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.

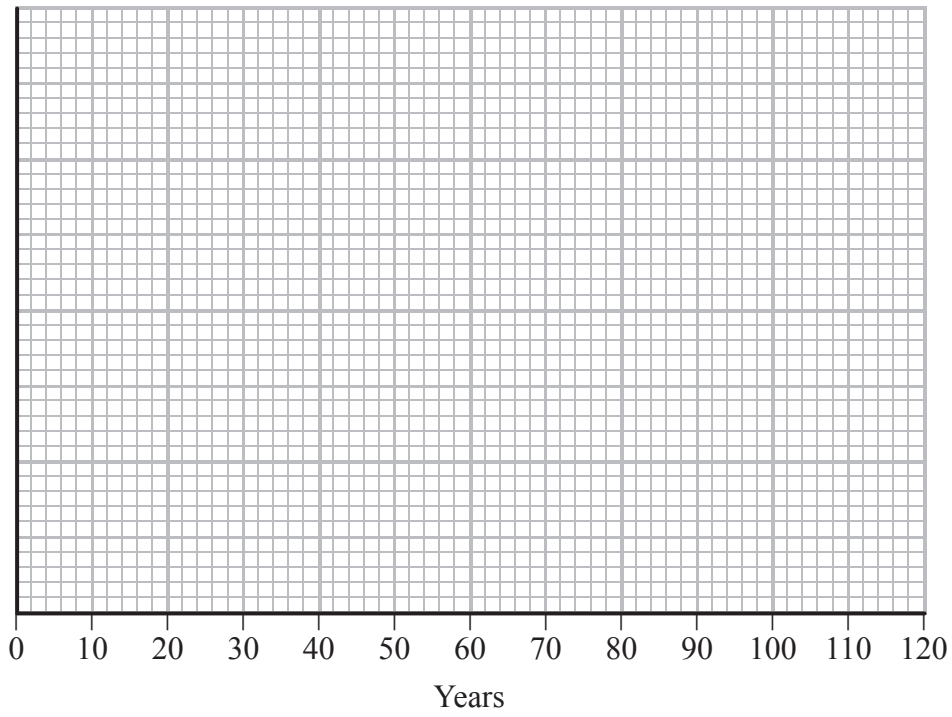
(4)

(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

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

(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

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

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

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)