

Please check the examination details below before entering your candidate information

Candidate surname

Other names

**Pearson Edexcel  
Level 3 GCE**

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

**Friday 14 June 2019**

Afternoon

Paper Reference **9MA0-31**

**Mathematics**

**Advanced**

**Paper 31: Statistics**

**You must have:**

Mathematical Formulae and Statistical Tables, calculator

Total Marks

**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

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with 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 spaces provided  
– *there may be more space than you need.*
- 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.

Turn over ►

P63358A

©2019 Pearson Education Ltd.

1/1/1/



  
Pearson

Answer ALL questions. Write your answers in the spaces provided.

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*.

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

(2)

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

(2)

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

(2)

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

(2)

Horizontal lines for writing answers.



**Question 1 continued**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



**Question 1 continued**

DO NOT WRITE IN THIS AREA

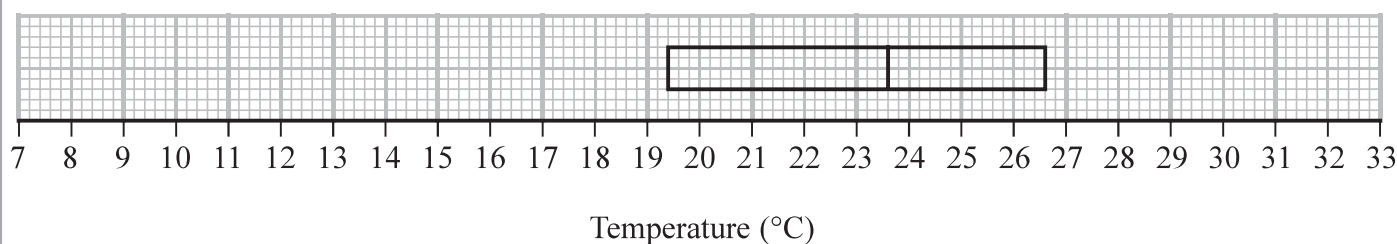
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA





2.

**Figure 1**

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

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 Figure 1 showing clearly any outliers. (4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come. (1)

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

$$n = 184 \quad \sum x = 4153.6 \quad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is  $5.19^\circ\text{C}$  (1)

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)

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)

---



---



---



**Question 2 continued**

Lined writing area for the answer to Question 2.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**Turn over for a spare grid if you need to redraw your box plot.**



Question 2 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA





DO NOT WRITE IN THIS AREA

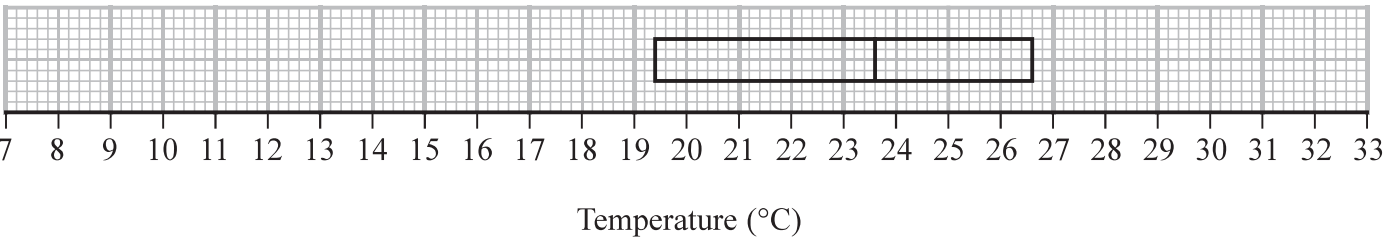
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

### Question 2 continued

Lined writing area for the student's response.

Only use this grid if you need to redraw your box plot.



(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)

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)

- (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)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Question 3 continued

Ruled area for writing answers, containing 25 horizontal lines.



**Question 3 continued**

Handwriting practice area consisting of 30 horizontal lines.

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA





4. 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 below.

<b>Daily mean total cloud cover (oktas)</b>	0	1	2	3	4	5	6	7	8
<b>Frequency (number of days)</b>	0	1	4	7	10	30	52	52	28

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)

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)

- (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)

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

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 below.

<b>Daily mean total cloud cover (oktas)</b>	0	1	2	3	4	5	6	7	8
<b>Frequency (number of days)</b>	0	0	1	1	2	1	5	9	9

- (d) Find the proportion of these days when the daily mean total cloud cover was 6 or greater. (1)
- (e) Comment on Magali's model in light of your answer to part (d). (2)

---



---



---



---



---



---



---



**Question 4 continued**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



P 6 3 3 5 8 A 0 1 5 2 0

**Question 4 continued**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA





**Question 4 continued**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Lined writing area for the answer.

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







**Question 5 continued**

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 5 is 13 marks)

**TOTAL FOR STATISTICS IS 50 MARKS**

