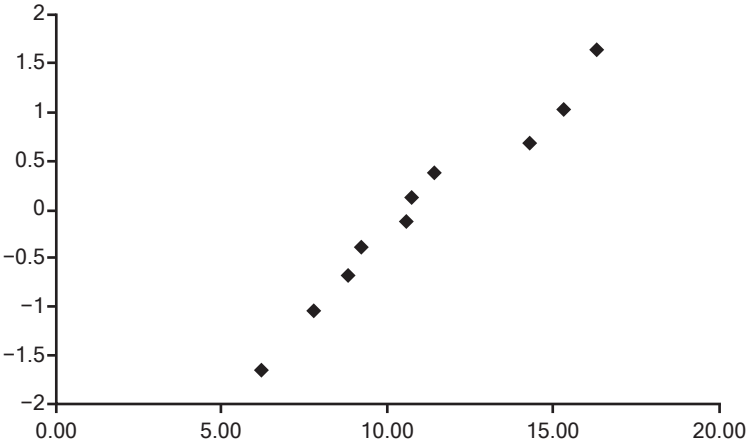
7 An environmental investigator wants to check whether the level of selenium in carrots in fields near a mine is different from the usual level in the country, which is 9.4 ng / g (nanograms per gram). She takes a random sample of 10 carrots from fields near the mine and measures the selenium level of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

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]**

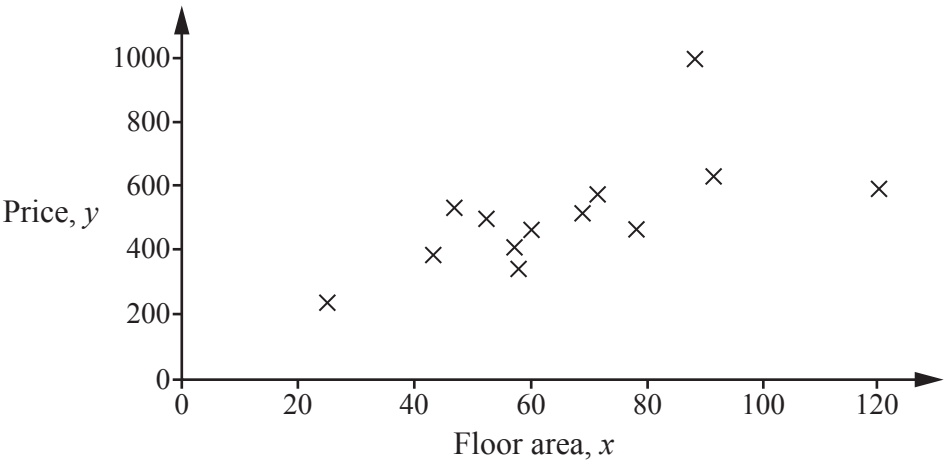
(c) State the null hypothesis for the Kolmogorov-Smirnov test for Normality. **[1]**

(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 ng / g. **[8]**

(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]**

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]**

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

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

(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]**

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

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

(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]**

(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]**

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]**

(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]**

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]**

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

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 $\mathrm{E}(Z)$. **[2]**

(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]**

(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 $\mathrm{P}(Z > 6)$. **[1]**

(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]**

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]**

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

(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]**

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

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

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]**

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

END OF QUESTION PAPER

BLANK PAGE

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 2024