Data

Leave blank

1. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(a) Calculate the width and height of the rectangle representing the group (16 – 25) hours.
(3)

(b) Use linear interpolation to estimate the median and interquartile range.
(5)

(c) Estimate the mean and standard deviation of the number of hours spent shopping.
(4)

(d) State, giving a reason, the skewness of these data.
(2)

(e) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data.
(2)

Leave blank

2. The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2|6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

(a) the median score,
(1)

(b) the interquartile range.
(3)

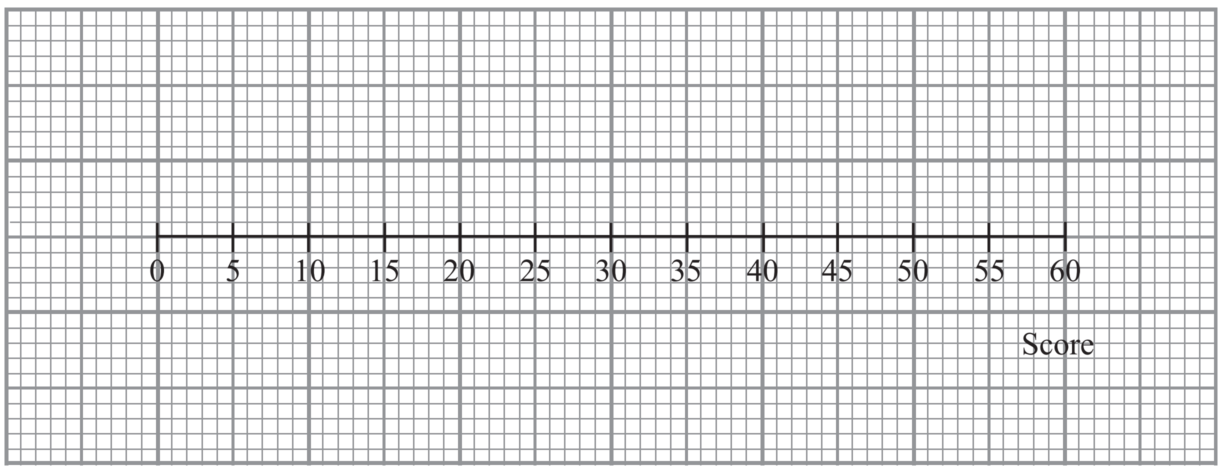
The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

(c) Explain why there is only one employee who will undergo retraining.
(2)

(d) On the graph paper on page 5, draw a box plot to illustrate the employees' scores.
(3)

Question 2 continued



Q2

(Total 9 marks)

Leave blank

3. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, x kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum \text{f}x = 4841$ and $\sum \text{f}x^2 = 15\,889.5$]

(a) Write down the missing midpoints in the table above. **(2)**

(b) Calculate an estimate of the mean birth weight. **(2)**

(c) Calculate an estimate of the standard deviation of the birth weight. **(3)**

(d) Use interpolation to estimate the median birth weight. **(2)**

(e) Describe the skewness of the distribution. Give a reason for your answer. **(2)**

Leave blank

4. The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

(a) the width,

(1)

(b) the height

(2)

of the bar representing this class.

Leave blank

5. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

Foot length, l, (cm)	**Number of children**
$10 \leqslant l < 12$	5
$12 \leqslant l < 17$	53
$17 \leqslant l < 19$	29
$19 \leqslant l < 21$	15
$21 \leqslant l < 23$	11
$23 \leqslant l < 25$	7

(a) Use interpolation to estimate the median of this distribution.
(2)

(b) Calculate estimates for the mean and the standard deviation of these data.
(6)

One measure of skewness is given by

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

(c) Evaluate this coefficient and comment on the skewness of these data.
(3)

Greg suggests that a normal distribution is a suitable model for the foot lengths of ten-year-old children.

(d) Using the value found in part (c), comment on Greg's suggestion, giving a reason for your answer.
(2)

Leave blank

6. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1 – 10	11 – 20	21 – 25	26 – 30	31 – 40	41 – 59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the 21 – 25 hour and 31 – 40 hour groups.
(2)

A histogram was drawn to represent these data. The 11 – 20 group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the 26 – 30 group.
(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.
(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by these students.
(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(e) State, giving a reason, the skewness of these data.
(2)

Leave blank

7. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

(a) Find the median and quartiles for these data. **(3)**

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(b) Show that 110 is the only outlier. **(2)**

(c) Using the graph paper on page 15 draw a box plot for these data indicating clearly the position of the outlier. **(3)**

The value of 110 is omitted.

(d) Show that S_{yy} for the remaining 10 students is 2966.9 **(3)**

These 10 students were each asked how many text messages, x, they sent in the same week.

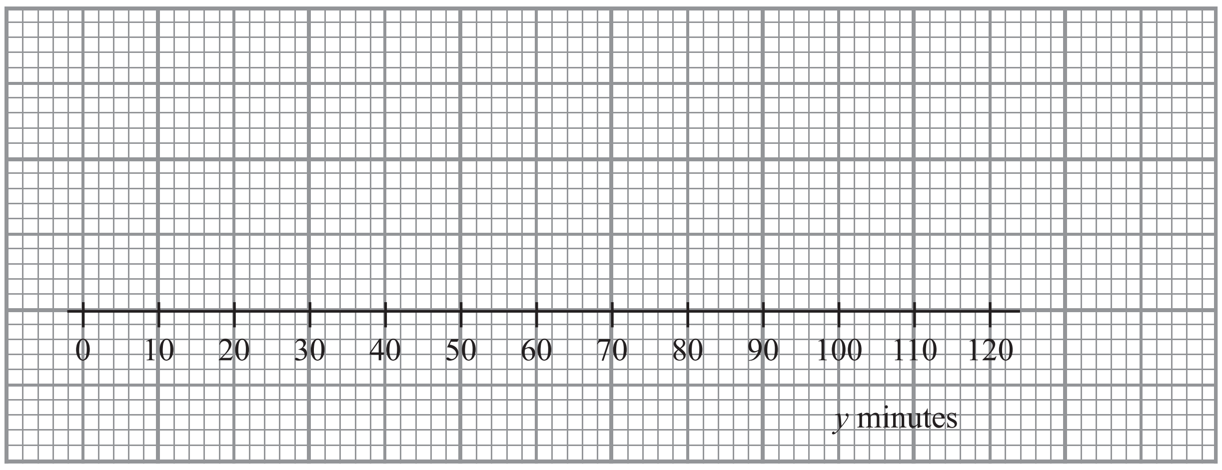
The values of S_{xx} and S_{xy} for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

(e) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students. **(2)**

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(f) Comment on this belief in the light of your calculation in part (e). **(1)**

Question 7 continued



Leave blank