OCR
Oxford Cambridge and RSA

Friday 13 June 2025 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

187041866O

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.

Section A (34 marks)

1 A fair eight sided dice has its sides labelled 1, 2, …, 8. The random variable X represents the score when the dice is rolled once.

(a) Find $P(X > 4)$. **[1]**

(b) Find each of the following.

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

The dice is rolled 3 times.

(c) Determine the exact probability that one of the scores is less than 4 and the other two are greater than 4. **[3]**

2 At a toy factory, wooden blocks of approximate heights 20 mm, 30 mm and 50 mm are made in red, yellow and green respectively. The heights of the blocks in mm are modelled by independent random variables which are Normally distributed with means and standard deviations as shown in the table.

Colour	Mean	Standard deviation
Red	20	0.8
Yellow	30	0.9
Green	50	1.2

In parts **(a)**, **(b)** and **(c)**, the blocks are selected randomly and independently of one another.

(a) Find the probability that the height of a red block is less than 19 mm. **[1]**

(b) A tower is made of 15 blocks stacked on top of each other consisting of 5 red blocks, 5 yellow blocks and 5 green blocks.

Determine the probability that the tower is at least 495 mm high. **[4]**

(c) Determine the probability that a tower made of 3 red blocks will be at least 1 mm higher than a tower made of 2 yellow blocks. **[3]**

3 Potatoes are sold in sacks of nominal weight 25 kg. As part of a trading standards check, a random sample of 50 sacks from a large batch is selected and software is used to produce a 95% confidence interval for the mean weight of potatoes per sack. An extract from the software is shown below.

Sample Mean	24.878
Standard Deviation	0.5664
Standard Error	0.0801
Sample Size	50
Confidence Level	0.95
Interval	24.878 ± 0.157

(a) Show how the standard error was calculated. **[1]**

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

(c) Explain whether the confidence interval suggests that the mean weight of potatoes per sack is different from 25 kg. **[1]**

(d) Determine whether your conclusion to part **(c)** would have been different if the confidence level had been 90%. **[4]**

4 Gabi is practising basketball. On each attempt to score a basket, the probability that Gabi scores is modelled as 0.4. It is assumed that each attempt is independent of previous attempts. The number of attempts which Gabi takes in order to score a basket for the first time is denoted by the random variable X.

(a) **(i)** Calculate $P(X = 5)$. **[1]**

(ii) Calculate $P(X > 5)$. **[1]**

(b) Determine the probability that X is within one standard deviation of its mean. **[5]**

(c) Determine the probability that Gabi scores at least 5 baskets in 20 attempts. **[2]**

(d) Determine the probability that Gabi scores the 5th basket on their 20th attempt. **[2]**

(e) Comment on the validity of the assumption that on each attempt to score a basket, the probability that Gabi scores is 0.4, independent of previous attempts. **[1]**

BLANK PAGE

Section B (86 marks)

5 The random variable X has a continuous uniform distribution over $[-4, 2]$.

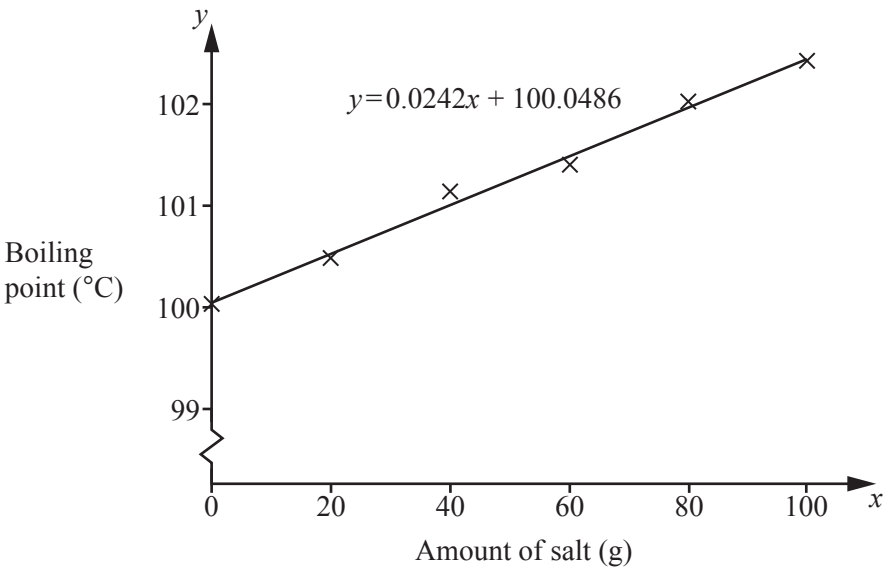
(a) Find the variance of X. **[2]**

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

6 A student is investigating the link between the amount of salt dissolved in water and the boiling point of the water. The student dissolves fixed amounts of salt, x g, into water and then measures the boiling point, y °C, of the water.

The data are shown in the table together with a scatter diagram to illustrate them. The equation of the regression line of y on x is also shown.

Amount of salt, x g	0	20	40	60	80	100
Boiling point, y °C	100.03	100.49	101.15	101.41	102.04	102.44



(a) State what type of variable 'Boiling point' is. **[1]**

(b) Explain why 'Amount of salt' has been plotted on the horizontal axis. **[1]**

(c) Calculate an estimate of the boiling point with the following amounts of salt.

- 50 g
- 200 g

[2]

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

(e) Determine the residual for the data point with 60 g of salt. **[2]**

7 A student is investigating the relationship between different electricity generation methods and cost of electricity in a particular country. The student first checks whether there is any correlation between the cost per unit of electricity, x euros, and the amount of electricity being generated by wind, y GW. The data from 30 observations are summarised as follows.

$n = 30 \quad \Sigma x = 2.219 \quad \Sigma y = 357.7 \quad \Sigma x^2 = 0.2368 \quad \Sigma y^2 = 4648 \quad \Sigma xy = 25.01$

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

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

(b) Carry out a hypothesis test at the 5% level to investigate whether there is any correlation between the cost per unit of electricity and the amount of electricity generated by wind. **[5]**

Table 7.1 shows the values of the product moment correlation coefficients between 5 variables. These variables are the cost per unit of electricity in euros and the amount of electricity generated in GW by four different types of generation, for a very large sample of observations.

For example, the correlation coefficient between the amount of electricity generated by wind and the amount of electricity generated by gas is −0.59. The 'storage' category consists of battery and other types of electricity taken from stored reserves. All of the correlations are significant at the 10% level.

Table 7.1

	Cost	Gas	Wind	Solar	Storage
Cost	1.00				
Gas	0.42	1.00			
Wind	−0.26	−0.59	1.00		
Solar	−0.18	−0.21	−0.07	1.00	
Storage	0.12	0.35	−0.38	−0.30	1.00

Table 7.2 shows standard guidelines for effect sizes.

Table 7.2

Product moment correlation coefficient	**Effect size**
$0.1 \leqslant r < 0.3$	Small
$0.3 \leqslant r < 0.5$	Medium
$r \geqslant 0.5$	Large

The student analyses these data for effect size.

(c) Explain how the very large sample size relates to the interpretation of the correlation coefficient of −0.07 between the amount of electricity generated by wind and the amount of electricity generated by solar, as shown in **Table 7.1**. **[2]**

(d) Comment briefly on what the student might infer from these tables regarding different electricity generation methods and cost of electricity. **[2]**

8 A teacher is investigating whether students run more quickly round an athletics track if they run in a clockwise direction rather than in an anti-clockwise direction. She thinks that students will run more quickly on average in a clockwise direction.

The teacher selects 40 students at random and records how long it takes for each of them to run one lap of the track in a clockwise direction. The next day she records how long it takes for each of these students to run one lap in an anti-clockwise direction.

(a) Explain why the teacher does **not** get the students to run the second lap immediately after they have completed their first lap. **[1]**

(b) Explain why the teacher notes the names of each student, so that she knows which student took which time on each day, rather than just noting the 40 times on each day. **[1]**

The differences in times, d seconds (anti-clockwise time ─ clockwise time), are summarised as follows:

$\sum d = 72.2, \quad \sum d^2 = 1510$

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

Carry out a hypothesis test at the 5% significance level to investigate whether there is any evidence to suggest that on average students run more quickly round the track in a clockwise direction. **[10]**

(d) Comment briefly on any possible problem with the teacher's method. **[1]**

9 A consultant is modelling the number of trades per minute of a particular share in a stock exchange. The number of trades per minute in a random sample of 200 minutes is shown in the table.

Number of trades per minute	0	1	2	3	4	5	6	7	⩾8
Frequency	23	42	46	45	20	15	7	2	0

The consultant thinks that the number of trades per minute may be modelled by a Poisson distribution.

(a) Use the data in the table to estimate the value of the parameter λ for an appropriate Poisson model. **[2]**

The consultant decides to carry out a chi-squared goodness of fit test to investigate further. The screenshot below shows part of a spreadsheet to assess the goodness of fit of the distribution Po(λ), using the value of λ estimated from the data.

	A	B	C	D	E
1	Number of trades per minute	Frequency	Poisson probability	Expected frequency	Chi-squared contribution
2	0	23	0.0907	18.1436	1.2999
3	1	42			
4	2	46			0.7484
5	3	45	0.2090	41.8028	0.2445
6	4	20	0.1254	25.0817	1.0296
7	5	15	0.0602	12.0392	0.7281
8	⩾6	9	0.0357	7.1345	0.4878
9					

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

- C3
- D3
- E3

[4]

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

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

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

10 Two independent random variables, X and Y, have distributions B(6, 0.4) and B(3, 0.4) respectively. The random variable $T = X - (Y_1 + Y_2)$ where Y_1 and Y_2 are two independent values of Y.

(a) Determine P($T = 5$). **[4]**

A spreadsheet has been constructed to simulate the situation, as shown below. There are 100 simulated values of each of X, Y_1, Y_2 and T.

Cells G2 and G3 in the spreadsheet show the number of values of T which are greater than 0, and less than or equal to 0, respectively. Cell G5 shows the mean value of the 100 simulated values of T.

	A	B	C	D	E	F	G
1	X	Y_1	Y_2	$T=X-(Y_1+Y_2)$			
2	3	2	1	0		Number > 0	42
3	2	0	0	2		Number ⩽ 0	58
4	3	1	0	2			
5	1	3	2	−4		Mean value of T	0.25
6	4	2	1	1			
7							
8							
99	4	1	1	2			
100	2	2	2	−2			
101	3	0	3	0			
102							
103							

(b) Use the information in the spreadsheet to write down an estimate of P($T > 0$). **[1]**

As shown in the spreadsheet, the mean value of the 100 simulated values of T is 0.25.

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

Use the Central Limit Theorem to determine an estimate of the probability that the mean of 100 independent values of T is greater than 0.25. **[8]**

11 The function F(x) is defined as follows.

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{ax+b}{9x+15} - \frac{4}{3} & 0 \leqslant x \leqslant 5 \\ 1 & x > 5 \end{cases}$$

where a and b are constants.

You are given that F(1) = 0.5 and F(2.5) = 0.8.

(a) Show that $a = 24$ and $b = 20$. **[3]**

(b) (i) Show that F(0) = 0. **[1]**

(ii) Show that F(x) is a non-decreasing function. **[3]**

(c) (i) State another property that is required for F(x) to be a cumulative distribution function. **[1]**

(ii) Show that the property in part **(c)(i)** is satisfied. **[1]**

You are given that F(x) is the cumulative distribution function of the continuous random variable X.

(d) Determine the upper quartile of X. Give your answer as an exact fraction. **[3]**

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

(f) Determine the mode of X by considering the gradient of the probability density function of X. **[2]**

12 The random variable X takes the value 1 with probability p and the value 0 with probability $1 - p$. You are given that E(X) = p and Var(X) = $p - p^2$.

The random variable $Y \sim B(64, 0.1)$ has mean μ and standard deviation σ.

Use the results given above to **prove** the following in either order.

- $\mu = 6.4$
- $\sigma = 2.4$ **[6]**

END OF QUESTION PAPER

BLANK PAGE