Oxford Cambridge and RSA

Wednesday 15 June 2022 – Afternoon

AS Level Further Mathematics B (MEI)

Y412/01 Statistics a

Time allowed: 1 hour 15 minutes

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

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 **60**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **8** pages.

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

Answer **all** the questions.

1 A fair five-sided spinner has sectors labelled 1, 2, 3, 4, 5. In a game at a stall at a charity event, the spinner is spun twice. The random variable X represents the lower of the two scores. The probability distribution of X is given by the formula

$$P(X = r) = k(11 - 2r) \text{ for } r = 1, 2, 3, 4, 5,$$

where k is a constant.

(a) Complete the copy of this table in the Printed Answer Booklet. **[1]**

r	1	2	3	4	5
$P(X = r)$		$7k$		$3k$	

(b) Determine the value of k. **[1]**

(c) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

(d) The stall-holder charges a player C pence to play the game, and then pays the player $50X$ pence, where X is the player's score.

Given that the average profit that the stall-holder makes on one game is 25 pence, find the value of C. **[2]**

2 On a car assembly line, a robot is used for a particular task.

(a) State the conditions under which a Poisson distribution is an appropriate model for the number of breakdowns of the robot in a week. **[2]**

It is given that the average number of breakdowns of the robot in a week is 1.7. For the remainder of this question, you should assume that a Poisson distribution is an appropriate model for the number of breakdowns of the robot in a week.

(b) (i) Find the probability that the number of breakdowns of the robot in a week is exactly 4. **[1]**

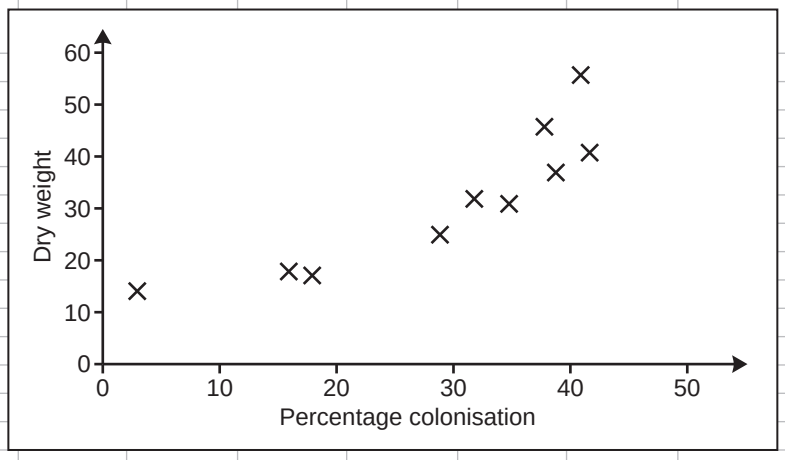
(ii) Determine the probability that the number of breakdowns of the robot in a week is at least 2. **[2]**

(c) Determine the probability that the number of breakdowns of the robot in 52 weeks is less than 100. **[2]**

3 A biology student is doing an experiment in which plants are inoculated with a particular microorganism in an attempt to help them grow. She is investigating whether there is any association between the percentage of roots which have been colonised by the microorganism and the dry weight of the plant shoots. After the plants have grown for a few weeks, the student takes a random sample of 10 plants and measures the percentage of roots which have been colonised by the microorganism and the dry weight of the plant shoots.

The spreadsheet output shows the data, together with a scatter diagram to illustrate the data.

	A	B
1	Percentage colonisation	Dry weight
2	3	14
3	18	17
4	16	18
5	29	25
6	35	31
7	32	32
8	39	37
9	42	41
10	38	46
11	41	56



(a) The student decides that a test based on Pearson's product moment correlation coefficient may not be valid.

Explain why she comes to this conclusion. **[2]**

(b) Calculate the value of Spearman's rank correlation coefficient. **[3]**

(c) Carry out a test based on this coefficient, at the 5% significance level, to investigate whether there is any association between percentage colonisation and shoot dry weight. **[5]**

4 A random number generator generates integers between 1 and 50 inclusive, with each number having an equal probability of being generated.

(a) State the probability distribution of the numbers generated. **[2]**

(b) Determine the probability that a number generated is within one standard deviation of the mean. **[4]**

5 A researcher is investigating whether there is any relationship between the overall performance of a student at GCSE and their grade in A Level Mathematics. Their A Level Mathematics grade is classified as A* or A, B, C or lower, and their overall performance at GCSE is classified as Low, Middle, High.

Data are collected for a sample of 80 students in a particular area. The researcher carries out a chi-squared test. The screenshot below shows part of a spreadsheet used to analyse the data. Some values in the spreadsheet have been deliberately omitted.

	A	B	C	D	E
1		**Observed frequency**			
2		**A* or A**	**B**	**C or lower**	**Totals**
3	**Low**	6	13	9	28
4	**Middle**	10	6	8	24
5	**High**	15	10	3	28
6	**Totals**	31	29	20	80
7					
8		**Expected frequency**			
9		**A* or A**	**B**	**C or lower**	
10	**Low**	10.85			
11	**Middle**	9.30			
12	**High**	10.85	10.15	7.00	
13					
14		**Contribution to the test statistic**			
15		**A* or A**	**B**	**C or lower**	
16	**Low**	2.1680	0.8002	0.5714	
17	**Middle**	0.0527	0.8379	0.6667	
18	**High**	1.5873		2.2857	
19					

(a) State what needs to be known about the sample for the test to be valid. **[1]**

For the remainder of this question, you should assume that the test is valid.

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

- C11
- C18

[3]

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

Carry out a hypothesis test at the 10% significance level to investigate whether there is any association between level of performance at GCSE and A Level Mathematics grade. **[6]**

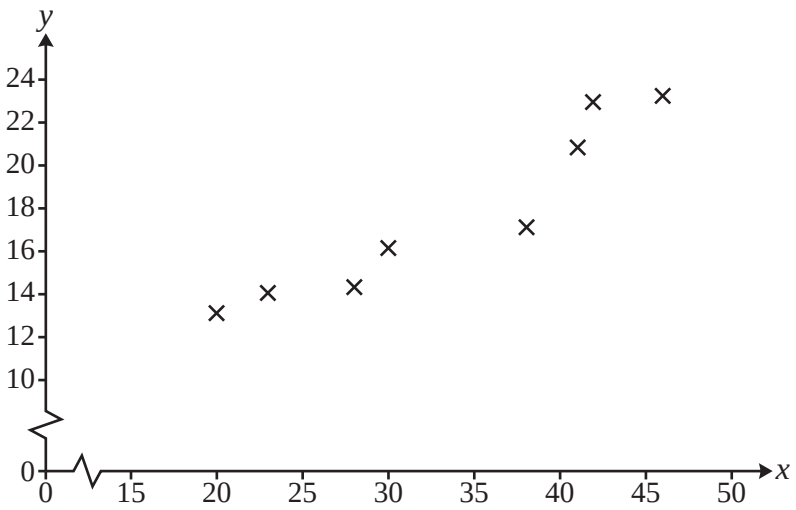
(d) Discuss briefly what the data suggest about A Level Mathematics grade for different levels of performance at GCSE. **[3]**

(e) State **one** disadvantage of using a 10% significance level rather than a 5% significance level in a hypothesis test. **[1]**

6 Tom has read in a newspaper that you can tell the air temperature by counting how often a cricket chirps in a period of 20 seconds. (A cricket is a type of insect.) He wants to know exactly how the temperature can be predicted. On 8 randomly selected days, when Tom can hear crickets chirping, he records the number of chirps, x, made by a cricket in a 20-second interval, and also the temperature, y °C, at that time. The data are summarised as follows.

$n = 8$ $\quad \Sigma x = 268$ $\quad \Sigma y = 141.9$ $\quad \Sigma x^2 = 9618$ $\quad \Sigma y^2 = 2630.55$ $\quad \Sigma xy = 5009.1$

These data are illustrated below.



(a) Determine the equation of the regression line of y on x. Give your answer in the form $y = ax + b$, giving the values of a and b correct to **3** significant figures. **[4]**

(b) Use the equation of the regression line to predict the temperature for the following values of x.

- 35
- 10

[2]

(c) Comment on the reliability of your predictions in part **(b)**. **[3]**

(d) State the coordinates of the point of intersection of the line whose equation you have calculated with the regression line of x on y. **[1]**

7 On average one in five packets of a breakfast cereal contains a voucher for a discount on the next packet bought. Whether or not a packet contains a voucher is independent of other packets, and can only be determined by opening the packet.

(a) State the distribution of the number of packets that need to be opened in order to find one which contains a voucher. **[2]**

(b) Determine the probability that exactly 4 packets have to be opened in order to find one which contains a voucher. **[1]**

(c) Determine the probability that exactly 10 packets have to be opened in order to find two which contain a voucher. **[2]**

(d) I have n packets, and I open them one by one until I find a voucher or until all the packets are open.

Given that the probability that I find a voucher is greater than 0.99, determine the least possible value of n. **[2]**

END OF QUESTION PAPER

BLANK PAGE