Centre No.						Paper Reference							Surname	Initial(s)
Candidate No.						6	6	8	3	/	0	1	Signature	

Paper Reference(s)

6683/01

Edexcel GCE

Statistics S1

Advanced/Advanced Subsidiary

Tuesday 10 June 2014 – Morning

Time: 1 hour 30 minutes

Examiner's use only

Team Leader's use only

Question Number	Leave Blank
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
Nil

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 or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates

In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
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 for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 28 pages in this question paper. Any blank pages are indicated.

Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.

This publication may be reproduced only in accordance with Pearson Education Ltd copyright policy.
©2014 Pearson Education Ltd.

Printer's Log. No.
P43017A
W850/R6683/57570 5/5/5/1/

P 4 3 0 1 7 A 0 1 2 8

Turn over

PEARSON

Leave blank

1. A random sample of 35 homeowners was taken from each of the villages Greenslax and Penville and their ages were recorded. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Greenslax									Penville											Totals
(2)							8	7	2	5	5	6	7	8	8	9					(7)
(3)						9	8	7	3	1	1	1	2	3	4	4	5	6	7	9	(11)
(4)					4	4	4	0	4	0	1	2	4	7							(5)
(5)				6	6	5	2	2	5	0	0	5	5	5							(5)
(7)		8	6	5	4	2	1	1	6	2	5	6	6								(4)
(8)	8	6	6	6	4	3	1	1	7	0	5										(2)
(5)				9	8	4	3	2	8												(0)
(1)								4	9	9											(1)

Key: 7 | 3 | 1 means 37 years for Greenslax and 31 years for Penville

Some of the quartiles for these two distributions are given in the table below.

	Greenslax	**Penville**
Lower quartile, Q_1	a	31
Median, Q_2	64	39
Upper quartile, Q_3	b	55

(a) Find the value of a and the value of b.

(2)

An outlier is a value that falls either

more than $1.5 \times (Q_3 - Q_1)$ above Q_3

or more than $1.5 \times (Q_3 - Q_1)$ below Q_1

(b) On the graph paper opposite draw a box plot to represent the data from Penville. Show clearly any outliers.

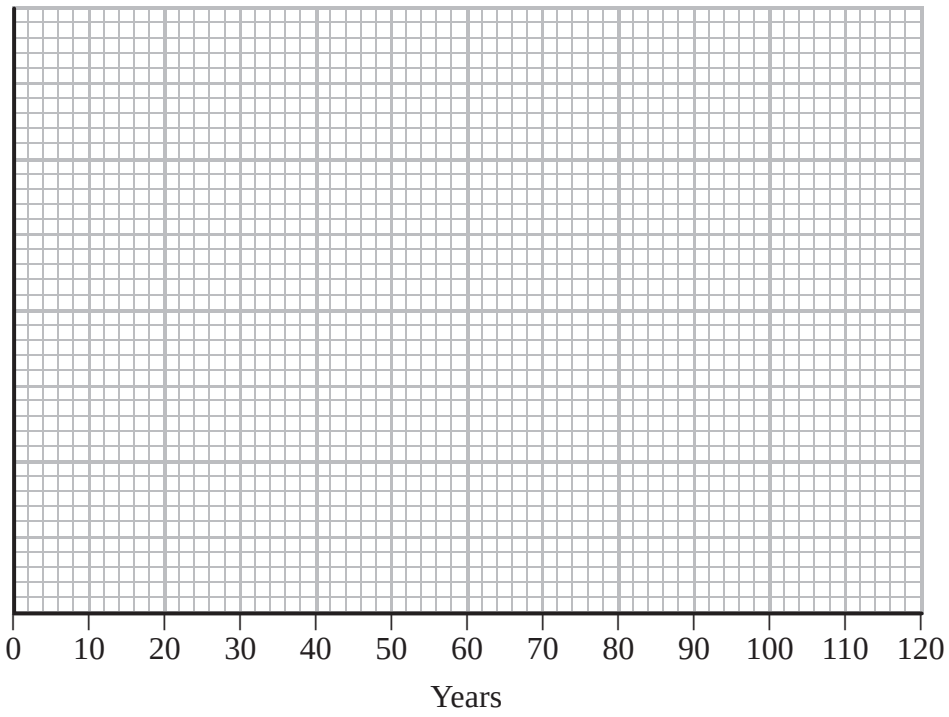
(4)

(c) State the skewness of each distribution. Justify your answers.

(3)

Leave blank

Question 1 continued



Leave blank

2. The mark, x, scored by each student who sat a statistics examination is coded using

$$y = 1.4x - 20$$

The coded marks have mean 60.8 and standard deviation 6.60

Find the mean and the standard deviation of x.

(4)

Q2

(Total 4 marks)

Leave blank

3. The table shows data on the number of visitors to the UK in a month, v (1000s), and the amount of money they spent, m (£ millions), for each of 8 months.

Number of visitors v (1000s)	2450	2480	2540	2420	2350	2290	2400	2460
Amount of money spent m (£ millions)	1370	1350	1400	1330	1270	1210	1330	1350

You may use

$S_{vv} = 42587.5 \quad S_{vm} = 31512.5 \quad S_{mm} = 25187.5 \quad \sum v = 19390 \quad \sum m = 10610$

(a) Find the product moment correlation coefficient between m and v.
(2)

(b) Give a reason to support fitting a regression model of the form $m = a + bv$ to these data.
(1)

(c) Find the value of b correct to 3 decimal places.
(2)

(d) Find the equation of the regression line of m on v.
(2)

(e) Interpret your value of b.
(2)

(f) Use your answer to part (d) to estimate the amount of money spent when the number of visitors to the UK in a month is 2 500 000
(2)

(g) Comment on the reliability of your estimate in part (f). Give a reason for your answer.
(2)

Leave blank

Question 3 continued

Leave blank

4. In a factory, three machines, J, K and L, are used to make biscuits.

Machine J makes 25% of the biscuits.

Machine K makes 45% of the biscuits.

The rest of the biscuits are made by machine L.

It is known that 2% of the biscuits made by machine J are broken, 3% of the biscuits made by machine K are broken and 5% of the biscuits made by machine L are broken.

(a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities. **(2)**

A biscuit is selected at random.

(b) Calculate the probability that the biscuit is made by machine J and is not broken. **(2)**

(c) Calculate the probability that the biscuit is broken. **(2)**

(d) Given that the biscuit is broken, find the probability that it was not made by machine K. **(3)**

Leave blank

Question 4 continued

Leave blank

5. The discrete random variable X has the probability function

$$\mathrm{P}(X = x) = \begin{cases} kx & x = 2, 4, 6 \\ k(x-2) & x = 8 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Show that $k = \dfrac{1}{18}$ **(2)**

(b) Find the exact value of F(5). **(1)**

(c) Find the exact value of E(X). **(2)**

(d) Find the exact value of E(X^2). **(2)**

(e) Calculate Var(3 – 4X) giving your answer to 3 significant figures. **(3)**

Leave blank

Question 5 continued

Leave blank

6. The times, in seconds, spent in a queue at a supermarket by 85 randomly selected customers, are summarised in the table below.

Time (seconds)	Number of customers, f
0 − 30	2
30 − 60	10
60 − 70	17
70 − 80	25
80 − 100	25
100 − 150	6

A histogram was drawn to represent these data. The 30 − 60 group was represented by a bar of width 1.5 cm and height 1 cm.

(a) Find the width and the height of the 70 − 80 group.

(3)

(b) Use linear interpolation to estimate the median of this distribution.

(2)

Given that x denotes the midpoint of each group in the table and

$$\sum fx = 6460 \qquad \sum fx^2 = 529\,400$$

(c) calculate an estimate for

(i) the mean,

(ii) the standard deviation,

for the above data.

(3)

One measure of skewness is given by

$$\text{coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

(d) Evaluate this coefficient and comment on the skewness of these data.

(3)

Leave blank

Question 6 continued

Leave blank

7. The heights of adult females are normally distributed with mean 160 cm and standard deviation 8 cm.

(a) Find the probability that a randomly selected adult female has a height greater than 170 cm.

(3)

Any adult female whose height is greater than 170 cm is defined as tall.

An adult female is chosen at random. Given that she is tall,

(b) find the probability that she has a height greater than 180 cm.

(4)

Half of tall adult females have a height greater than h cm.

(c) Find the value of h.

(5)

Leave blank

Question 7 continued

Leave blank

8. For the events A and B,

$$P(A' \cap B) = 0.22 \text{ and } P(A' \cap B') = 0.18$$

(a) Find $P(A)$.

(1)

(b) Find $P(A \cup B)$.

(1)

Given that $P(A \mid B) = 0.6$

(c) find $P(A \cap B)$.

(3)

(d) Determine whether or not A and B are independent.

(2)

Leave blank

Question 8 continued

Q8

(Total 7 marks)

TOTAL FOR PAPER: 75 MARKS

END