Write your name here

Surname	Other names

Pearson Edexcel
International
Advanced Level

Centre Number | | | | | |

Candidate Number | | | | |

Statistics S1

Advanced/Advanced Subsidiary

Friday 17 January 2014 – Afternoon
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 ▶

P43140A

5/5/5/5/5/

P 4 3 1 4 0 A 0 1 2 8

PEARSON

Leave blank

1. A price comparison website publishes data on the cost per month, £c, and the level of satisfaction, s, of a random sample of six internet service providers. A low value of s corresponds to a low level of satisfaction. The data are given in the table below.

Internet service provider	*A*	*B*	*C*	*D*	*E*	*F*
c	20	15	12	30	9	25
s	5	3	4	2	3	4

(You may use $\sum c = 111$, $\sum c^2 = 2375$, $\sum s = 21$, $\sum s^2 = 79$, $\sum cs = 380$, $S_{cc} = 321.5$)

(a) Calculate the value of S_{cs} and the value of S_{ss}

(3)

(b) Calculate the product moment correlation coefficient for these data.

(2)

Brad is not satisfied with his current internet service and decides to change his provider. He decides to pay a lot more for his new internet service.

(c) On the basis of your calculation in part (b), comment on Brad's decision. Give a reason for your answer.

(2)

Leave blank

Question 1 continued

Q1

(Total 7 marks)

Leave blank

2. A rugby club coach uses club records to take a random sample of 15 players from 1990 and an independent random sample of 15 players from 2010. The body weight of each player was recorded to the nearest kg and the results from 2010 are summarised in the table below.

Body weight (kg)	75–79	80–84	85–89	90–94	95–99	100–104	105–109
Number of Players (2010)	1	2	2	4	3	2	1

(a) Find the estimated values in kg of the summary statistics *a*, *b* and *c* in the table below.

	Estimate in 1990	**Estimate in 2010**
Mean	83.0	*a*
Median	82.0	*b*
Variance	44.0	*c*

Give your answers to 3 significant figures.

(6)

The rugby coach claims that players' body weight increased between 1990 and 2010.

(b) Using the table in part (a), comment on the rugby coach's claim.

(2)

Leave blank

Question 2 continued

Q2

(Total 8 marks)

Leave blank

3. Jean works for an insurance company. She randomly selects 8 people and records the price of their car insurance, £p, and the time, t years, since they passed their driving test. The data is shown in the table below.

t	10	13	17	18	22	24	25	27
p	720	650	430	490	500	390	280	300

(You may use $\bar{t} = 19.5$, $\bar{p} = 470$, $S_{tp} = -6080$, $S_{tt} = 254$, $S_{pp} = 169\,200$)

(a) On the graph below draw a scatter diagram for these data.

(2)

(b) Comment on the relationship between p and t.

(1)

(c) Find the equation of the regression line of p on t.

(4)

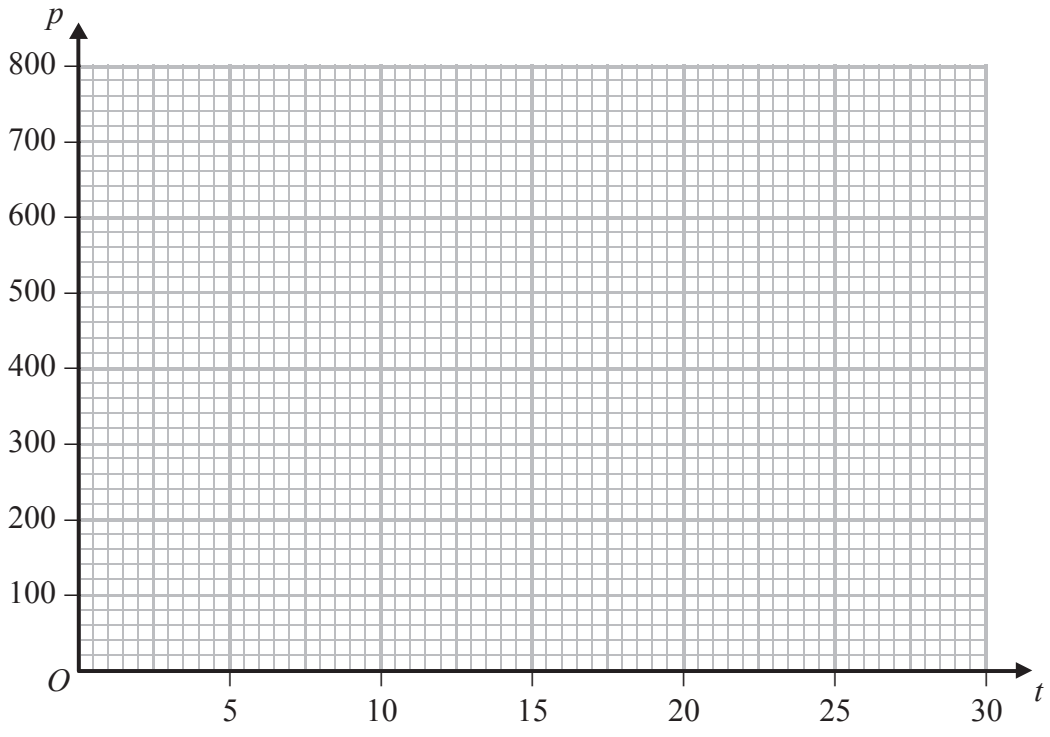
(d) Use your regression equation to estimate the price of car insurance for someone who passed their driving test 20 years ago.

(2)

Jack passed his test 39 years ago and decides to use Jean's data to predict the price of his car insurance.

(e) Comment on Jack's decision. Give a reason for your answer.

(2)



Leave blank

Question 3 continued

Leave blank

Question 3 continued

Leave blank

Question 3 continued

Q3

(Total 11 marks)

Leave blank

4. A discrete random variable X has the probability distribution given in the table below, where a and b are constants.

x	-1	0	1	2	3
$P(X = x)$	a	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	b

Given $E(X) = \frac{9}{5}$

(a) (i) find two simultaneous equations for a and b,

(ii) show that $a = \frac{1}{20}$ and find the value of b.

(4)

(b) Specify the cumulative distribution function $F(x)$ for $x = -1, 0, 1, 2$ and 3

(2)

(c) Find $P(X < 2.5)$.

(1)

(d) Find $\text{Var}(3 - 2X)$.

(4)

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Leave blank

Question 4 continued

Q4

(Total 11 marks)

Leave blank

5. A group of 100 students are asked if they like folk music, rock music or soul music.

All students who like folk music also like rock music

No students like both rock music and soul music

75 students do not like soul music

12 students who like rock music do not like folk music

30 students like folk music

(a) Draw a Venn diagram to illustrate this information.

(4)

(b) State two of these types of music that are mutually exclusive.

(1)

Find the probability that a randomly chosen student

(c) does not like folk music, rock music or soul music,

(1)

(d) likes rock music,

(1)

(e) likes folk music or soul music.

(1)

Given that a randomly chosen student likes rock music,

(f) find the probability that he or she also likes folk music.

(2)

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Leave blank

Question 5 continued

Q5

(Total 10 marks)

Leave blank

6. A manufacturer has a machine that fills bags with flour such that the weight of flour in a bag is normally distributed. A label states that each bag should contain 1 kg of flour.

(a) The machine is set so that the weight of flour in a bag has mean 1.04 kg and standard deviation 0.17 kg. Find the proportion of bags that weigh less than the stated weight of 1 kg.

(3)

The manufacturer wants to reduce the number of bags which contain less than the stated weight of 1 kg. At first she decides to adjust the mean but not the standard deviation so that only 5% of the bags filled are below the stated weight of 1 kg.

(b) Find the adjusted mean.

(3)

The manufacturer finds that a lot of the bags are overflowing with flour when the mean is adjusted, so decides to adjust the standard deviation instead to make the machine more accurate. The machine is set back to a mean of 1.04 kg. The manufacturer wants 1% of bags to be under 1 kg.

(c) Find the adjusted standard deviation. Give your answer to 3 significant figures.

(3)

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Leave blank

Question 6 continued

Q6

(Total 9 marks)

Leave blank

7. In a large college, $\frac{3}{5}$ of the students are male, $\frac{3}{10}$ of the students are left handed and $\frac{1}{5}$ of the male students are left handed.

A student is chosen at random.

(a) Given that the student is left handed, find the probability that the student is male.

(2)

(b) Given that the student is female, find the probability that she is left handed.

(3)

(c) Find the probability that the randomly chosen student is male and right handed.

(2)

Two students are chosen at random.

(d) Find the probability that one student is left handed and one is right handed.

(2)

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Leave blank

Question 7 continued

Q7

(Total 9 marks)

Leave blank

8. A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

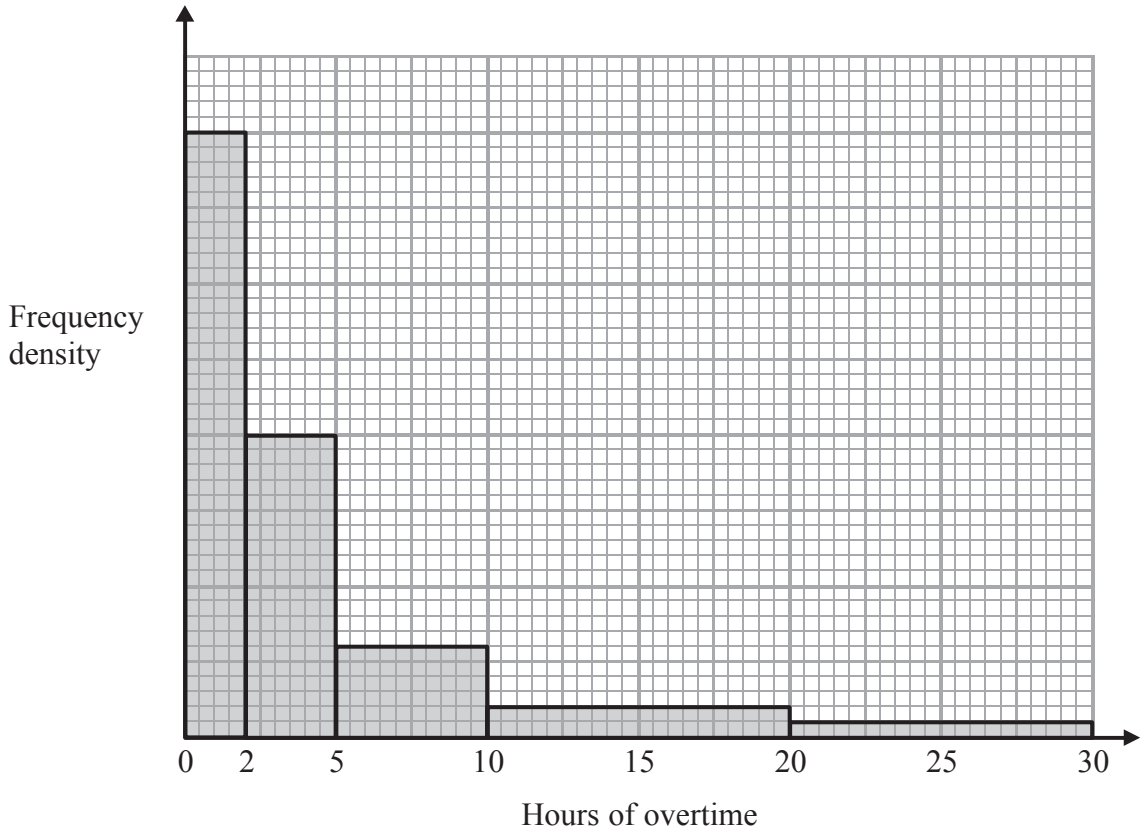


Figure 1

(a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month.

(4)

(b) Estimate the median number of hours of overtime claimed by these 40 staff in the month.

(2)

(c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month.

(2)

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

(d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff.

(2)

Leave blank

Question 8 continued

Leave blank

Question 8 continued

Q8

(Total 10 marks)

TOTAL FOR PAPER: 75 MARKS

END