

Please check the examination details below before entering your candidate information

Candidate surname

Other names

**Pearson Edexcel
Level 3 GCE**

Centre Number

--	--	--	--	--

Candidate Number

--	--	--	--	--

Monday 19 October 2020

Afternoon

Paper Reference **9MA0/31**

Mathematics

Advanced

Paper 31: Statistics Shadow Paper Set 1

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.

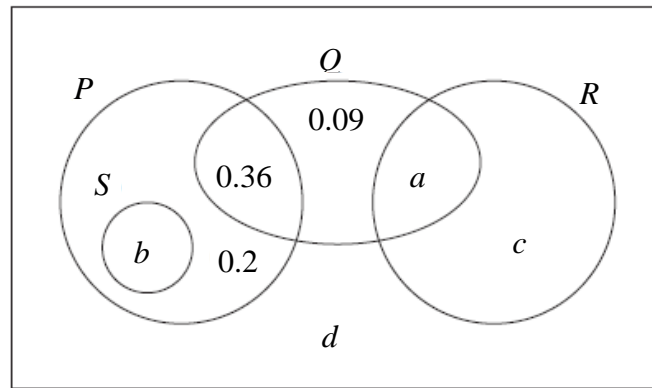
Information

- 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.

1 The Venn diagram shows the probabilities associated with four events, P , Q , R and S



(a) Write down any pair of mutually exclusive events from P , Q , R and S (1)

Given that $P(Q) = 0.5$

(b) find the value of a (1)

Given also that P and Q are independent

(c) find the value of b (2)

Given further that $P(Q' | R) = 0.6$

(d) find (4)

- (i) the value of c
- (ii) the value of d

(Total for Question 1 is 8 marks)

- 2 A random sample of 15 days is taken from the large data set for Perth in June and July 1987. The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

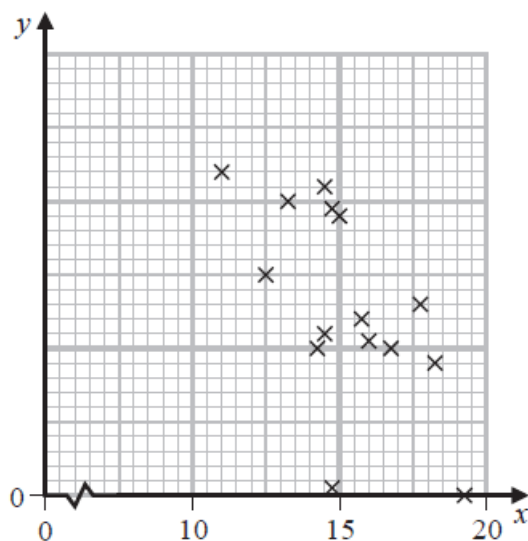


Figure 1

- (a) Describe the correlation. (1)

The variable on the x -axis is Daily Mean Temperature measured in $^{\circ}\text{C}$.

- (b) Using your knowledge of the large data set,
 (i) suggest which variable is on the y -axis,
 (ii) state the units that are used in the large data set for this variable. (2)

Priti believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

She calculates the product moment correlation coefficient between these two variables for a random sample of 25 days and obtains $r = -0.377$

- (c) Carry out a suitable test to investigate Priti's belief at a 5% level of significance. State clearly
 • your hypotheses
 • your critical value (3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 95%

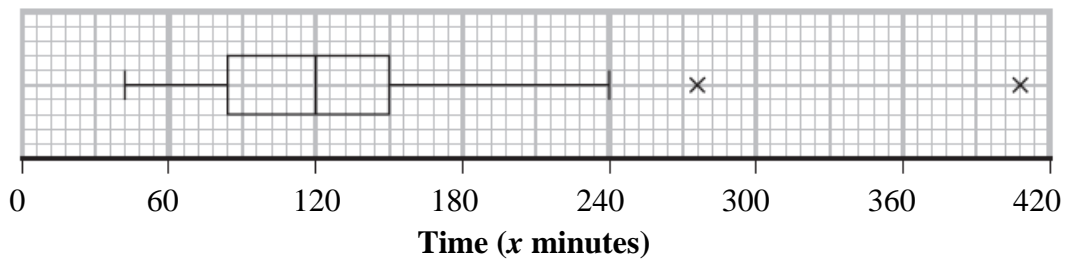
- (d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer. (1)

(Total for Question 2 is 7 marks)

3 Each member of a group of 35 people went walking on a weekend.

The time taken for the walk, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

For these 35 people $\sum x = 3850$ and $\sum x^2 = 575960$

(c) calculate the mean time taken on the walk,

(1)

(d) calculate the standard deviation of the times taken on the walk.

(2)

Louise defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Louise would say there are in these data, giving a reason for your answer.

(1)

Alana and Buda also went walking for a minutes and b minutes respectively, where $a > b$.

When their times are included with the data of the other 35 people

- the median time decreases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b , explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 37 times will be lower than your answer to part (d).

(1)

(Total for Question 3 is 10 marks)

4 The discrete random variable D has the following probability distribution

d	5	10	15	20	25
$P(D = d)$	$\frac{c}{5}$	$\frac{c}{10}$	$\frac{c}{15}$	$\frac{c}{20}$	$\frac{c}{25}$

where c is a constant.

(a) Show that the value of c is $\frac{300}{137}$ (2)

The random variables D_1 and D_2 are independent and each have the same distribution as D .

(b) Find $P(D_1 + D_2 = 40)$.
Give your answer to 3 significant figures. (3)

A single observation of D is made.

The value obtained, d , is the common difference of an arithmetic sequence.

The first 4 terms of this arithmetic sequence are the angles, measured in degrees, of quadrilateral Q

(c) Find the exact probability that the smallest angle of Q is more than 65° (5)

(Total for Question 4 is 10 marks)

5 A health centre claims that the time a dentist spends with a patient can be modelled by a normal distribution with a mean of 25 minutes and a standard deviation of 12 minutes.

- (a) Using this model, find the probability that the time spent with a randomly selected patient is more than 40 minutes. (1)

Some patients complain that the mean time the dentist spends with a patient is more than 25 minutes.

The receptionist takes a random sample of 30 patients and finds that the mean time the doctor spends with a patient is 28 minutes.

- (b) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint. (4)

The health centre also claims that the time a receptionist spends registering a patient for an appointment, T minutes, can be modelled by the normal distribution where $T \sim N(4, 2.5^2)$

- (c) Using this model,
- (i) find the probability that a registration with the receptionist takes less than 3 minutes, (1)
 - (ii) find $P(T < 3 \mid T > 0)$ (3)
 - (iii) hence explain why this normal distribution may not be a good model for T . (1)

The receptionist believes that she cannot complete registration in less than 3 minutes.

She suggests that the dental centre should use a refined model only including values of $T > 3$.

- (d) Find the median time to be registered using this new model, giving your answer correct to one decimal place. (5)

(Total for Question 5 is 15 marks)

TOTAL FOR STATISTICS IS 50 MARKS