

**Paper Reference 9MA0–31**  
**Pearson Edexcel**  
**Level 3 GCE**

**Mathematics**  
**Advanced**  
**Paper 31: Statistics**

**Friday 14 June 2019 – Afternoon**

**MATERIALS REQUIRED FOR EXAMINATION**  
**Mathematical Formulae and Statistical Tables,**  
**calculator**

**ITEMS INCLUDED WITH QUESTION PAPER**  
**Diagram Book**  
**Answer Book**

**X63358A**

**Candidates may use any calculator allowed by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.**

## **INSTRUCTIONS**

**In the boxes on the Answer Book and on the Diagram Book, write your name, centre number and candidate number.**

**Answer ALL questions and ensure that your answers to parts of questions are clearly labelled.**

**Answer the questions in the Answer Book or on the separate diagrams – there may be more space than you need.**

**Do NOT write on the Question Paper.**

**You should show sufficient working to make your methods clear. Answers without working may not gain full credit.**

**Answers should be given to three significant figures unless otherwise stated.**

## **INFORMATION**

**A booklet 'Mathematical Formulae and Statistical Tables' is provided.**

**The total mark for this part of the examination is 50**

**There are 5 questions.**

**The marks for EACH question are shown in brackets  
– use this as a guide as to how much time to spend on each question.**

## **ADVICE**

**Read each question carefully before you start to answer it.**

**Try to answer every question.**

**Check your answers if you have time at the end.**

---

**Answer ALL questions.**

**Write your answers in the Answer Book.**

1. Three bags, **A**, **B** and **C**, each contain **1** red marble and some green marbles.

Bag **A** contains **1** red marble and **9** green marbles only

Bag **B** contains **1** red marble and **4** green marbles only

Bag **C** contains **1** red marble and **2** green marbles only

Sasha selects at random one marble from bag **A**

If he selects a red marble, he stops selecting.

If the marble is green, he continues by selecting at random one marble from bag **B**

If he selects a red marble, he stops selecting.

If the marble is green, he continues by selecting at random one marble from bag **C**

(continued on the next page)

**1. continued.**

**(a) Draw a tree diagram to represent this information.**

**(2 marks)**

**(b) Find the probability that Sasha selects 3 green marbles.**

**(2 marks)**

**(c) Find the probability that Sasha selects at least 1 marble of each colour.**

**(2 marks)**

**(d) Given that Sasha selects a red marble, find the probability that he selects it from bag B**

**(2 marks)**

**(Total for Question 1 is 8 marks)**

---

2. Refer to the diagrams for Question 2 in the Diagram Book.

The partially completed box plot in Diagram 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

Diagram 2 shows the same partially completed box plot.

An outlier is defined as a value more than  $1.5 \times \text{IQR}$  below  $Q_1$  or more than  $1.5 \times \text{IQR}$  above  $Q_3$

The three lowest air temperatures in the data set are  $7.6^\circ\text{C}$ ,  $8.1^\circ\text{C}$  and  $9.1^\circ\text{C}$

The highest air temperature in the data set is  $32.5^\circ\text{C}$

- (a) Complete the box plot in Diagram 1 showing clearly any outliers.  
(4 marks)

(continued on the next page)

**2. continued.**

- (b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.**

**(1 mark)**

**Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature,  $x^{\circ}\text{C}$ , for Beijing in 2015**

$$n = 184 \quad \sum x = 4153 \cdot 6 \quad S_{xx} = 4952 \cdot 906$$

- (c) Show that, to 3 significant figures, the standard deviation is  $5 \cdot 19^{\circ}\text{C}$**

**(1 mark)**

**(continued on the next page)**



**2. continued.**

**Simon decides to model the air temperatures with the random variable**

$$T \sim N(22.6, 5.19^2)$$

- (d) Using Simon's model, calculate the 10th to 90th interpercentile range.**  
**(3 marks)**

**Simon wants to model another variable from the large data set for Beijing using a normal distribution.**

- (e) State two variables from the large data set for Beijing that are NOT suitable to be modelled by a normal distribution.**  
**Give a reason for each answer.**  
**(2 marks)**

**(Total for Question 2 is 11 marks)**

---

3. Barbara is investigating the relationship between average income (GDP per capita),  $x$  US dollars, and average annual carbon dioxide ( $\text{CO}_2$ ) emissions,  $y$  tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual  $\text{CO}_2$  emissions and average income to be 0.446

- (a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero.

(3 marks)

(continued on the next page)

3. continued.

Barbara believes that a non-linear model would be a better fit to the data.

She codes the data using the coding  $m = \log_{10} x$  and  $c = \log_{10} y$  and obtains the model

$$c = -1.82 + 0.89m$$

The product moment correlation coefficient between  $c$  and  $m$  is found to be  $0.882$

(b) Explain how this value supports Barbara's belief.

(1 mark)

(c) Show that the relationship between  $y$  and  $x$  can be written in the form  $y = ax^n$  where  $a$  and  $n$  are constants to be found.

(5 marks)

(Total for Question 3 is 9 marks)

---

- 4. Refer to the table for Question 4 in the Diagram Book.**

**Magali is studying the mean total cloud cover, in oktas, for Leuchars in 1987 using data from the large data set.**

**The daily mean total cloud cover for all 184 days from the large data set is summarised in the table.**

**One of the 184 days is selected at random.**

- (a) Find the probability that it has a daily mean total cloud cover of 6 or greater.**

**(1 mark)**

**(continued on the next page)**

4. continued.

Magali is investigating whether the daily mean total cloud cover can be modelled using a binomial distribution.

She uses the random variable  $X$  to denote the daily mean total cloud cover and believes that  $X \sim B(8, 0.76)$

Using Magali's model,

(b) (i) find  $P(X \geq 6)$   
(2 marks)

(ii) find, to 1 decimal place, the expected number of days in a sample of 184 days with a daily mean total cloud cover of 7  
(2 marks)

(c) Explain whether or not your answers to part (b) support the use of Magali's model.  
(1 mark)

(continued on the next page)

Turn over

**4. continued.**

**Refer to the table for Question 4(c) in the Diagram Book.**

**There were 28 days that had a daily mean total cloud cover of 8**

**For these 28 days the daily mean total cloud cover for the FOLLOWING day is shown in the table.**

**(d) Find the proportion of these days when the daily mean total cloud cover was 6 or greater.  
(1 mark)**

**(e) Comment on Magali's model in light of your answer to part (d)  
(2 marks)**

**(Total for Question 4 is 9 marks)**

---

5. A machine puts liquid into bottles of perfume.  
The amount of liquid put into each bottle,  $D$  ml,  
follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than  
24.63 ml

- (a) find, to 2 decimal places, the value of  $k$  such  
that  $P(24.63 < D < k) = 0.45$   
(5 marks)

A random sample of 200 bottles is taken.

- (b) Using a normal approximation, find the  
probability that fewer than half of these bottles  
contain between 24.63 ml and  $k$  ml  
(3 marks)

(continued on the next page)

**5. continued.**

**The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml**

**Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml**

**She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml**

**(c) Test Hannah's belief at the 5% level of significance.**

**You should state your hypotheses clearly.  
(5 marks)**

**(Total for Question 5 is 13 marks)**

---

**TOTAL FOR STATISTICS IS 50 MARKS**

**END OF PAPER**

---