2. Two youth clubs, *Eastyou* and *Westyou*, decided to raise money for charity by running a 5 km race. All the members of the youth clubs took part and the time, in minutes, taken for each member to run the 5 km was recorded.

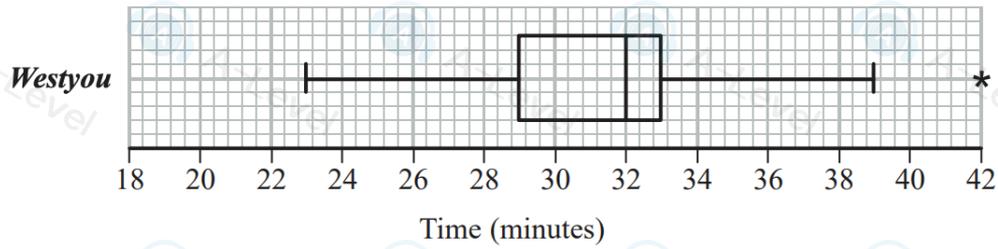The times for the *Westyou* members are summarised in Figure 1.

Westyou

Figure 1

(a) Write down the time that is exceeded by 75% of *Westyou* members.

(1)

The times for the *Eastyou* members are summarised by the stem and leaf diagram below.

Stem	Leaf													
2	0	2	3	4										(4)
2	5	6	8	8	8	9	9							(7)
3	0	0	0	0	0	1	1	1	2	2	2	2	3 4	(14)
3	5	5	5	7	9									(5)

Key: 2|0 means 20 minutes

(b) Find the value of the median and interquartile range for the *Eastyou* members.
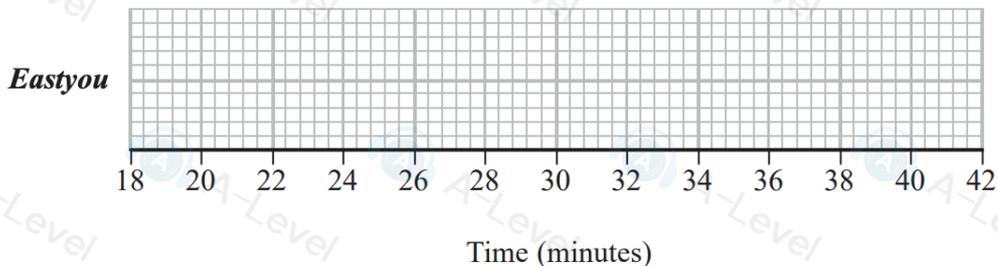
(3)

An outlier is a value that falls either

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

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

(c) On the grid on page 7, draw a box plot to represent the times of the *Eastyou* members.

(4)

(d) State the skewness of each distribution. Give reasons for your answers.
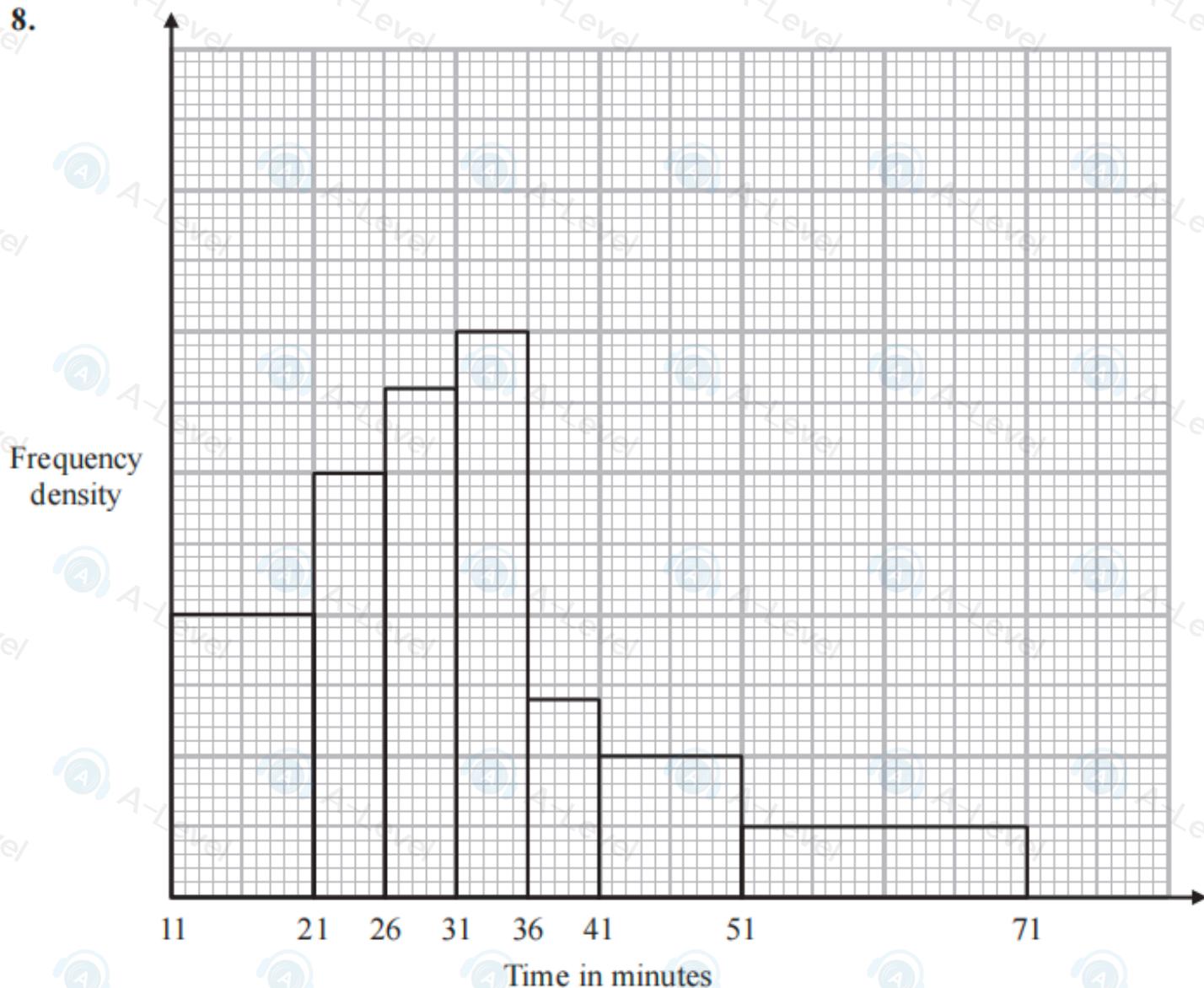
(3)

Eastyou

Time (minutes)

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)

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)

1. A researcher is investigating the growth of two types of tree, Birch and Maple. The height, to the nearest cm, a seedling grows in one year is recorded for 35 Birch trees and 32 Maple trees. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Birch		Maple	Totals
(2)	9 8	2	5 7 7 8 9	(5)
(8)	9 9 9 6 5 3 1 1	3	0 2 6 6 8 9 9	(7)
(9)	9 8 8 7 6 3 1 1 1	4	1 1 1 *k* 7 8	(6)
(9)	7 7 7 5 4 3 2 1 0	5	0 1 2 3 4 4 4	(7)
(3)	7 6 5	6	3 4 6	(3)
(3)	6 5 4	7	0 7	(2)
(1)	5	8	0 0	(2)

    Key: 5 | 6 | 3 means 65 cm for a Birch tree and 63 cm for a Maple tree

The median height that these **Maple** trees grow in one year is 45 cm.

(a) Find the value of k, used in the stem and leaf diagram.

(1)

(b) Find the lower quartile and the upper quartile of the height grown in one year for these **Birch** trees.

(2)

The researcher defines an outlier as an observation that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1) \text{ or less than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

(c) Show that there is only one outlier amongst the Birch trees.

(2)

The grid on page 3 shows a box plot for the heights that the Maple trees grow in one year.

(d) On the same grid draw a box plot for the heights that the Birch trees grow in one year.

(4)

(e) Comment on any difference in the distributions of the growth of these Birch trees and the growth of these Maple trees.
State the values of any statistics you have used to support your comment.

(1)

The researcher realises he has missed out 4 pieces of data for the **Maple** trees. The heights each seedling grows in one year, to the nearest cm, in ascending order, for these 4 Maple trees are $27\,cm$, $a\,cm$, $48\,cm$, $2a\,cm$.

Given that there is no change to the box plot for the **Maple** trees given on page 3

(f) find the range of possible values for a
Show your working clearly.

(3)

Height grown (cm)
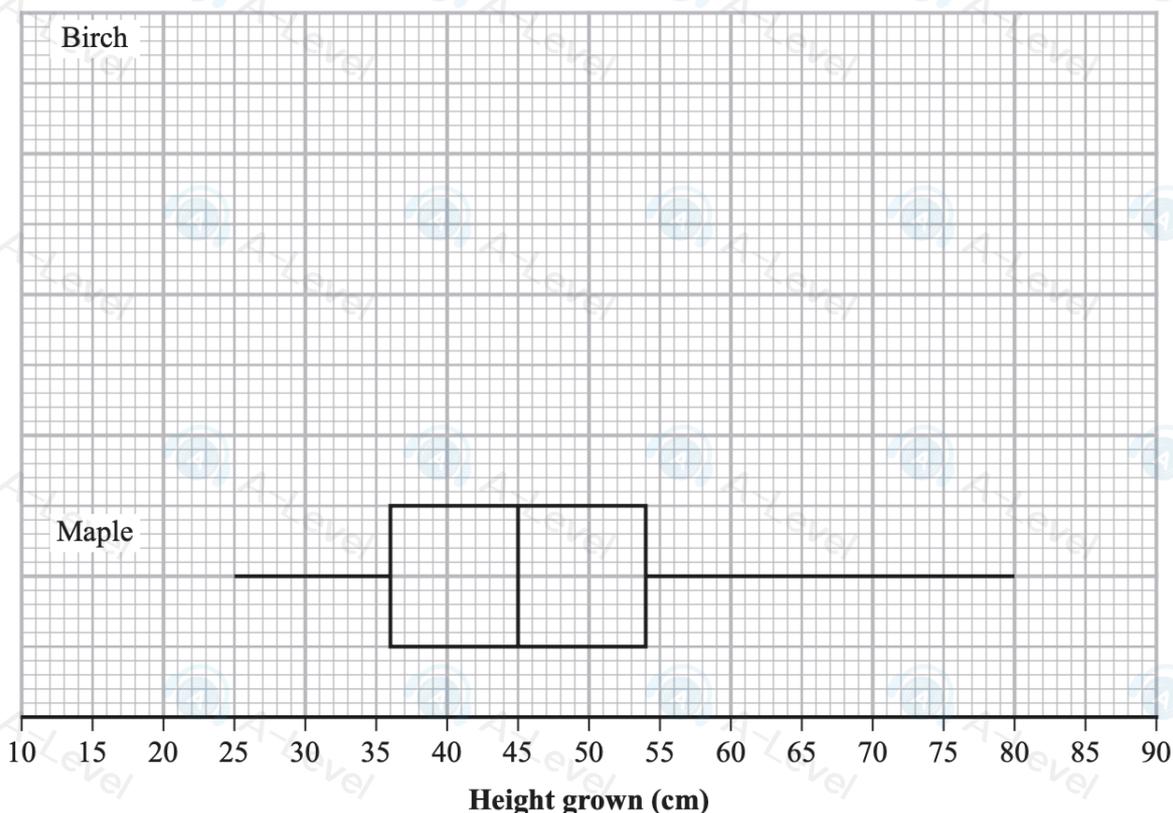
Turn over for spare grid if you need to redraw your answer for part (d)

3. The lengths, x mm, of 50 pebbles are summarised in the table below.

Length	Frequency
$20 \leqslant x < 30$	2
$30 \leqslant x < 32$	16
$32 \leqslant x < 36$	20
$36 \leqslant x < 40$	8
$40 \leqslant x < 45$	3
$45 \leqslant x < 50$	1

A histogram is drawn to represent these data.

The bar representing the class $32 \leqslant x < 36$ is 2.5 cm wide and 7.5 cm tall.

(a) Calculate the width and the height of the bar representing the class $30 \leqslant x < 32$

(3)

(b) Using linear interpolation, estimate the median of x

(2)

The weight, w grams, of each of the 50 pebbles is coded using $10y = w - 20$
These coded data are summarised by

$$\sum y = 104 \qquad \sum y^2 = 233.54$$

(c) Show that the mean of w is 40.8

(2)

(d) Calculate the standard deviation of w

(4)

The weight of a pebble recorded as 40.8 grams is added to the sample.

(e) Without carrying out any further calculations, state, giving a reason, what effect this would have on the value of

(i) the mean of w

(ii) the standard deviation of w

(3)

2. The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

| Age		Key: $4\mid 3$ is an age of 43
4	3	(1)
5	4	(1)
6	2 3 5 6 8 8 8 9 9	(9)
7	1 1 3 4 4 6 6 6 8 8 9	(11)
8	0 0 2 7 8 8 9	(7)
9	3 7	(2)

(a) Find the median age of the residents.

(1)

(b) Find the interquartile range (IQR) of the ages of the residents.

(2)

An outlier is defined as a value that is either

more than $1.5 \times$ (IQR) below the lower quartile or

more than $1.5 \times$ (IQR) above the upper quartile.

(c) Determine any outliers in these data. Show clearly any calculations that you use.

(3)

(d) On the grid on page 5, draw a box plot to summarise these data.

(3)

30   40   50   60   70   80   90   100

Ages

Q2
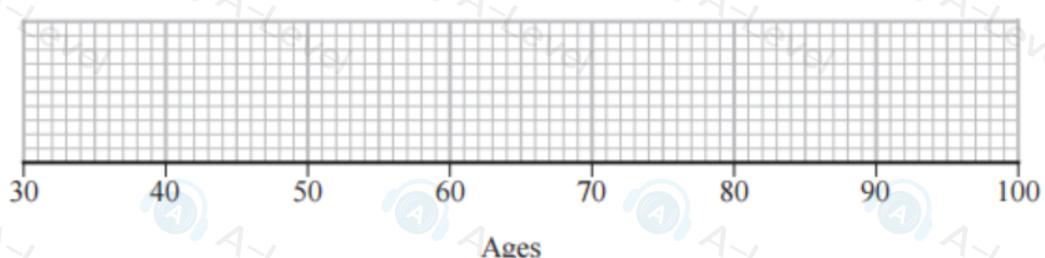
(Total 9 marks)

2. A sports teacher recorded the number of press-ups done by his students in two minutes. He recorded this information for a Year 7 class and for a Year 11 class.

The back-to-back stem and leaf diagram shows this information.

Totals	Year 7 class		Year 11 class	Totals
(6)	8 7 6 5 5 4	1		
(10)	9 7 7 6 5 4 4 4 2 2	2	0 5 6 9	(4)
(7)	8 7 5 4 3 3 0	3	3 4 5 8 8	(5)
(5)	9 9 7 2 2	4	0 5 6 7 9	(5)
(3)	8 4 0	5	0 3 5 5 6 6 7 7 7 9 9	(11)
		6	0 3 3 3 3 4 8	(7)

Key: 2|4|0 means 42 press-ups for a Year 7 student and 40 press-ups for a Year 11 student

(a) Find the median number of press-ups for each class.

(2)

For the Year 11 class, the lower quartile is 38 and the upper quartile is 59

(b) Find the lower quartile and the upper quartile for the Year 7 class.

(2)

(c) Use the medians and quartiles to describe the skewness of each of the two distributions.

(3)

(d) Give two reasons why the normal distribution should not be used to model the number of press-ups done by the Year 11 class.

(2)

6. The stem and leaf diagram gives the blood pressure, x mmHg, for a random sample of 19 female patients.

$$
\begin{array}{r|llllllllll}
10 & 1 & 2 & & & & & & & & (2)\\
11 & 2 & 7 & 7 & 8 & 8 & & & & & (5)\\
12 & 0 & 2 & 2 & 3 & 4 & 4 & 5 & 5 & 7 & (9)\\
13 & 1 & 2 & 9 & & & & & & & (3)
\end{array}
$$

Key: 10 | 1 means blood pressure of 101 mmHg

(a) Find the median and the quartiles for these data.

(3)

(b) Find the interquartile range $(Q_3 - Q_1)$

(1)

An outlier is a value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or less than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(c) Showing your working clearly, identify any outliers for these data.

(3)

(d) On the grid on page 21 draw a box and whisker plot to represent these data. Show any outliers clearly.

(4)

The above data can be summarised by

$$\sum x = 2299 \quad \text{and} \quad \sum x^2 = 279709$$

(e) Calculate the mean and the standard deviation for these data.

(3)

For a random sample taken from a normal distribution, a rule for determining outliers is:

an outlier is more than $2.7 \times$ standard deviation above or below the mean.

(f) Find the limits to determine outliers using this rule.

(2)

(g) State, giving a reason based on some of the above calculations, whether or not a normal distribution is a suitable model for these data.

(1)

90   100   110   120   130   140   150
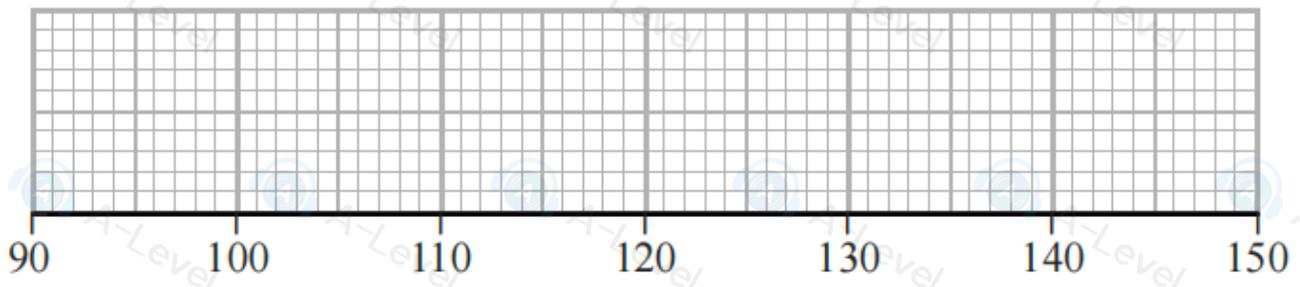
Turn over for a spare diagram if you need to redraw your plot.