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

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

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

(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)

a) $\frac{1}{4}(35) = 8.75 \quad \therefore Q_1 = x_9 = 44$

$\frac{3}{4}(35) = 26.25 \quad \therefore Q_3 = x_{27} = 76$

b) lower limit $\Rightarrow 1.5(55-31) = 36 \quad Q_1 - 36 = -5$ LL

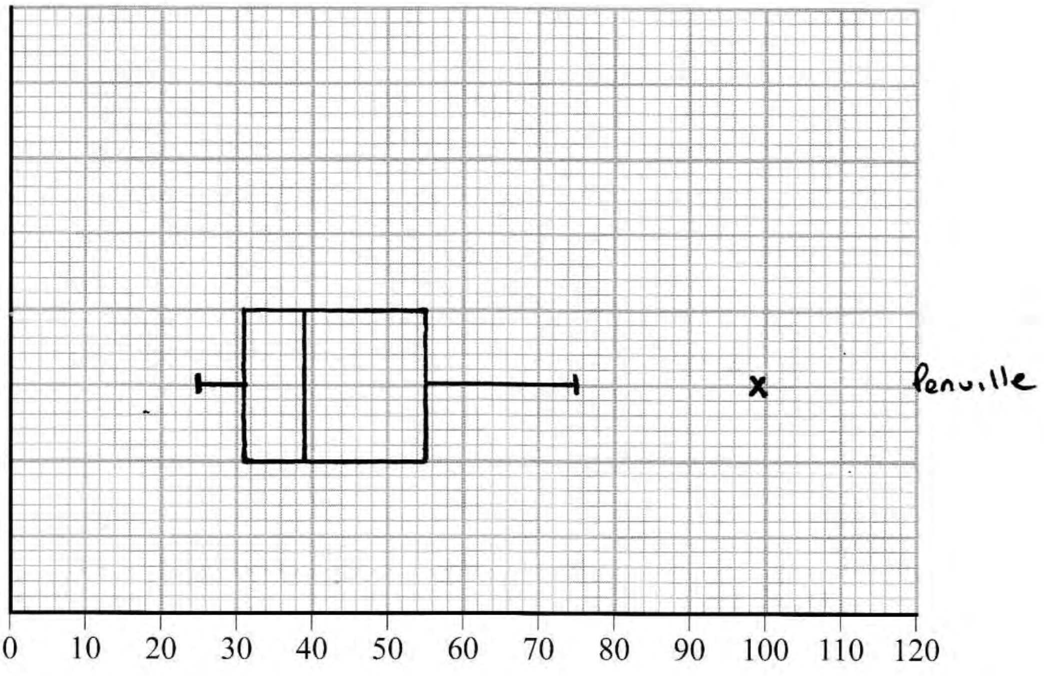
$Q_3 + 36 = 91$ UL

$\therefore$ there are no lower outliers, 99 is an outlier.

c) Greenslax $\underset{(20)}{Q_2 - Q_1} > \underset{(12)}{Q_3 - Q_2}$ negatively skewed

Penville $\underset{(8)}{Q_2 - Q_1} < \underset{(16)}{Q_3 - Q_2}$ positively skewed.

b)



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)

code $x \rightarrow \times 1.4 \rightarrow -20 \rightarrow y$

$\div 1.4 \leftarrow +20 \leftarrow$

decode.

mean $x = (60.8 + 20) \div 1.4 = 57.7$

$sd_x = 6.6 \div 1.4 = 4.71$

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)**

a) $r = \dfrac{S_{vm}}{\sqrt{S_{vv} \times S_{mm}}} = \dfrac{31512.5}{\sqrt{42587.5 \times 25187.5}} = 0.962$

2

b) r is very close to 1, suggesting a linear regression model would be appropriate as there is evidence to suggest v and m are positively correlated.

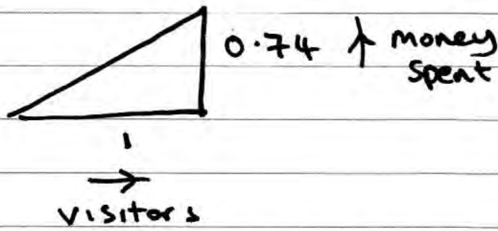
c) $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{S_{vm}}{S_{vv}} = \dfrac{31512.5}{42587.5} = \dfrac{2521}{3407}$ $\quad y = a + bx$, $\quad m = a + bv$

$a = \bar{y} - b\bar{x} = \bar{m} - b\bar{v} = \dfrac{10610}{8} - \dfrac{2521}{3407} \times \dfrac{19390}{8}$

c) $\therefore \; a = -467.197 \qquad b = 0.740$

d) $\Rightarrow \; m = -467.197 + 0.740v$

e)



every extra 1000 visitors results in an extra £740000 spent.

f) $v = 2500 \qquad m = -467.197 + 0.74 \times 2500$

$m = 1382.803$

$m = £1\,382\,803\,000$

g) reliable as 2500000 is within the data set of v and r was close to 1.

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)**

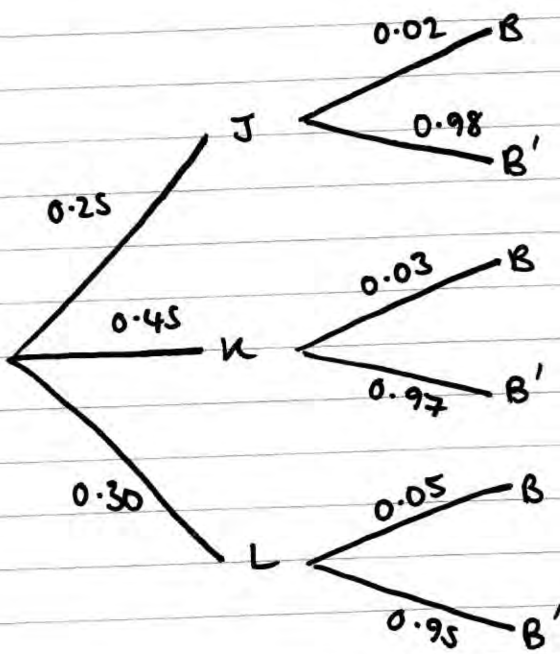
A biscuit is selected at random.

(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)**

a)



b) 0.25×0.98

$= 0.245$

2

c) $0.25 \times 0.02 + 0.45 \times 0.03 + 0.30 \times 0.05$

$= 0.0335$

2

d) $P(K' \mid B) = \dfrac{P(J \cap B) + P(L \cap B)}{P(B)} = \dfrac{0.25 \times 0.02 + 0.30 \times 0.05}{0.0335}$

$= \dfrac{40}{67} = 0.597$

2

5. The discrete random variable X has the probability function

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

where k is a constant.

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

a)

x	2	4	6	8
p	$2k$	$4k$	$6k$	$6k$

$\Sigma p = 1 \therefore 18k = 1 \therefore k = \frac{1}{18}$

b) $F(5) = P(x \leq 5) = 6k = \frac{6}{18} = \frac{1}{3}$

c)

x	2	4	6	8
p	$\frac{2}{18}$	$\frac{4}{18}$	$\frac{6}{18}$	$\frac{6}{18}$

$E(X) = \frac{4}{18} + \frac{16}{18} + \frac{36}{18} + \frac{48}{18} = \frac{104}{18} = \frac{52}{9}$

d) $E(X^2) = \frac{8}{18} + \frac{64}{18} + \frac{216}{18} + \frac{384}{18} = \frac{672}{18} = \frac{112}{3}$

e) $V(X) = E(X^2) - E(X)^2 = 3.9506\ldots$

$V(3-4X) = (-4)^2 V(X) = 63.2$

6. The times, in seconds, spent in a queue at a supermarket by 85 randomly 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

0
2
12
29 → 42.5th
54
79
85

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)**

a) 30 60 $\boxed{f = 10 \quad A = 1.5}$ 1cm

cw = 30

width = 1.5cm $\} \div 20$

$f \rightarrow \times \frac{1.5}{10} \rightarrow A$

70 80 $\boxed{f = 25}$

cw = 10

width = 0.5cm

$25 \times \frac{1.5}{10} = 3.75\text{cm}^2$

$\therefore h = \frac{3.75}{0.5} = 7.5\text{cm}$

b) $\frac{1}{2}n = 42.5$

70 Q_2 80

29 42.5 54

$\frac{Q_2 - 70}{10} = \frac{13.5}{25}$ ×, + $\therefore Q_2 = 75.4$

c) i) $\bar{x} = \frac{\Sigma fx}{n} = \frac{6460}{85} = 76$

ii) $Sxx = \Sigma fx^2 - (\Sigma fx)^2 \div n = 529400 - (6460)^2 \div 85$

$Sxx = 38440$

$sd = \sqrt{\frac{Sxx}{n}} = 21.3$

d) $\text{Skew} = \frac{3(76 - 75.4)}{21.3} = 0.0845$

very little positive skew, almost symmetrical

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.

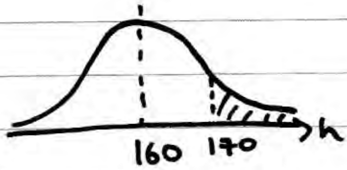
(4)

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

(c) Find the value of h.

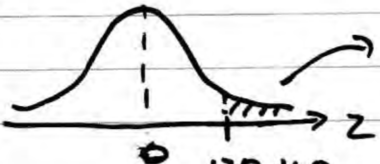
(5)

a)



$h \sim N(160, 8^2)$

$P(h > 170)$

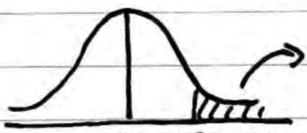


$\frac{170-160}{8} = 1.25$

$1 - \Phi(1.25) = 1 - 0.8944$
$= 0.1056$

b) $P(h > 180 \mid h > 170) = \frac{P(h > 180)}{P(h > 170)}$

$P(h > 180)$

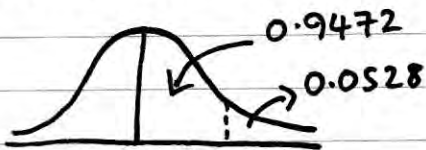


$\frac{180-160}{8} = 2.5$

$1 - \Phi(2.5) = 0.0062$ $\therefore \frac{0.0062}{0.1056} = 0.0587$

c) $P(H > h \mid H > 170) = 50\%$ $\Rightarrow \frac{P(H > h)}{P(H > 170)} = \frac{1}{2}$

$\therefore P(H > h) = \frac{1}{2} \times 0.1056 = 0.0528$



$\frac{h-160}{8}$

$\Phi\left(\frac{h-160}{8}\right) = 0.9472$

$\therefore \frac{h-160}{8} \simeq 1.62$

$h = 172.96$

8. For the events A and B,

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

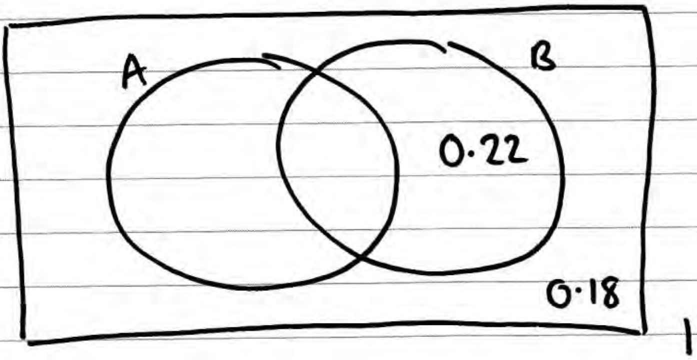
(a) Find $P(A)$.
(1)

(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)



a) $P(A) = 1 - 0{\cdot}18 - 0{\cdot}22 = \underline{0{\cdot}60}$

b) $P(A \cup B) = 0{\cdot}82 \qquad (1 - 0{\cdot}18)$

c) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = 0{\cdot}6 \qquad \therefore P(A \cap B) = 0{\cdot}6P(B)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0{\cdot}6 + P(B) - 0{\cdot}82$

$\Rightarrow 0{\cdot}6P(B) = 0{\cdot}6 + P(B) - 0{\cdot}82 \Rightarrow 0{\cdot}4P(B) = 0{\cdot}22$

$P(B) = 0{\cdot}55$

$\therefore P(A \cap B) = 0{\cdot}6 \times 0{\cdot}55 = \underline{\underline{0{\cdot}33}}$

d) $P(A \cap B) = 0{\cdot}33 \qquad P(A) \times P(B) = 0{\cdot}6 \times 0{\cdot}55 = 0{\cdot}33$

$\therefore P(A \cap B) = P(A) \times P(B) \quad \therefore$ A and B are independent.