OCR
Oxford Cambridge and RSA

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Practice Paper – Set 3

Time allowed: 2 hours 15 minutes

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You may use:
- a scientific or graphical calculator

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Booklet. The question number(s) must be clearly shown.
- Do **not** write in the barcodes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION
- The total number of marks for this paper is **120**.
- The marks for each question are shown in brackets **[ ]**.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is used. You should communicate your method with correct reasoning.
- The Printed Answer Booklet consists of **16** pages. The Question Paper consists of **12** pages.

Section A (30 marks)

Answer **all** the questions

1 The probability distribution of a random variable X is given by the formula

$$P(X = r) = k(0.5r^2 - 1.5r + 6) \text{ for } r = 1, 2, 3, 4, 5, 6.$$

(a) Show that $k = 0.02$. **[2]**

(b) Draw a graph to illustrate the distribution. **[2]**

(c) Comment briefly on the shape of the distribution. **[1]**

(d) Find each of the following.

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

2 Athletes are regularly tested for banned performance-enhancing drugs. For a particular drug, on average 1 test in 2000 is positive (a positive test suggests that the athlete has taken the drug). The random variable X represents the number of positive tests for this drug in a total of 1000 tests. You should assume that positive tests occur randomly and independently of each other.

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

Use either of the distributions you have identified in part **(a)** to answer part **(b)**.

(b) Calculate each of the following probabilities.

- $P(X = 2)$
- $P(X > 2)$ **[3]**

(c) A researcher wishes to find an athlete who has tested positive for this drug. Find the probability that she has to look through more than 1500 tests in order to find one positive test. **[2]**

3 A large consignment of pears is purchased for a supermarket chain. A random sample of 80 pears is selected and their weights in grams are measured. Software is used to produce a 95% confidence interval for the mean weight of the pears in the consignment. The output from the software is shown in Fig. 3.

Z Estimate of a Mean ▼

Confidence level: 0.95

Sample

Mean	145.3
s	8.4
N	80

Result

Z Estimate of a Mean

Mean	145.3
s	8.4
SE	0.93915
N	80
Interval	145.3 ± 1.8407

Fig. 3

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

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

(c) Show how the value of 1.8407 in the confidence interval was calculated. **[2]**

(d) Discuss whether you could have constructed a valid confidence interval for the average weight of the pears if the sample size had been 10 instead of 80. **[2]**

4 The length of a particular type of paving stone manufactured by a company is modelled by a Normal distribution with mean 598 mm and standard deviation 3.4 mm.

(a) Find the probability that the mean length of 10 randomly selected paving stones is at least 600 mm. **[3]**

(b) When the paving stones are laid, there is a gap between each stone and the next one. The size of the gap is modelled by a Normal distribution with mean 16 mm and standard deviation 1.6 mm. Find the probability that the total length of 5 laid paving stones and 4 gaps is less than 3.06 m. **[5]**

Section B (90 marks)

Answer **all** the questions

5 A doctor thinks that doing yoga will reduce the blood pressure of people with hypertension (high blood pressure). The doctor selects a random sample of 10 patients and enrols them on an 8-week yoga course. Fig. 5 is a screenshot of part of a spreadsheet used to analyse the results. The spreadsheet shows the SBP (systolic blood pressure), rounded to the nearest whole number, before and after the 8-week course.

	A	B	C	D
1	**Patient**	**SBP at start**	**SBP at end**	**Difference**
2	1	160	153	7
3	2	137	135	2
4	3	159	161	–2
5	4	158	157	1
6	5	162	159	3
7	6	166	164	2
8	7	130	128	2
9	8	160	150	10
10	9	141	136	5
11	10	174	168	6
12		**Mean**		3.6
13		**Standard deviation**		3.4383
14				

Fig. 5

(a) Give one possible reason why the doctor only used a sample of 10 patients with hypertension rather than all of her patients with hypertension. **[1]**

(b) State any assumptions necessary in order to construct a confidence interval for the mean difference between SBP before and after the course. **[2]**

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

Given that any necessary assumptions are satisfied, construct a 95% confidence interval for the mean difference. **[4]**

(d) Explain whether the confidence interval that you have calculated in part **(c)** supports the doctor's belief. **[1]**

(e) A 95% confidence interval for the diastolic blood pressures of the 10 people in the sample is (87.4, 92.1). For this sample, find

- the mean diastolic blood pressure,
- the standard deviation of the diastolic blood pressures. **[3]**

(f) Explain the meaning of a 95% confidence interval. **[2]**

6 Deepak is checking the fuel consumption of his car. On a quiet stretch of motorway, he drives his car at a constant speed of 50 mph for a distance of 6 miles and notes the fuel consumption in miles per gallon (mpg) indicated on his dashboard for this journey. He then repeats this process at speeds of 55, 60, 65 and 70 mph.

(a) Explain why it would not be appropriate to carry out a test for correlation based on the product moment correlation coefficient. **[2]**

The scatter diagram in Fig. 6 shows the data that Deepak collected, together with the equation of the regression line of fuel consumption on speed.

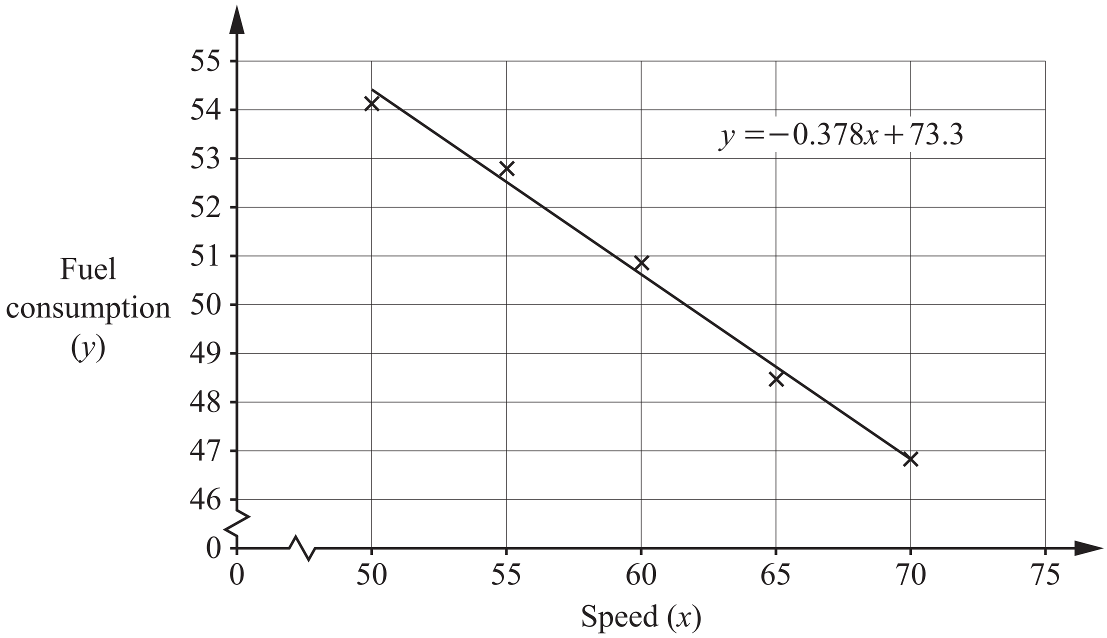


Fig. 6

(b) Explain why speed has been plotted on the horizontal axis. **[1]**

(c) Using the equation shown in Fig. 6, estimate the fuel consumption when the car is driven at the following speeds.

- 56 mph
- 90 mph **[2]**

(d) Comment on the reliability of each of your estimates. **[2]**

(e) Calculate the residual for the point where the speed is 60 mph and the fuel consumption is 50.9 mpg. **[2]**

7 A gardener wonders if there is positive correlation between lengths and widths in the leaves of bay trees. He decides to carry out a hypothesis test to investigate this and so collects a random sample of 18 bay leaves. Summary statistics for the lengths l mm and widths w mm of the leaves are as follows.

$$\Sigma l = 1713 \quad \Sigma w = 948 \quad \Sigma l^2 = 165817 \quad \Sigma w^2 = 50408 \quad \Sigma lw = 90942$$

The gardener plots the scatter diagram shown in Fig. 7 to check which type of test to carry out.

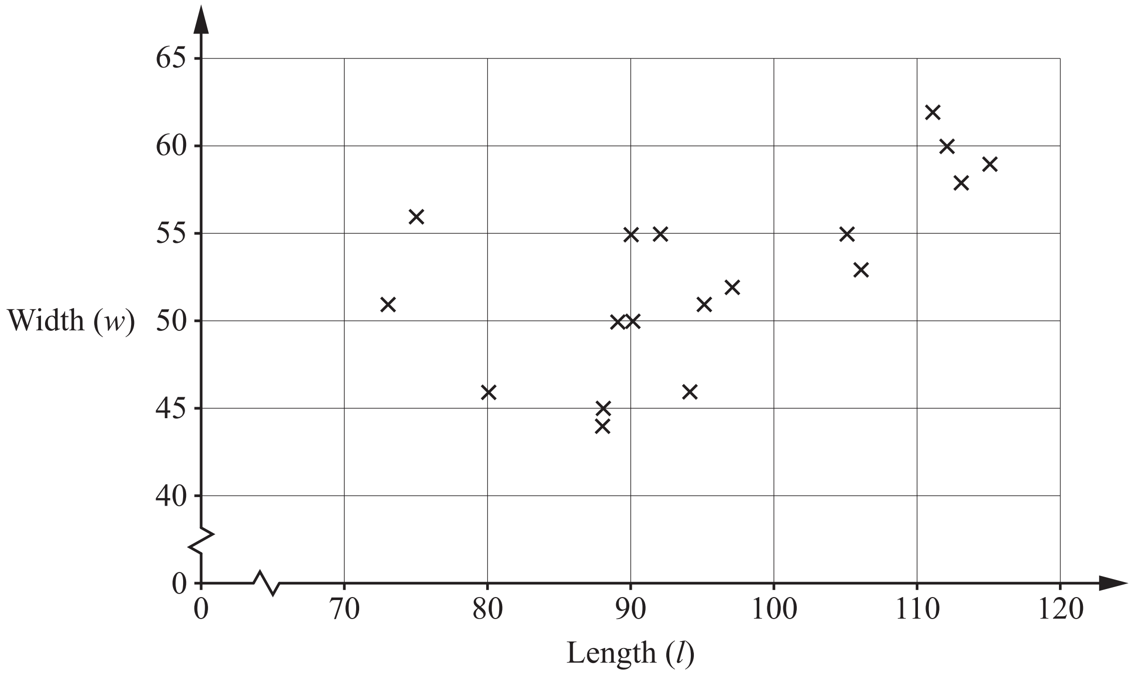


Fig. 7

(a) The gardener decides to carry out a test based on Pearson's product moment correlation coefficient. Explain why he came to this conclusion. **[2]**

(b) Find the value of Pearson's product moment correlation coefficient. **[4]**

(c) Carry out a test at the 5% significance level to investigate whether there is positive correlation between length and width. **[5]**

8 An ornithologist thinks that visits of birds to a bird feeder can be modelled by a Poisson distribution. In order to investigate this, she selects a random sample of 120 five-minute intervals, and records the number of birds visiting the feeder in each interval, with the following results.

Number of birds	0	1	2	3	4	5	6	7	$\geqslant 8$
Frequency	15	21	22	28	24	7	1	2	0

(a) Use the values of the sample mean and sample variance to discuss the suitability of a Poisson distribution as a model. **[3]**

The ornithologist decides to carry out a goodness of fit test to investigate further. The screenshot in Fig. 8 shows part of a spreadsheet to assess the goodness of fit of the distribution $\text{Po}(\lambda)$, where λ is the sample mean.

	A	B	C	D	E
1	Number of birds	Observed frequency	Poisson probability	Expected frequency	Chi-squared contribution
2	0	15	0.0821	9.8502	2.6924
3	1	21	0.2052	24.6255	0.5338
4	2	22	0.2565	30.7819	2.5054
5	3	28			
6	4	24			3.9599
7	5	7	0.0668	8.0161	0.1288
8	⩾6	3	0.0420	5.0425	0.8273

Fig. 8

(b) Calculate the missing values in each of the following cells.

- C5
- D5
- E5 **[4]**

(c) Explain why the numbers for 6, 7 and at least 8 birds have been combined into the single category of at least 6 birds, as shown in the spreadsheet. **[1]**

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

9 Kirsten commutes to her office in London every day, cycling to the railway station then taking a train.

- The time in minutes that she takes to cycle to the station is modelled by a Normal distribution with mean 11 and standard deviation 1.2.
- The time in minutes that she has to wait for the train is modelled by a continuous uniform distribution over the interval [2, 10].
- The time in minutes for the train journey is modelled by a Normal distribution with mean 42 and standard deviation 3.
- The time it takes Kirsten to walk to her office is 8 minutes.

The total time in minutes that Kirsten takes to travel from home to office is denoted by T. You should assume that the times for the separate stages of her journey are independent.

(a) Find each of the following.

- $\mathrm{E}(T)$
- $\mathrm{Var}(T)$ **[4]**

The spreadsheet in Fig. 9 shows the first 25 rows of a simulation of 1000 journeys. It also shows in column H the number of values of T that are less than the values in column G, so for example there are 217 out of the 1000 simulated values of T where T was less than 64.

	A	B	C	D	E	F	G	H	I
1	Cycle to station	Wait for train	Train time	Walk to office	Total time *T*		Total *T*	Number < *T*	
2	9.33	8.86	40.63	8	66.83		50	0	
3	10.34	2.66	41.48	8	62.47		52	0	
4	11.87	4.91	47.49	8	72.28		54	1	
5	9.86	3.77	43.30	8	64.93		56	2	
6	9.06	3.48	46.23	8	66.77		58	9	
7	11.55	8.06	39.43	8	67.05		60	30	
8	11.01	3.49	49.99	8	72.49		62	98	
9	12.49	8.70	42.68	8	71.86		64	217	
10	11.38	8.21	43.74	8	71.33		66	390	
11	9.80	7.28	39.22	8	64.30		68	576	
12	11.13	9.57	42.25	8	70.95		70	752	
13	13.14	8.65	38.17	8	67.97		72	889	
14	11.40	5.06	46.09	8	70.55		74	968	
15	9.93	4.41	39.30	8	61.64		76	991	
16	8.63	2.50	43.30	8	62.43		78	997	
17	11.87	2.32	41.81	8	64.00		80	1000	
18	8.44	3.52	42.10	8	62.07		82	1000	
19	10.19	3.65	40.73	8	62.56		84	1000	
20	9.66	3.60	44.42	8	65.68				
21	11.88	2.43	37.35	8	59.66				
22	11.44	5.12	39.61	8	64.18				
23	10.18	8.24	46.21	8	72.63				
24	10.71	3.12	40.08	8	61.91				
25	12.11	9.13	35.55	8	64.79				

Fig. 9

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

- $\mathrm{P}(T < 60)$
- $\mathrm{P}(T \geqslant 68)$ **[3]**

(c) The random variable W is Normally distributed with the same mean and variance as T. Find each of the following.

- $\mathrm{P}(W < 60)$
- $\mathrm{P}(W \geqslant 68)$ **[2]**

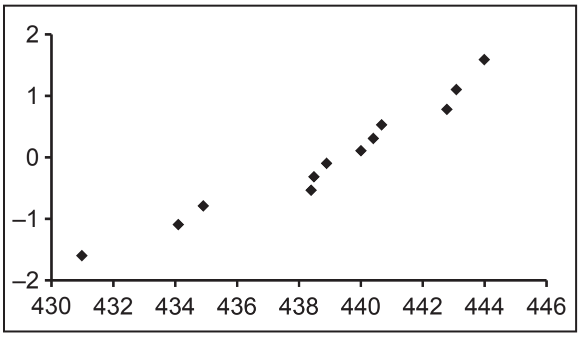
(d) Explain why, even if more and more journeys were simulated, you would not expect the probabilities in part **(b)** from the simulation to get closer and closer to those calculated in part **(c)**. **[2]**

(e) Kirsten wishes to investigate whether or not the distribution of T is symmetrical. Explain how data from Fig. 9 could be used to investigate this. **[2]**

10 The temperature of the exhaust gas in a particular type of industrial jet engine is specified as 441 °C. An engineer is testing an engine to ensure that it is working correctly. The temperature of the exhaust gas in °C is measured on a random sample of 12 occasions, with the following results.

431.0 438.4 443.1 440.0 434.9 440.7 443.1 438.9 438.5 440.4 434.1 442.8

The engineer uses software to draw a Normal probability plot for these data, and to carry out a Normality test as shown in Fig. 10.



Kolmogorov-Smirnov Test for Normality

p-value = 0.210

Fig. 10

(a) The engineer carries out a hypothesis test to check whether the exhaust gas temperature is 441 °C on average. Explain which test the engineer should use. **[3]**

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

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

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

$$F(x) = \begin{cases} 0 & x < a, \\ q - \frac{p}{x} & a \leqslant x \leqslant b, \\ 1 & x > b, \end{cases}$$

where a, b, p and q are constants.

(a) Show that $p = \frac{ab}{b-a}$. **[5]**

The time that it takes for my mobile phone to take a picture is modelled by the random variable T milliseconds. The cumulative distribution function of T is the same as that of X, with $a = 100$ and $b = 500$.

(b) Find $P(150 \leqslant T \leqslant 200)$. **[3]**

(c) **(i)** Find $E(T)$. **[3]**

(ii) Find the standard deviation of T. **[3]**

(d) Sketch the graph of the probability density function of T. **[2]**

(e) Explain why the mode of T is 100. **[1]**

END OF QUESTION PAPER

BLANK PAGE