Write your name here

Surname	Other names

Pearson Edexcel
International
Advanced Level

Centre Number

Candidate Number

Statistics S1

Advanced/Advanced Subsidiary

Wednesday 7 June 2017 – Morning
Time: 1 hour 30 minutes

Paper Reference
WST01/01

You must have:
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. 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). Coloured pencils and highlighter pens must not be used.
- **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 the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

Information

- The total mark for this paper is 75.
- 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

Pearson

Leave blank

1. Nina weighed a random sample of 50 carrots from her shop and recorded the weight, in grams to the nearest gram, for each carrot. The results are summarised below.

Weight of carrot	**Frequency (f)**	**Weight midpoint (*x* grams)**
45–54	5	49.5
55–59	10	57
60–64	22	62
65–74	13	69.5

(You may use $\sum fx^2 = 192\,102.5$)

(a) Use linear interpolation to estimate the median weight of these carrots.
(2)

(b) Find an estimate for the mean weight of these carrots.
(2)

(c) Find an estimate for the standard deviation of the weights of these carrots.
(2)

A carrot is selected at random from Nina's shop.

(d) Estimate the probability that the weight of this carrot is more than 70 grams.
(2)

Leave blank

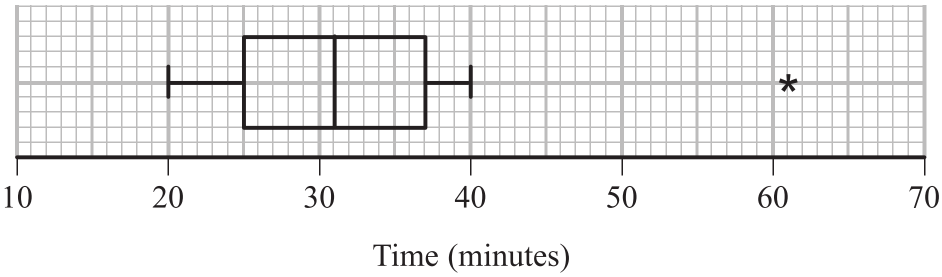
Question 1 continued

Q1

(Total 8 marks)

Turn over

2. The box plot shows the times, t minutes, it takes a group of office workers to travel to work.



(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

(c) Using the quartiles, describe the skewness of these data. Give a reason for your answer.

(2)

Chetna believes that house prices will be higher if the time to travel to work is shorter. She asks a random sample of these office workers for their house prices £x, where x is measured in thousands, and obtains the following statistics

$$S_{xx} = 5514 \qquad S_{xt} = 10 \qquad S_{tt} = 1145.6$$

(d) Calculate the product moment correlation coefficient between x and t.

(2)

(e) State, giving a reason, whether or not your correlation coefficient supports Chetna's belief.

(2)

Adam and Betty are part of the group of office workers and they have both moved house.
Adam's time to travel to work changes from 32 minutes to 36 minutes.
Betty's time to travel to work changes from 38 minutes to 58 minutes.
Outliers are defined as values that are more than 1.5 times the interquartile range above the upper quartile.

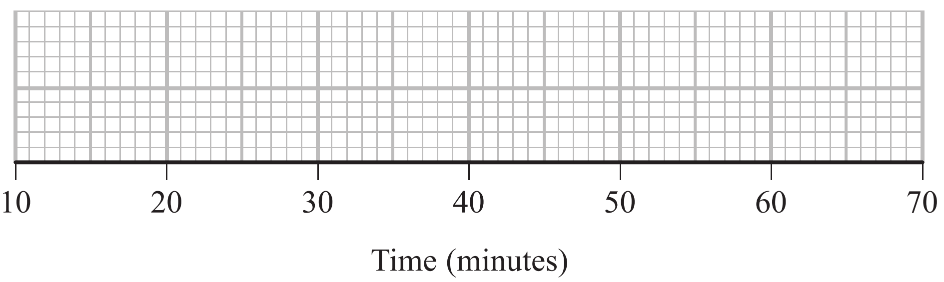
(f) Showing all necessary calculations, determine how the box plot of times to travel to work will change and draw a new box plot on the grid on page 5.

(3)

P 5 0 7 1 5 A 0 4 2 4

Leave blank

Question 2 continued



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

Leave blank

Question 2 continued

DO NOT WRITE IN THIS AREA

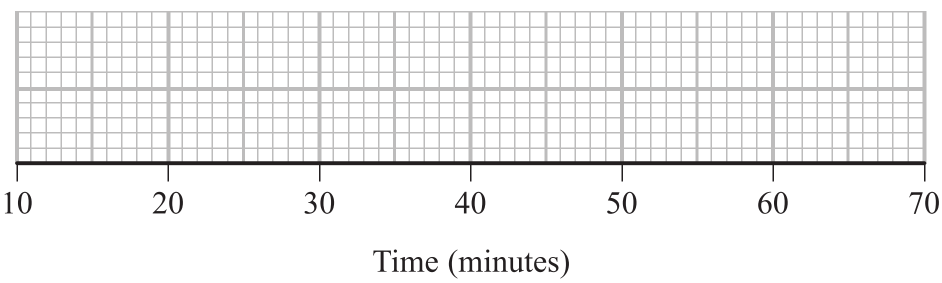
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 2 continued

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



Q2

(Total 11 marks)

P50715A0724

Turn over

Leave blank

3. At a school athletics day, the distances, in metres, achieved by students in the long jump are modelled by the normal distribution with mean 3.3 m and standard deviation 0.6 m

(a) Find an estimate for the proportion of students who jump less than 2.5 m

(3)

The long jump competition consists of 2 jumps. All the students can take part in the first jump and the 40% who jump the greatest distance in their first jump qualify for the second jump.

(b) Find an estimate for the minimum distance achieved in the first jump in order to qualify for the second jump.
Give your answer correct to 4 significant figures.

(3)

(c) Find an estimate for the median distance achieved in the first jump by those who qualify for the second jump.

(3)

The distance of the second jump is independent of the distance of the first jump and is modelled with the same normal distribution. Students who jump a distance greater than 4.1 m in their second jump receive a certificate.

At the start of the long jump competition, a student is selected at random.

(d) Find the probability that this student will receive a certificate.

(3)

Leave blank

Question 3 continued

Leave blank

Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

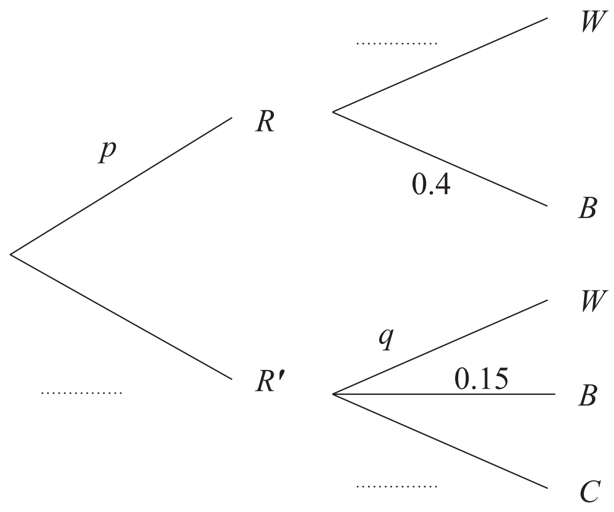
Question 3 continued

Q3

(Total 12 marks)

Leave blank

4. The partially completed tree diagram, where p and q are probabilities, gives information about Andrew's journey to work each day.



R represents the event that it is raining

W represents the event that Andrew walks to work

B represents the event that Andrew takes the bus to work

C represents the event that Andrew cycles to work

Given that $P(B) = 0.26$

(a) find the value of p

(3)

Given also that $P(R' \mid W) = 0.175$

(b) find the value of q

(4)

(c) Find the probability that Andrew cycles to work.

(2)

Given that Andrew did not cycle to work on Friday,

(d) find the probability that it was raining on Friday.

(3)

Leave blank

Question 4 continued

P 5 0 7 1 5 A 0 1 3 2 4

Leave blank

Question 4 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 4 continued

(Total 12 marks)

Q4

Leave blank

5. Tomas is studying the relationship between temperature and hours of sunshine in *Seapron*. He records the midday temperature, t °C, and the hours of sunshine, s hours, for a random sample of 9 days in October. He calculated the following statistics

$$\sum s = 15 \qquad \sum s^2 = 44.22 \qquad \sum t = 127 \qquad S_{tt} = 10.89$$

(a) Calculate S_{ss}

(2)

Tomas calculated the product moment correlation coefficient between s and t to be 0.832 correct to 3 decimal places.

(b) State, giving a reason, whether or not this correlation coefficient supports the use of a linear regression model to describe the relationship between midday temperature and hours of sunshine.

(1)

(c) State, giving a reason, why the hours of sunshine would be the explanatory variable in a linear regression model between midday temperature and hours of sunshine.

(1)

(d) Find S_{st}

(3)

(e) Calculate a suitable linear regression equation to model the relationship between midday temperature and hours of sunshine.

(4)

(f) Calculate the standard deviation of s

(1)

Tomas uses this model to estimate the midday temperature in *Seapron* for a day in October with 5 hours of sunshine.

(g) State the value of Tomas' estimate.

(1)

Given that the values of s are all within 2 standard deviations of the mean,

(h) comment, giving your reason, on the reliability of this estimate.

(2)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

Question 5 continued

Q5

(Total 15 marks)

Leave blank

6. A biased coin has probability 0.4 of showing a head. In an experiment, the coin is spun until a head appears. If a head has not appeared after 4 spins, the coin is not spun again. The random variable X represents the number of times the coin is spun.

For example, $X = 3$ if the first two spins do not show a head but the third spin does show a head. The coin would not then be spun a fourth time since the coin has already shown a head.

(a) Show that $P(X = 3) = 0.144$

(1)

The table gives some values for the probability distribution of X

x	1	2	3	4
$P(X = x)$		0.24	0.144	

(b) (i) Write down the value of $P(X = 1)$

(ii) Find $P(X = 4)$

(3)

(c) Find $E(X)$

(2)

(d) Find $Var(X)$

(3)

The random variable H represents the number of heads obtained when the coin is spun in the experiment.

(e) Explain why H can only take the values 0 and 1 and find the probability distribution of H.

(2)

(f) Write down the value of

(i) $P(\{X = 3\} \cap \{H = 0\})$

(ii) $P(\{X = 4\} \cap \{H = 0\})$

(2)

The random variable $S = X + H$

(g) Find the probability distribution of S

(4)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

P 5 0 7 1 5 A 0 2 3 2 4

Leave blank

Question 6 continued

Q6

(Total 17 marks)

TOTAL FOR PAPER: 75 MARKS

END

P 5 0 7 1 5 A 0 2 4 2 4