

Please check the examination details below before entering your candidate information

Candidate surname					Other names						
<b>Pearson Edexcel</b>		Centre Number					Candidate Number				
		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
<b>Level 3 GCE</b>											
<b>Friday 14 June 2019</b>											
Afternoon					Paper Reference <b>9MA0-31</b>						
<b>Mathematics</b>					<b>DRAFT Shadow Set 1</b>						
<b>Advanced</b>											
<b>Paper 31: Statistics</b>											
<b>You must have:</b> 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).
- 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.
- There are 5 questions in this question paper. The total mark for this paper is 50.
- 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.**

- 1.** Three boxes, *A*, *B* and *C*, each contain 1 red scarf and some green scarves.

Box *A* contains 1 red scarf and 7 green scarves only

Box *B* contains 1 red scarf and 7 green scarves only

Box *C* contains 1 red scarf and 9 green scarves only

Julia selects at random one scarf from Box *A*.

If she selects a red scarf, she stops selecting.

If the scarf is green, she continues by selecting at random one scarf from Box *B*.

If she selects a red scarf, she stops selecting.

If the scarf is green, she continues by selecting at random one scarf from Box *C*.

(a) Draw a tree diagram to represent this information. (2)

(b) Find the probability that Julia selects 3 green scarves. (2)

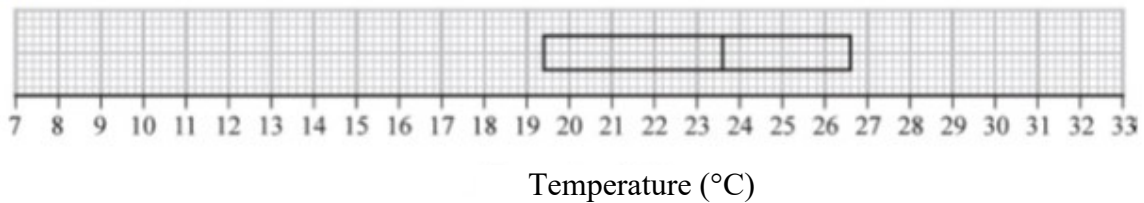
(c) Find the probability that Julia selects at least 1 scarf of each colour. (2)

(d) Given that Julia selects a red scarf, find the probability that she selects it from Box *B*. (2)

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

---

2.



**Figure 1**

The partially complete box plot in Figure 1 above shows the distribution of daily mean air temperatures using the data from a 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  $10.9^\circ\text{C}$ ,  $7.6^\circ\text{C}$  and  $9.5^\circ\text{C}$ .  
 The highest air temperature in the data set is  $31.6^\circ\text{C}$ .

(a) Complete the box plot in Figure 1. Write down 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 same large set, Craig produced the following summary statistics for the daily mean air temperature,  $x^\circ\text{C}$ , for Beijing in 2015.

$$n = 166 \quad \sum x = 4222.8 \quad S_{xx} = 4877.585$$

(c) Show that, to 3 significant figures, the standard deviation is  $5.42^\circ\text{C}$  (1)

Craig decides to model the air temperature with the random variable  $T \sim N(25.44, 5.42^2)$ .

(d) Using Craig's model, calculate the 10th to 90th interpercentile range. (3)

Craig 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)

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

- 
3. Helen is investigating the relationship between average income (GDP per capita),  $x$  US dollars, and average annual carbon dioxide (CO<sub>2</sub>) emissions,  $y$  tonnes, for different countries.

She takes a random sample of the 24 countries and finds the product moment correlation average annual CO<sub>2</sub>, emission and the average income to be 0.4771.

- (a) Taking the null hypothesis to be that there is no correlation, test, at the 5% level of significance, whether to reject the null hypothesis and say that the product moment correlation coefficient for all the countries is greater than zero.

**(3)**

Helen 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  $c = -0.47 + 0.92m$ .

The product moment correlation coefficient between  $c$  and  $m$  is found to be 0.8432.

- (b) Explain how this value supports Helen'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. Give your answers to 2 decimal places.

**(5)**

**(Total for Question 3 is 9 marks)**

---

4. Chen is studying the mean total cloud cover, in oktas, for Leuchars in 1987. The daily mean total cloud cover for all 195 days 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>	1	2	5	6	13	28	50	59	31

One of the 195 days is selected at random.

- (a) Find the probability that it has a daily mean total cloud cover of 7 or greater. (1)

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

He uses the random variable  $X$  to denote the daily mean total cloud cover and believes  $X \sim B(8, 0.78)$

- (b) Using Chen's model, find,

- (i)  $P(X \geq 7)$  to 2 decimal places, (2)

- (ii) the expected number of days in a sample of 195 days with a daily mean total cloud cover of 7, to 1 decimal place. (2)

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

There were 31 days that had a daily mean total cloud cover of 8. For these 31 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	0	1	2	2	6	9	11

- (d) Find the proportion of these days when the daily mean total cloud cover was 7 or greater. (1)

- (e) Comment on Chen's model in light of your answer to part (d). (2)

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

5. A machine puts liquid into bottles of medicine. The amount of liquid put into each bottle,  $D$  ml, follows a normal distribution with mean 28 ml.

Given that 5% of the bottles contain less than 27.29 ml,

- (a) find, to 2 decimal places, the value of  $k$  such that  $P(27.29 < D < k) = 0.55$ . (5)

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 27.29 ml and  $k$  ml. (3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.7 ml.

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 28 ml.

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 27.95 ml.

- (c) Showing all your working, test Hannah's belief at the 5% level of significance. (5)

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

---

**TOTAL 50 MARKS**