Write your name here

Surname	Other names

Pearson Edexcel
International
Advanced Level

Centre Number | | | | | |

Candidate Number | | | | |

Statistics S1

Advanced/Advanced Subsidiary

Friday 5 June 2015 – 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 ▶

P44844A

5/1/1/1/1/

P 4 4 8 4 4 A 0 1 2 8

PEARSON

Leave blank

1. The discrete random variable X can only take the values 1, 2, 3 and 4
For these values the cumulative distribution function is defined by

$$F(x) = kx^2 \quad \text{for} \quad x = 1, 2, 3, 4$$

where k is a constant.

(a) Find the value of k.

(2)

(b) Find the probability distribution of X.

(2)

Leave blank

Question 1 continued

Q1

(Total 4 marks)

Leave blank

2. Paul believes there is a relationship between the value and the floor size of a house. He takes a random sample of 20 houses and records the value, £v, and the floor size, s m^2

The data were coded using $x = \dfrac{s - 50}{10}$ and $y = \dfrac{v}{100\,000}$ and the following statistics obtained.

$$\sum x = 441.5, \quad \sum y = 59.8, \quad \sum x^2 = 11\,261.25, \quad \sum y^2 = 196.66, \quad \sum xy = 1474.1$$

(a) Find the value of S_{xy} and the value of S_{xx}

(3)

(b) Find the equation of the least squares regression line of y on x in the form $y = a + bx$

(3)

The least squares regression line of v on s is $v = c + ds$

(c) Show that $d = 1020$ to 3 significant figures and find the value of c

(3)

(d) Estimate the value of a house of floor size 130 m^2

(2)

(e) Interpret the value d

(1)

Paul wants to increase the value of his house. He decides to add an extension to increase the floor size by 31 m^2

(f) Estimate the increase in the value of Paul's house after adding the extension.

(1)

Leave blank

Question 2 continued

Leave blank

Question 2 continued

Leave blank

Question 2 continued

(Total 13 marks)

Q2

Leave blank

3. A company employs 90 administrators. The length of time that they have been employed by the company and their gender are summarised in the table below.

Length of time employed, x years	Female	Male
$x < 4$	9	16
$4 \leqslant x < 10$	14	20
$10 \leqslant x$	7	24

One of the 90 administrators is selected at random.

(a) Find the probability that the administrator is female.

(1)

(b) Given that the administrator has been employed by the company for less than 4 years, find the probability that this administrator is male.

(2)

(c) Given that the administrator has been employed by the company for less than 10 years, find the probability that this administrator is male.

(2)

(d) State, with a reason, whether or not the event 'selecting a male' is independent of the event 'selecting an administrator who has been employed by the company for less than 4 years'.

(3)

Leave blank

Question 3 continued

Q3

(Total 8 marks)

Leave blank

4. A bag contains 19 red beads and 1 blue bead only.

Linda selects a bead at random from the bag. She notes its colour and replaces the bead in the bag. She then selects a second bead at random from the bag and notes its colour.

Find the probability that

(a) both beads selected are blue,

(1)

(b) exactly one bead selected is red.

(2)

In another bag there are 9 beads, 4 of which are green and the rest are yellow.

Linda selects 3 beads from this bag at random without replacement.

(c) Find the probability that 2 of these beads are yellow and 1 is green.

(3)

Linda replaces the 3 beads and then selects another 4 at random without replacement.

(d) Find the probability that at least 1 of the beads is green.

(3)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 9 marks)

Leave blank

5. Police measure the speed of cars passing a particular point on a motorway. The random variable X is the speed of a car.

X is modelled by a normal distribution with mean 55 mph (miles per hour).

(a) Draw a sketch to illustrate the distribution of X. Label the mean on your sketch.

(2)

The speed limit on the motorway is 70 mph. Car drivers can choose to travel faster than the speed limit but risk being caught by the police.

The distribution of X has a standard deviation of 20 mph.

(b) Find the percentage of cars that are travelling faster than the speed limit.

(3)

The fastest 1% of car drivers will be banned from driving.

(c) Show that the lowest speed, correct to 3 significant figures, for a car driver to be banned is 102 mph. Show your working clearly.

(3)

Car drivers will just be given a caution if they are travelling at a speed m such that

$$P(70 < X < m) = 0.1315$$

(d) Find the value of m. Show your working clearly.

(4)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 12 marks)

Leave blank

6. The random variable X has a discrete uniform distribution and takes the values 1, 2, 3, 4

Find

(a) F(3), where F(x) is the cumulative distribution function of X,

(1)

(b) E(X).

(1)

(c) Show that $\text{Var}(X) = \frac{5}{4}$

(3)

The random variable Y has a discrete uniform distribution and takes the values

$$3,\ 3 + k,\ 3 + 2k,\ 3 + 3k$$

where k is a constant.

(d) Write down $\text{P}(Y = y)$ for $y = 3,\ 3 + k,\ 3 + 2k,\ 3 + 3k$

(1)

The relationship between X and Y may be written in the form $Y = kX + c$ where c is a constant.

(e) Find $\text{Var}(Y)$ in terms of k.

(2)

(f) Express c in terms of k.

(1)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 9 marks)

Leave blank

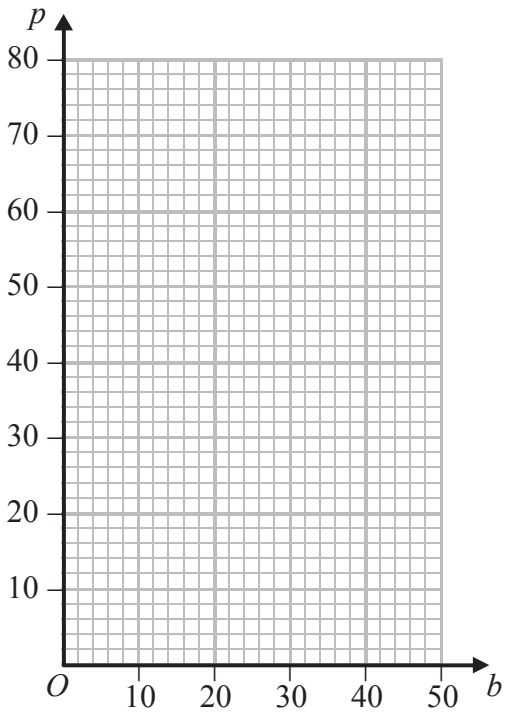
7. A doctor is investigating the correlation between blood protein, p, and body mass index, b.

He takes a random sample of 8 patients and the data are shown in the table below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
b	32	36	40	44	42	21	27	37
p	18	21	31	39	21	12	19	70

(a) Draw a scatter diagram of these data on the axes provided.

(2)



The doctor decides to leave out patient H from his calculations.

(b) Give a reason for the doctor's decision.

(1)

For the 7 patients A, B, C, D, E, F and G,

$$S_{bp} = 369, \quad S_{pp} = 490 \quad \text{and} \quad S_{bb} = 423\frac{5}{7}$$

(c) Find the product moment correlation coefficient, r, for these 7 patients.

(2)

(d) Without any further calculations, state how r would differ from your answer in part (c) if it was calculated for all 8 patients.

(1)

Leave blank

Question 7 continued

Q7

(Total 6 marks)

Leave blank

8.

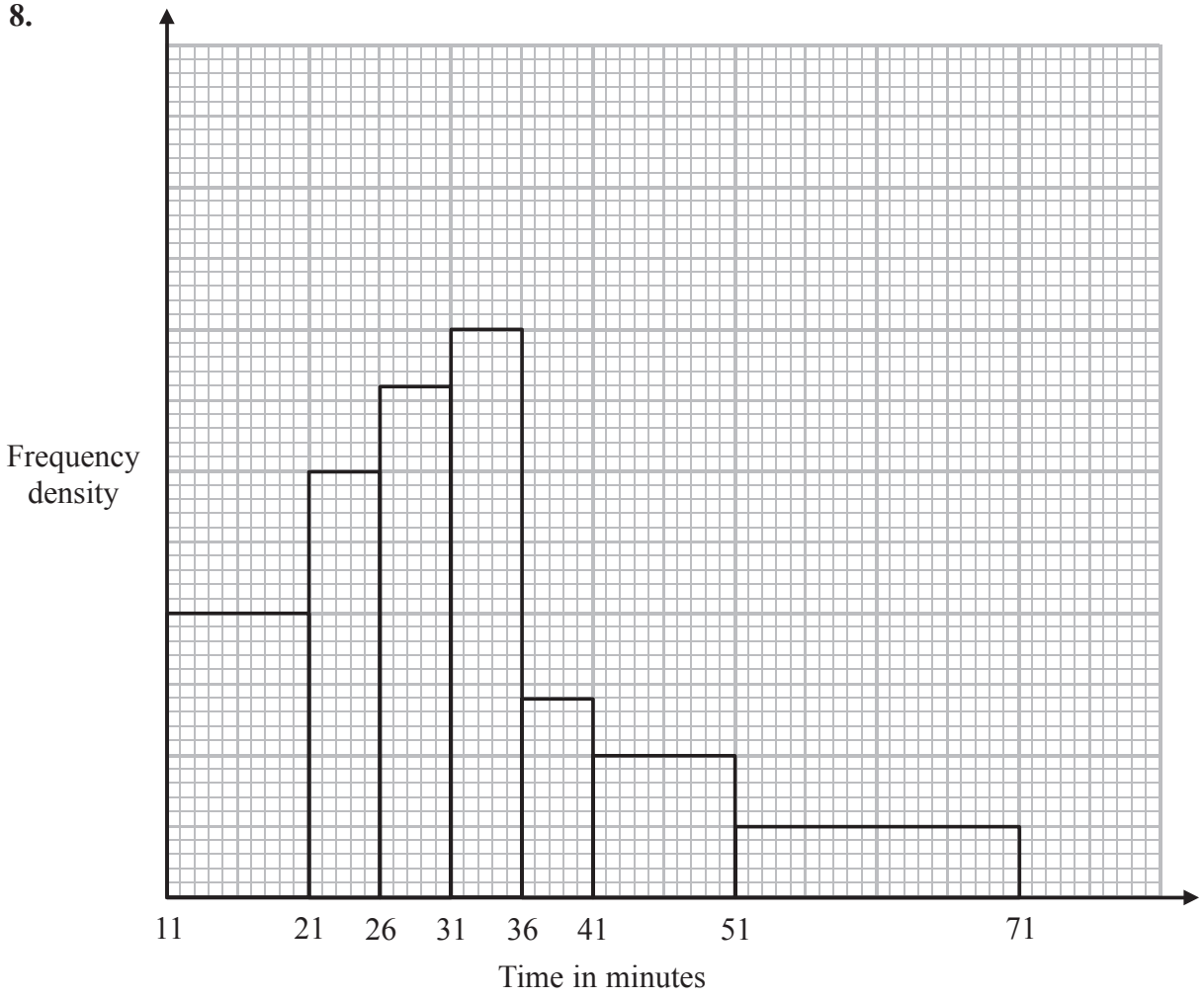


Figure 1

The histogram in Figure 1 summarises the times, in minutes, that 200 people spent shopping in a supermarket.

(a) Give a reason to justify the use of a histogram to represent these data.

(1)

Given that 40 people spent between 11 and 21 minutes shopping in the supermarket, estimate

(b) the number of people that spent between 18 and 25 minutes shopping in the supermarket,

(3)

(c) the median time spent shopping in the supermarket by these 200 people.

(2)

The mid-point of each bar is represented by x and the corresponding frequency by f.

(d) Show that $\sum fx = 6390$

(2)

Leave blank

Question 8 continued

Given that $\sum fx^2 = 238\,430$

(e) for the data shown in the histogram, calculate estimates of

(i) the mean,

(ii) the standard deviation.

(3)

A coefficient of skewness is given by $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

(f) Calculate this coefficient of skewness for these data.

(1)

The manager of the supermarket decides to model these data with a normal distribution.

(g) Comment on the manager's decision. Give a justification for your answer.

(2)

Leave blank

Question 8 continued

Leave blank

Question 8 continued

Leave blank

Question 8 continued

Q8

(Total 14 marks)

TOTAL FOR PAPER: 75 MARKS

END