Please check the examination details below before entering your candidate information

Candidate surname	Other names

Pearson Edexcel Level 3 GCE

Centre Number	Candidate Number

Mock Paper – Set 2

Paper Reference **9MA0/31**

Mathematics

Advanced
Paper 31: Statistics

You must have:
Mathematical Formulae and Statistical Tables (Green), 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.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact 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 ▶

S67802A

1/

S 6 7 8 0 2 A 0 1 2 0

Pearson

1. The partially completed table below summarises the times taken by 120 job applicants to complete a task.

Time, *t* (minutes)	$5 < t \leqslant 7$	$7 < t \leqslant 10$	$10 < t \leqslant 14$	$14 < t \leqslant 18$	$18 < t \leqslant 30$
Frequency	10	23	51		

A histogram is drawn.
The bar representing the $5 < t \leqslant 7$ has a width of 1 cm and a height of 5 cm.

(a) Given that the bar representing the group $14 < t \leqslant 18$ has a height of 4 cm, find the frequency of this group.

(2)

(b) Showing your working, estimate the mean time taken by the 120 job applicants.

(3)

The lower quartile of the times is 9.6 minutes and the upper quartile of the times is 15.5 minutes.
For these data, an outlier is classified as any value greater than $Q_3 + 1.5 \times \text{IQR}$.

(c) Showing your working, explain whether or not any of the times taken by these 120 job applicants might be classified as outliers.

(2)

Candidates with the fastest 5% of times for the task are given interviews.

(d) Estimate the time taken by a job applicant, below which they might be given an interview.

(2)

Question 1 continued

(Total for Question 1 is 9 marks)

2. A local sports centre has showers with two temperature settings, warm and hot.

On the warm setting, the water temperature may be modelled by a Normal distribution with mean 30 °C and standard deviation 2 °C

(a) Using the model, find the probability that the next time the shower is used on the warm setting, the water temperature is

(i) exactly 31°C

(ii) more than 31°C

(2)

The sports centre manager thinks that a water temperature of more than 33 °C is too high for the warm setting.
She tests the water temperature on the warm setting on 5 randomly selected days.

Given that the probability of the water temperature being more than 33 °C is 0.0668

(b) find the probability of the water temperature being more than 33 °C

(i) on only the first of these 5 days,

(2)

(ii) on more than 1 of these 5 days.

(3)

On the hot setting, the water temperature may be modelled by a Normal distribution with standard deviation 1.5 °C

The probability that the water temperature is more than 42 °C is 0.0005

(c) Find the mean water temperature on this setting, giving your answer to 1 decimal place.

(4)

Question 2 continued

Question 2 continued

Question 2 continued

(Total for Question 2 is 11 marks)

3. Sam is investigating the weather throughout the year in Hurn.
He uses the large data set to investigate the daily mean wind direction.

Sam believes that each cardinal wind direction is equally likely in Hurn.

(a) Assuming that Sam is correct,

(i) state the probability that the cardinal wind direction in Hurn on a randomly selected day is NNE,

(1)

(ii) state the distribution that Sam should use to model the probability of each cardinal wind direction in Hurn on a randomly selected day.

(1)

Sam decides to investigate the daily mean wind direction throughout the year.

(b) State a limitation of using the data for Hurn from the large data set as a sampling frame.

(1)

(c) Explain how to use simple random sampling to select 36 days from a year.

(2)

Sam defines the random variable X as the number of days out of 36 on which the daily mean wind direction in Hurn is between the bearings 135° and 225°

Sam collects data from 36 randomly selected days and finds that $x = 15$

Sam carries out a hypothesis test at the 10% level of significance.

(d) Given that $H_0 : p = 0.25$ and that the critical region is $\{X \leqslant 4 \cup X \geqslant 14\}$

(i) state the alternative hypothesis, H_1

(ii) giving a reason for your answer, explain what Sam should conclude about the daily mean wind direction in Hurn.

(3)

Question 3 continued

Question 3 continued

Question 3 continued

(Total for Question 3 is 8 marks)

4. A large number of cyclists take part in a cycling time trial.
A random sample of these cyclists are selected and their times, in minutes, are summarised in the following statistics

$$\sum x = 1680 \quad \sum x^2 = 47654.4 \quad n = 60$$

(a) Calculate, for this sample, the value of

(i) the mean time,

(1)

(ii) the standard deviation of the times.

(2)

Historically, the mean time for cyclists on this time trial has been 27 minutes and 30 seconds. Lucy is watching the time trial and believes that the mean time of cyclists in this time trial is greater than the mean time of cyclists in previous time trials.

The times of cyclists on this time trial are modelled by a Normal distribution with standard deviation 3 minutes.

(b) Test, at the 5% level of significance, whether or not this sample provides evidence to support Lucy's belief. You should state your hypotheses and show your working clearly.

(5)

Speedy Wheels cycling club entered its 5 fastest riders and 5 beginners to take part in the time trial.
The fastest 20% of the cyclists in the time trial are invited to compete in a race the following week.

(c) (i) Explain, with specific reference to the parameter p, why the distribution B(10, 0.2) might not be reasonable to model the number of these *Speedy Wheels* cycling club members who are invited to compete in the race.

(2)

(ii) Suggest how to improve the model for the number of these *Speedy Wheels* cycling club members invited to compete in the race.

(1)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 11 marks)

5. (i) Two events A and B are mutually exclusive.

Given that $P(B) = p \neq 0$ and

$$P(A) = 3 \times P(B)$$

(a) draw a Venn diagram to illustrate this information,

(b) find the possible values of $P(B)$

(3)

(ii) Two events C and D are such that

$$P(C \mid D) = 3 \times P(C) \text{ where } P(C) \neq 0$$

(a) Explain whether or not events C and D could be independent events.

(1)

Given also that

$$P(C \cap D) = \frac{1}{2} \times P(C) \text{ and } P(C' \cap D') = \frac{7}{10}$$

(b) find $P(C)$, showing your working clearly.

(7)

Question 5 continued

Question 5 continued

Question 5 continued

Question 5 continued

(Total for Question 5 is 11 marks)

TOTAL FOR STATISTICS IS 50 MARKS