41. Categorical, numerical and grouped data	• collect, classify and tabulate statistical data • read, interpret and draw simple inferences from tables and statistical diagrams • calculate the mean, median, mode and range for individual and discrete data and distinguish between the purposes for which they are used • calculate an estimate of the mean for grouped and continuous data • identify the modal class from a grouped frequency distribution	
42. Statistical diagrams	• construct and interpret bar charts, pie charts, pictograms, simple frequency distributions, frequency polygons, histograms with equal and unequal intervals and scatter diagrams • construct and use cumulative frequency diagrams • estimate and interpret the median, percentiles, quartiles and interquartile range for cumulative frequency diagrams • calculate with frequency density • understand what is meant by positive, negative and zero correlation with reference to a scatter diagram • draw a straight line of best fit by eye.	For unequal intervals on histograms, areas are proportional to frequencies and the vertical axis is labelled 'Frequency density'.

Videos for understanding

https://www.youtube.com/watch?v=h8EYEJ32oQ8

https://www.youtube.com/watch?v=JAyYsEfp_HA

https://www.youtube.com/watch?v=M-ijIZOcU0Y
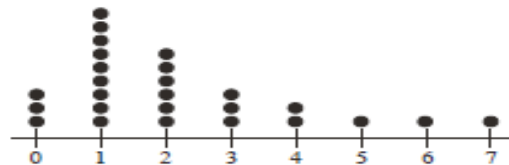
UNIT 3.2 | Data Analysis

Dot Diagram

1. A dot diagram consists of a horizontal number line and dots placed above the number line, representing the values in a set of data.

Example 1

The table shows the number of goals scored by a soccer team during the tournament season.

Number of goals	0	1	2	3	4	5	6	7
Number of matches	3	9	6	3	2	1	1	1

The data can be represented on a dot diagram.

Example 2

The marks scored by twenty students in a placement test are as follows:

| 42 | 42 | 49 | 31 | 34 | 42 | 40 | 43 | 35 | 38 |
| 34 | 35 | 39 | 45 | 42 | 42 | 35 | 45 | 40 | 41 |

(a) Illustrate the information on a dot diagram.
(b) Write down
  (i)   the lowest score,
  (ii) the highest score,
  (iii) the modal score.

Solution

(a)
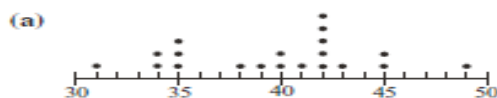
(b) (i)   Lowest score = 31
    (ii) Highest score = 49
    (iii) Modal score = 42

- -

2. An advantage of a dot diagram is that it is an easy way to display small sets of data which do not contain many distinct values.

Stem-and-Leaf Diagram

3. In a stem-and-leaf diagram, the stems must be arranged in numerical order and the leaves must be arranged in ascending order.

4. An advantage of a stem-and-leaf diagram is that the individual data values are retained.

e.g. The ages of 15 shoppers are as follows:

32	34	13	29	38
36	14	28	37	13
42	24	20	11	25

The data can be represented on a stem-and-leaf diagram.

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

Key: 1 | 3 means 13 years old

The tens are represented as stems and the ones are represented as leaves.
The values of the stems and the leaves are arranged in ascending order.

Stem-and-Leaf Diagram with Split Stems

5. If a stem-and-leaf diagram has more leaves on some stems, we can break each stem into two halves.

e.g. The stem-and-leaf diagram represents the number of customers in a store.

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

Key: 4 | 0 means 40 customers

The information can be shown as a stem-and-leaf diagram with split stems.

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

Key: 4 | 0 means 40 customers

Back-to-Back Stem-and-Leaf Diagram

6. If we have two sets of data, we can use a back-to-back stem-and-leaf diagram with a common stem to represent the data.

e.g. The scores for a quiz of two classes are shown in the table.

Class A	55	98	67	84	85	92	75	78	89	64
	72	60	86	91	97	58	63	86	92	74
Class B	56	67	92	50	64	83	84	67	90	83
	68	75	81	93	99	76	87	80	64	58

A back-to-back stem-and-leaf diagram can be constructed based on the given data.

Leaves for Class B					Stem	Leaves for Class A				
		8	6	0	5	5	8			
	8	7	7	4	4	6	0	3	4	7	
				6	5	7	2	4	5	8	
7	4	3	3	1	0	8	4	5	6	6	9
		9	3	2	0	9	1	2	2	7	8

Key: 58 means 58 marks

Note that the leaves for Class B are arranged in ascending order from the right to the left.

Measures of Central Tendency

7. The three common measures of central tendency are the mean, median and mode.

Mean

8. The mean of a set of n numbers $x_1, x_2, x_3, \ldots, x_n$ is denoted by $\bar{x}$.

9. For ungrouped data,
$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x}{n}.$$

10. For grouped data,
$$\bar{x} = \frac{\sum fx}{\sum f}$$
where f is the frequency of data in each class interval and x is the mid-value of the interval.

Median

11. The median is the value of the middle term of a set of numbers arranged in ascending order.

    Given a set of n terms, if n is odd, the median is the $\left(\frac{n+1}{2}\right)^{th}$ term;

    if n is even, the median is the average of the two middle terms.

    e.g. Given the set of data: 5, 6, 7, 8, 9.
    There is an odd number of data.
    Hence, median is 7.

    e.g. Given a set of data 5, 6, 7, 8.
    There is an even number of data.
    Hence, median is 6.5.

Example 3

The table records the number of mistakes made by 60 students during an exam.

Number of students	24	x	13	y	5
Number of mistakes	5	6	7	8	9

(a) Show that $x + y = 18$.
(b) Find an equation of the mean, given that the mean number of mistakes made is 6.3. Hence, find the values of x and y.
(c) State the median number of mistakes made.

Solution

(a) Since there are 60 students in total,
$$24 + x + 13 + y + 5 = 60$$
$$x + y + 42 = 60$$
$$x + y = 18$$

(b) Since the mean number of mistakes made is 6.3,
$$\text{Mean} = \frac{\text{Total number of mistakes made by 60 students}}{\text{Number of students}}$$
$$6.3 = \frac{24(5) + 6x + 13(7) + 8y + 5(9)}{60}$$
$$6.3(60) = 120 + 6x + 91 + 8y + 45$$
$$378 = 256 + 6x + 8y$$
$$6x + 8y = 122$$
$$3x + 4y = 61$$

To find the values of x and y, solve the pair of simultaneous equations obtained above.
$$x + y = 18 \quad\text{---- (1)}$$
$$3x + 4y = 61 \quad\text{---- (2)}$$
$3 \times$ (1):
$$3x + 3y = 54 \quad\text{---- (3)}$$
(2) − (3): $y = 7$
When $y = 7$, $x = 11$.

(c) Since there are 60 students, the 30th and 31st students are in the middle. The 30th and 31st students make 6 mistakes each. Therefore, the median number of mistakes made is 6.

Mode

12. The mode of a set of numbers is the number with the highest frequency.

13. If a set of data has two values which occur the most number of times, we say that the distribution is bimodal.

    e.g. Given a set of data: 5, 6, 6, 6, 7, 7, 8, 8, 9.
    6 occurs the most number of times.
    Hence, the mode is 6.

Standard Deviation

14. The standard deviation, s, measures the spread of a set of data from its mean.

15. For ungrouped data,

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \quad \text{or} \quad s = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

16. For grouped data,

$$s = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} \quad \text{or} \quad s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Example 4

The following set of data shows the number of books borrowed by 20 children during their visit to the library.

$$0, 2, 4, 3, 1, 1, 2, 0, 3, 1$$
$$1, 2, 1, 1, 2, 3, 2, 2, 1, 2$$

Calculate the standard deviation.

Solution

Represent the set of data in the table below.

Number of books borrowed	0	1	2	3	4
Number of children	2	7	7	3	1

Standard deviation

$$= \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$= \sqrt{\frac{0^2(2) + 1^2(7) + 2^2(7) + 3^2(3) + 4^2(1)}{20} - \left(\frac{0(2) + 1(7) + 2(7) + 3(3) + 4(1)}{20}\right)^2}$$

$$= \sqrt{\frac{78}{20} - \left(\frac{34}{20}\right)^2}$$

$$= 1.00 \text{ (to 3 s.f.)}$$

Cumulative Frequency Curve

17. The following figure shows a cumulative frequency curve.
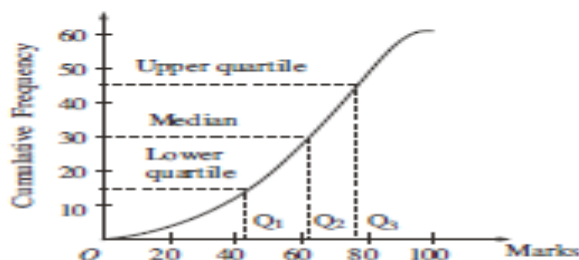
18. Q_1 is called the lower quartile or the 25th percentile.

19. Q_2 is called the median or the 50th percentile.

20. Q_3 is called the upper quartile or the 75th percentile.

21. $Q_3 - Q_1$ is called the interquartile range.

1 A group of 80 students took a physics test.
   This table shows the distribution of their marks.   SP18/01/8

Mark (m)	$0 < m \leqslant 10$	$10 < m \leqslant 20$	$20 < m \leqslant 30$	$30 < m \leqslant 40$	$40 < m \leqslant 50$	$50 < m \leqslant 60$
Frequency	4	12	14	22	18	10

(a) Complete the cumulative frequency table.

Mark (m)	$m \leqslant 10$	$m \leqslant 20$	$m \leqslant 30$	$m \leqslant 40$	$m \leqslant 50$	$m \leqslant 60$
Cumulative frequency						

[1]

(b) Draw a cumulative frequency curve for this information.

Mark (m)
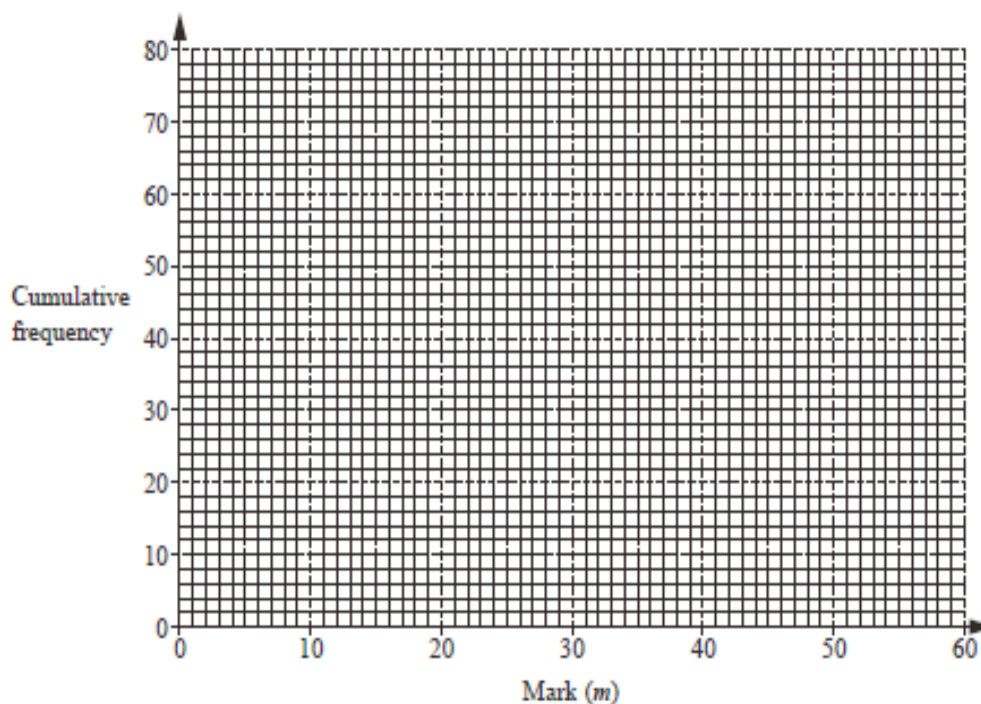
[2]

(c) The pass mark for the test is 45.
     Use your cumulative frequency curve to estimate the number of students who passed.

Answer .... [2]

2   The table shows the test marks in mathematics and physics for 10 students.   SP18/02/2

Mathematics mark	30	50	35	25	5	39	48	40	10	15
Physics mark	26	39	35	28	9	37	45	33	16	12

(a) (i) On the grid, complete the scatter diagram to show the mathematics and physics marks for the
         10 students. The first four points have been plotted for you.

Mathematics mark

[2]
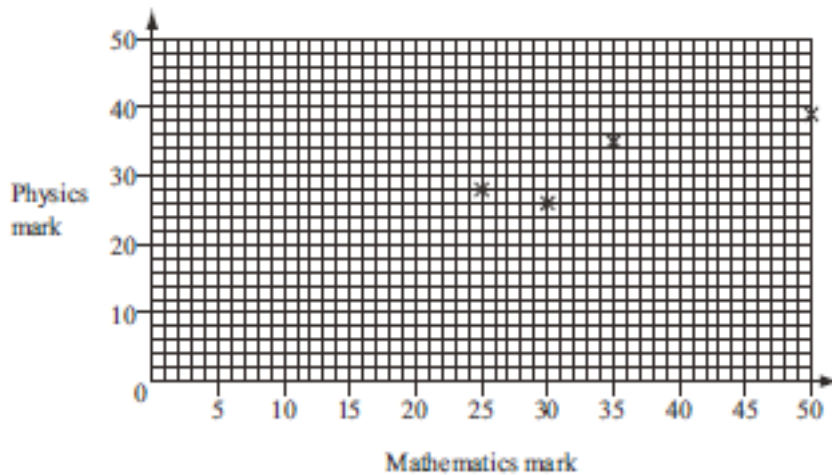
(ii) What type of correlation does your scatter diagram show?

Answer .... [1]

(iii) Draw a line of best fit on the grid.   [1]

(iv) Ann missed the physics test but scored 22 marks in the mathematics test.
      Use your line of best fit to estimate a possible physics mark for Ann.

Answer .... [1]

(b) Find the range of marks in the mathematics test.

Answer .... [1]

(c) Show that the mean physics mark for the 10 students is 28.

[1]

(d) Two new students do the physics test. They both score the **same** mark.
The mean physics mark for the 12 students is 31.

Calculate the physics mark for the new students.

Answer .... [3]

(e) Two of the original 10 students are chosen at random.

Work out the probability that **both** students scored less than 20 in the mathematics test.

Answer .... [2]

3   The students in a school each choose a piece of fruit to eat with their lunch.
They can choose from either an apple, a banana or an orange.   M/J19/11/9

(a)   On Monday, Klaudia records the fruit chosen by 30 of her classmates.
Their choices are given below.

Orange	Apple	Orange	Apple	Apple	Orange
Banana	Orange	Apple	Banana	Orange	Orange
Apple	Banana	Orange	Apple	Orange	Banana
Orange	Apple	Banana	Apple	Banana	Orange
Apple	Orange	Orange	Banana	Apple	Banana

(i)   Complete the frequency table for the data.

Fruit	Apple	Banana	Orange
Frequency			

[1]

(ii)   On the grid below, draw a bar chart to show the data.
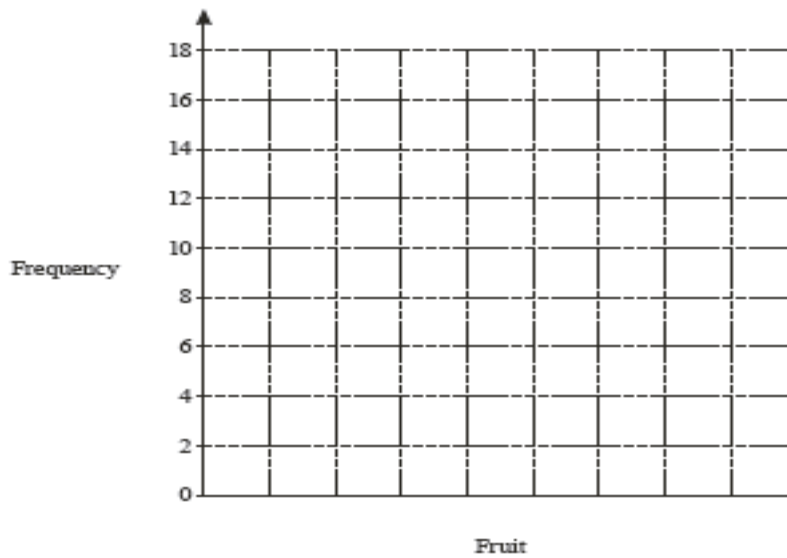
[2]

(b)   On Tuesday, Ahmed records the fruit chosen by a random sample of 30 students in the lunch queue.
His results are shown in the table below.

Fruit	Apple	Banana	Orange
Frequency	8	15	7

(i)   Use Ahmed's results to estimate the probability that a student selected at random chooses an **orange**.

.... [1]

(ii)   There is a total of 180 students in the school.

Use Ahmed's results to estimate the number of students in the school who chose an **apple** on Tuesday.

.... [1]

4   The table summarises the distances, d m, that 80 women threw the javelin.

M/J19/21/8

Distance (d m)	Frequency
$20 < d \leqslant 25$	6
$25 < d \leqslant 30$	16
$30 < d \leqslant 35$	25
$35 < d \leqslant 40$	18
$40 < d \leqslant 45$	13
$45 < d \leqslant 50$	2

(a) One of these women is chosen at random.

Find the probability that she threw the javelin 30 metres or less.

.... [1]

(b) Calculate an estimate of the mean distance the javelin was thrown.

.... m [3]

(c) Draw the cumulative frequency diagram for this data on the grid on the next page.

Cumulative frequency

Distance (m)

[3]
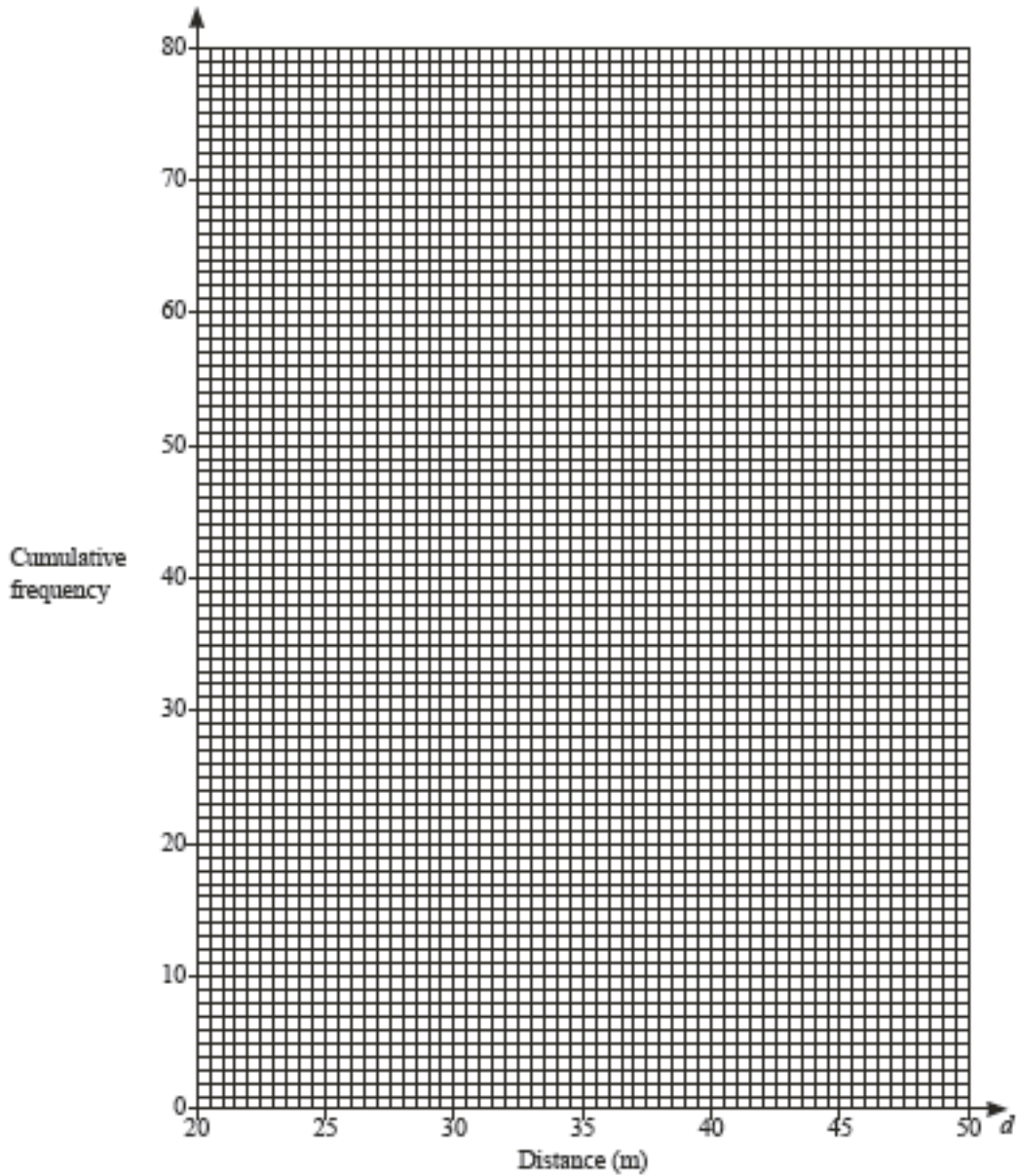
(d) Use your graph to find an estimate for

(i) the median,

.... m [1]

(ii) the interquartile range.

.... m [2]

(e) Women who threw the javelin more than 43 m qualified for a regional competition.

Use your graph to estimate the number of women who qualified for this competition.

.... [2]

5 The table shows the average monthly temperatures (°C) in Tokyo and in Sydney one

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Sydney (°C)	23	23	22	19	16	14	13	14	16	18	20	21
Temperature in Tokyo (°C)	5	6	9	14	18	21	25	26	23	18	12	8

(a) Complete the scatter diagram.
The first eight points have been plotted for you.   M/J19/22/2   [2]
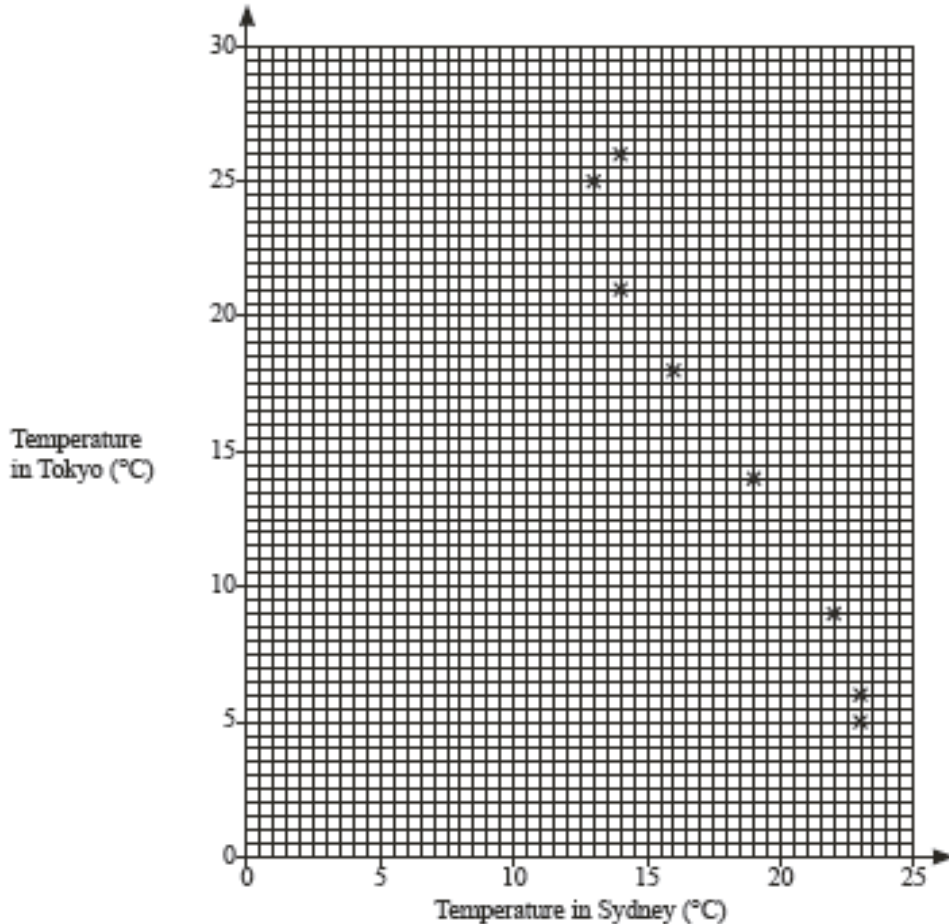
(b) What type of correlation is shown by the scatter diagram?

.... [1]

(c) Draw a line of best fit.   [1]

(d) The following year, the average temperature in Sydney during May was 15 °C.

By using your line of best fit, estimate the average temperature in Tokyo that May.

.... °C [1]

6 The mass, in grams, of each of 75 oranges is measured. The results are summarised in the table.   M/J19/22/8

Mass (m g)	Frequency
$160 < m \leqslant 170$	4
$170 < m \leqslant 180$	10
$180 < m \leqslant 190$	15
$190 < m \leqslant 200$	23
$200 < m \leqslant 210$	18
$210 < m \leqslant 220$	5

(a) One of these oranges is chosen at random.

Find the probability that it has a mass greater than 200 g.

.... [1]

(b) Calculate an estimate of the mean mass.

.... g [3]

(c) On the grid on the next page, draw a cumulative frequency diagram for the mass of the oranges.

[3]
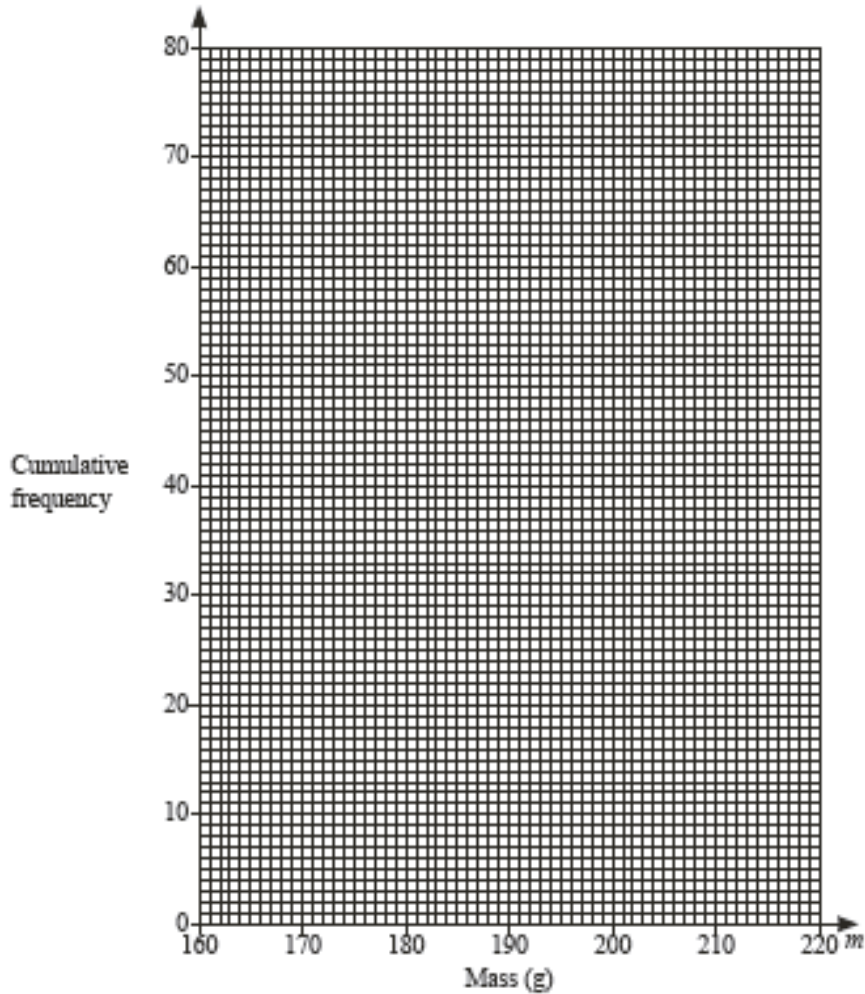
(d) Use your graph to find an estimate for the median.

.... g [1]

(e) Oranges with a mass of more than T g are labelled 'Class A'.
20% of the oranges are labelled 'Class A'.

Use your graph to find T.

$T = $ .... [2]

(f) Oranges with a mass of 178 g or less are used to make orange juice.

Use your graph to find an estimate for the number of these oranges that are used to make orange juice.

.... [1]

7 A five-sided spinner is numbered 1, 2, 3, 4 and 5.   O/N18/11/15
   Ashraf spun the spinner 200 times.
   The results are shown in the table.

Number spinner lands on	1	2	3	4	5
Frequency	30	25	50	55	40

(a) Calculate the relative frequency that the spinner lands on 3.

                                        Answer .... [1]

(b) Meriam spins the spinner 20 times.

     How many times would you expect the spinner to land on 3?

                                        Answer .... [1]

(c) Ashraf claims: "My results show that the spinner is fair".
     Is his claim correct?
     Give a reason for your answer.

     .... because ....

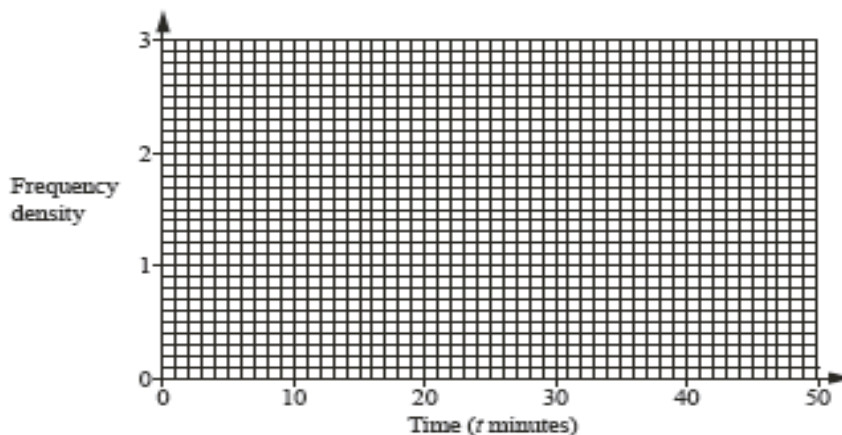   .... [1]

7 The lengths of the times of telephone calls made by Ellie during one week are summarised in the table.

Time (t minutes)	$0 < t \leqslant 10$	$10 < t \leqslant 20$	$20 < t \leqslant 25$	$25 < t \leqslant 30$	$30 < t \leqslant 50$
Frequency	10	15	10	12	16

O/N18/11/13

On the grid, draw a histogram to illustrate the distribution of these times.

[3]

8   Each week, Henri records the number of units of gas used in his house and the mean temperature outside.
Ten of his results are shown in the table.

Mean temperature (°C)	12.2	13.8	0.6	2.2	2.8	4.4	5.6	6.8	9.0	10.6
Number of units	140	100	740	570	660	600	500	560	410	320

(a) On the grid, complete the scatter diagram.
The first five results have been plotted for you.   O/N18/12/15

[2]
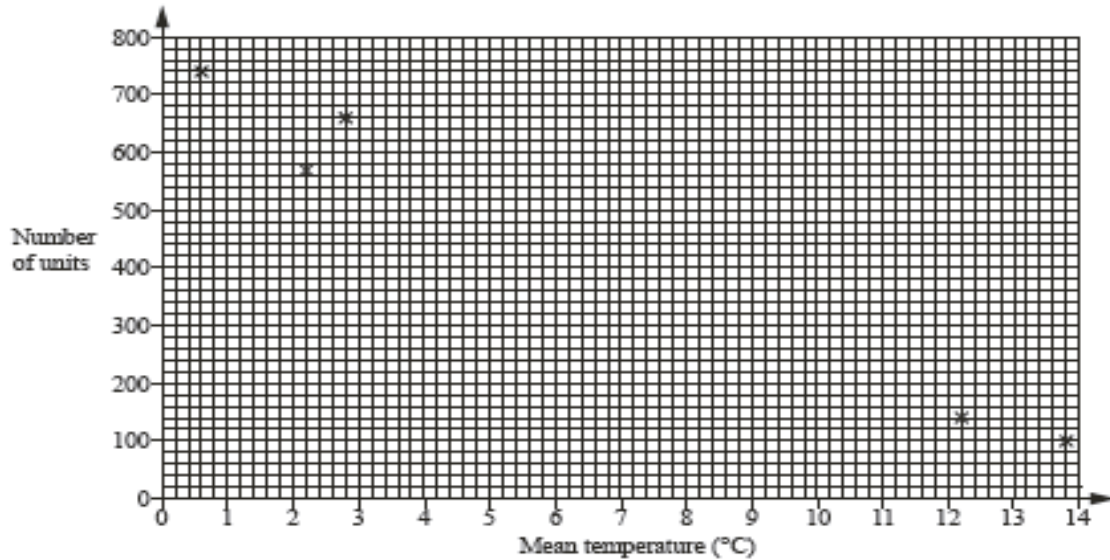
(b) What type of correlation does your scatter diagram show?

Answer .... [1]

(c) Draw a line of best fit on the grid.   [1]

(d) Use your line of best fit to estimate the number of units used for one week when the mean temperature outside is 7.6 °C.
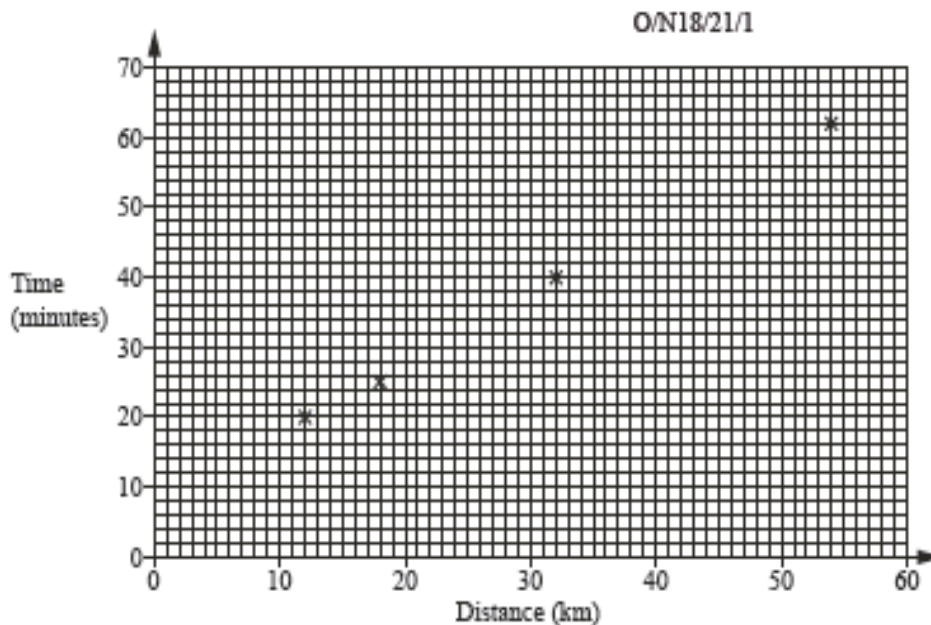
Answer .... [1]

9   (a)   The table shows the distances 10 people drive to work and the times they take.

Distance (km)	12	32	18	54	26	36	29	15	20	42
Time (minutes)	20	40	25	62	35	34	32	18	32	50

(i)   On the grid, complete the scatter diagram to show this information.
The first four points have been plotted for you.

O/N18/21/1

[2]

(ii)   What type of correlation does the scatter diagram show?

Answer .... [1]

(iii)   One of these 10 people is selected at random.

Find the probability that this person drove less than 30 km to work.
Give your answer as a fraction in its simplest form.

Answer .... [2]

(iv)   Ateeq drives 48 km to work.

By drawing a line of best fit on the scatter diagram, estimate the time Ateeq takes to travel to work.

(b) The table summarises the times taken by the 120 employees in a company to travel to work.

Time (t minutes)	$0 < t \leqslant 20$	$20 < t \leqslant 40$	$40 < t \leqslant 60$	$60 < t \leqslant 80$	$80 < t \leqslant 100$
Frequency	29	38	26	21	6

(i) Write down the modal class.

Answer .... [1]

(ii) Calculate an estimate of the mean time.

Answer .... minutes [3]

(iii) Work out the percentage of employees who took more than 1 hour to travel to work.

Answer .... % [2]

10   Lim grows tomatoes.   O/N18/22/2
      The masses, m grams, of 200 of her tomatoes are recorded.
      The cumulative frequency table shows the results.

Mass (m grams)	$m \leqslant 80$	$m \leqslant 100$	$m \leqslant 110$	$m \leqslant 120$	$m \leqslant 130$	$m \leqslant 140$	$m \leqslant 160$
Cumulative frequency	0	20	48	112	158	184	200

(a) On the grid, draw a cumulative frequency diagram to represent these results.
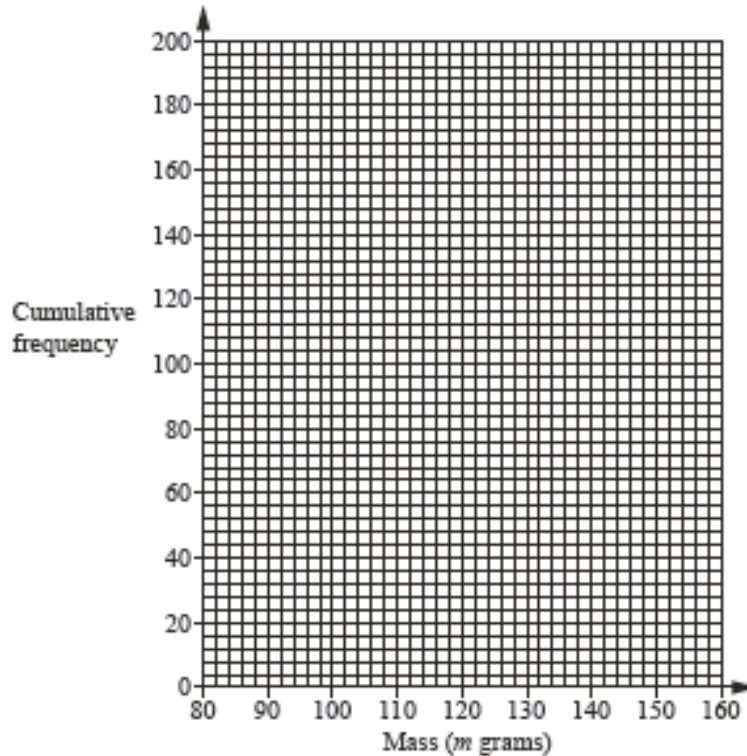
[2]

(b) Use your diagram to estimate

(i) the median,

Answer .... g [1]

(ii) the interquartile range.

Answer .... g [2]

(c) Ravi also grows tomatoes.

The masses of 200 of his tomatoes are also recorded.

For Ravi's tomatoes, the median mass was 124 g and the interquartile range of the masses was 12 g.

Make two comments comparing the masses of tomatoes grown by Lim with those grown by Ravi.

1 ....

....

2 ....

.... [2]

(d) (i) Complete the frequency table for the masses of tomatoes grown by Lim.

Mass (m grams)	Frequency
$80 < m \leqslant 100$	20
$100 < m \leqslant 110$	28
$110 < m \leqslant 120$	64
$120 < m \leqslant 130$	
$130 < m \leqslant 140$	
$140 < m \leqslant 160$	16

[1]

(ii) Write down the modal class.

Answer .... [1]

(iii) Calculate an estimate for the mean mass of these tomatoes.

Answer .... g [3]

11 The table below shows the masses of 10 mothers and their babies at birth.

Mass of mother (kg)	64	90	54	102	57	105	70	89	57	75
Mass of baby (kg)	4.1	4.5	3.6	4.5	3.9	5.5	3.9	4.3	3.2	4.4

M/J18/11/13

Mass of baby (kg) vs Mass of mother (kg)
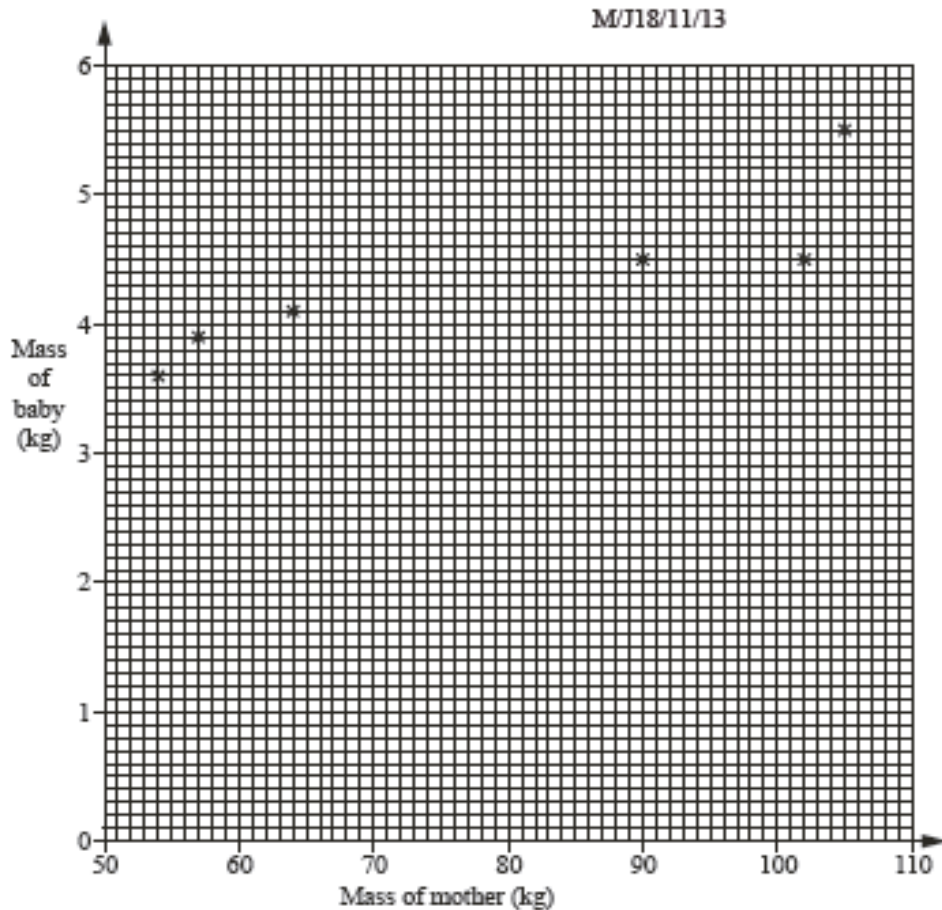
(a) On the grid, complete the scatter diagram.
The first six points have been plotted for you. [1]

(b) What type of correlation is shown on the scatter diagram?

Answer .... [1]

(c) On the scatter diagram, draw a line of best fit. [1]

(d) Anna has a mass of 82 kg and gives birth to a baby.

Use your line of best fit to estimate the mass of her baby.

Answer .... kg [1]

12 In a school of 270 children, the distance each child can swim was recorded. M/J18/12/13
The distances are summarised in the table.

| Distance (d metres)	$0 \leqslant d < 100$	$100 \leqslant d < 200$	$200 \leqslant d < 500$	$500 \leqslant d < 1000$
Number of children	110	50	60	50
Frequency density				

(a) Complete the table to show the frequency densities. [2]

(b) Calculate an estimate for the number of children who could swim more than 400 metres.

Answer .... [1]

13 The masses of 200 beetles were measured.
The results are summarised in the cumulative frequency table and part of the cumulative frequency curve is drawn. M/J18/12/16

Mass (m grams)	$m \leqslant 0.5$	$m \leqslant 1$	$m \leqslant 1.5$	$m \leqslant 2$	$m \leqslant 2.25$	$m \leqslant 2.5$	$m \leqslant 3$
Cumulative frequency	0	25	75	150	170	185	200

Mass (m grams)
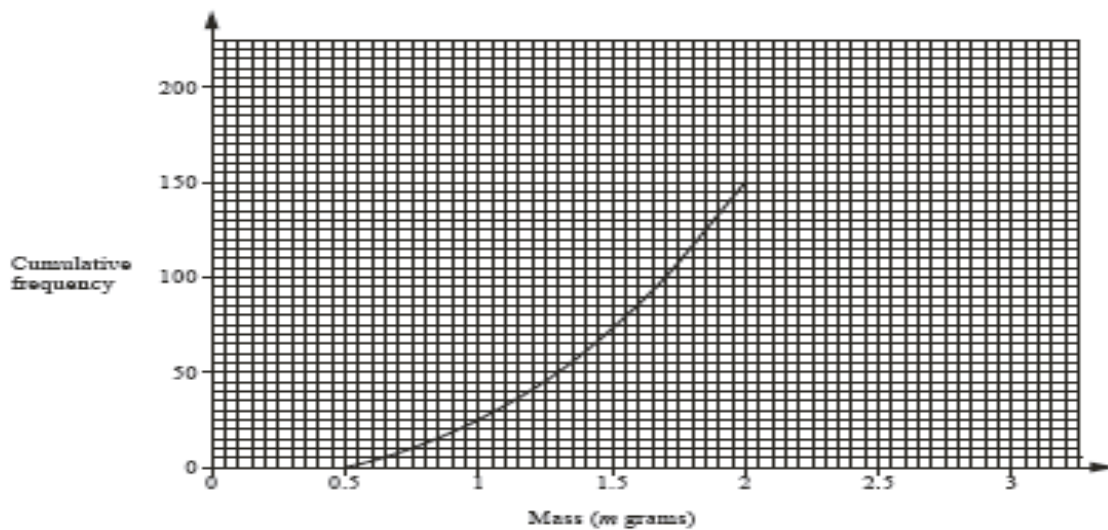
(a) Complete the cumulative frequency curve. [1]

(b) Use the curve to find an estimate for

(i) the median,

Answer .... g [1]

(ii) the lower quartile,

Answer .... g [1]

(iii) the number of beetles that have a mass greater than 1.85 grams.

Answer .... [2]

14 (a) Jenny recorded the time, in minutes, of 40 movies.
The table summarises her results.   M/J18/22/2

Time (t minutes)	$60 < t \leqslant 80$	$80 < t \leqslant 100$	$100 < t \leqslant 120$	$120 < t \leqslant 140$	$140 < t \leqslant 160$
Frequency	2	7	15	11	5

On the grid, draw a frequency polygon to represent this information.
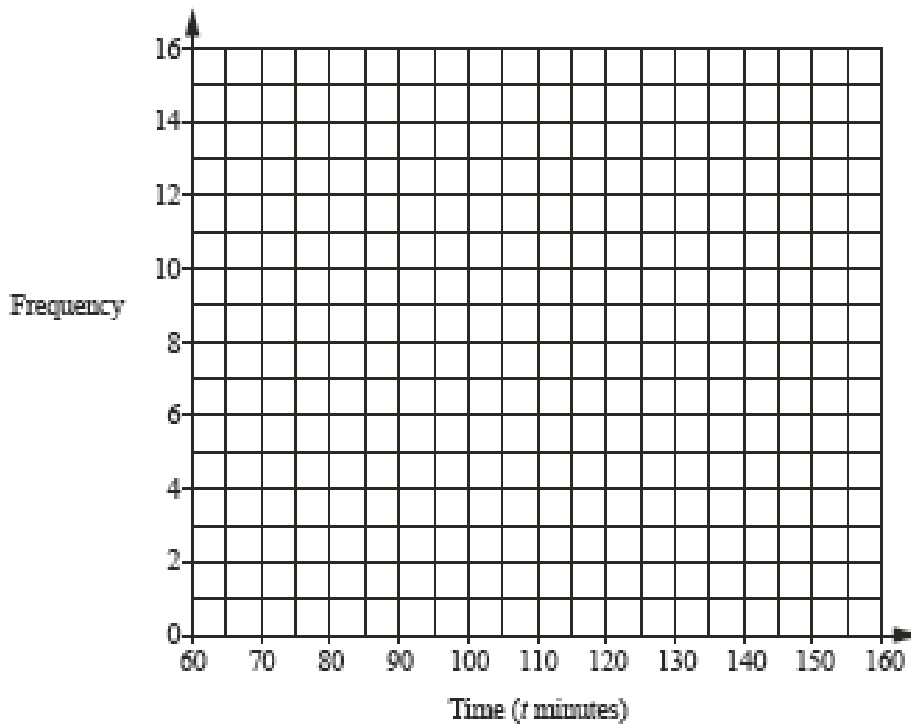
[2]

(b) Jenny asked 60 people how many movies they had each watched in the last month.
The table summarises her results.

Number of movies	0	1	2	3	4	5	6
Frequency	p	14	15	7	q	5	2

The mean number of movies watched is 2.3 .

Find the value of p and the value of q.

Answer p = ...

q = ... [3]

(c) Jenny also asked which type of movie each of the 60 people preferred.

The table summarises her results.

Type of movie	Action	Comedy	Drama	Horror
Frequency	24	15	9	12

(i) Complete the pie chart to represent the results.
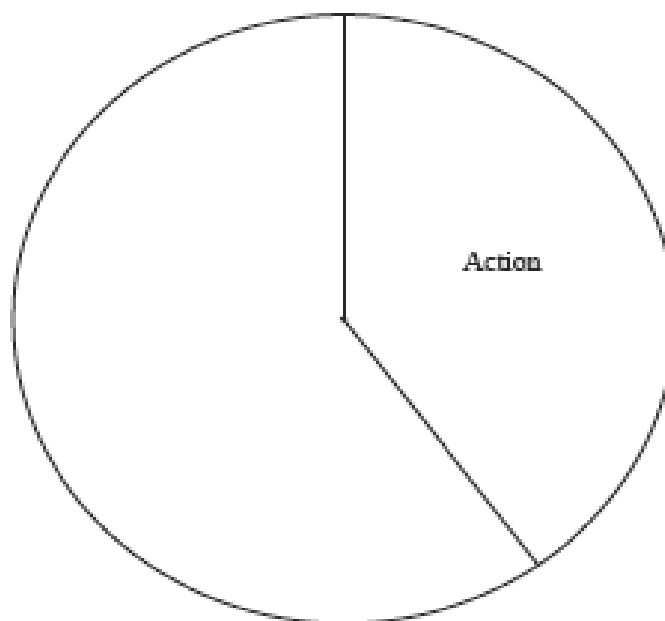
[3]

(ii) One of the 60 people is chosen at random.

Find the probability that this person preferred drama or horror movies.

Answer .... [1]

(iii) Two of the 60 people are chosen at random.

Calculate the probability that they both preferred comedy movies.

Answer .... [2]

15 One week the temperatures, in degrees Celsius, at midnight were recorded.
The results are given below.

$$-1 \qquad -3 \qquad 2 \qquad 5 \qquad -2 \qquad 1 \qquad -2$$

ON/17/11/10

Use these results to find

(a) the mode,

Answer .... [1]

(b) the median,

Answer .... [1]

(c) the mean.

Answer .... [1]
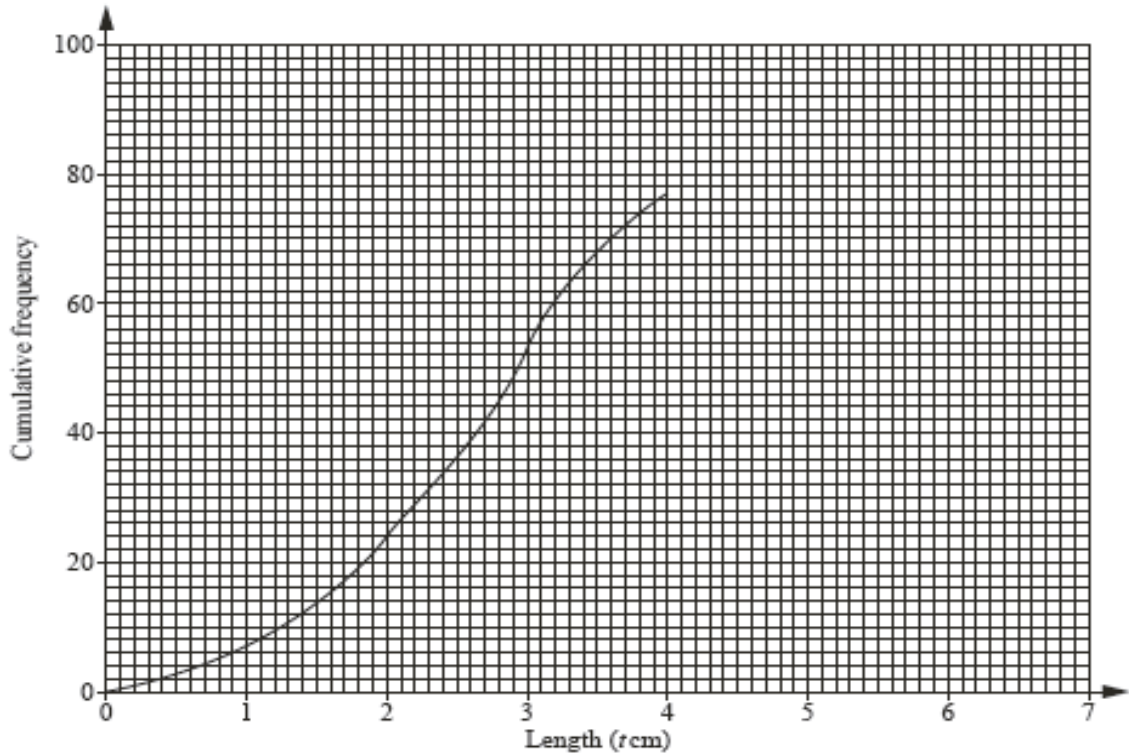
16 The lengths of 90 leaves of a plant were measured.   O/N1711/17
    The results are given in the table.
    The diagram shows part of the cumulative frequency curve.

Length (t cm)	$0 < t \leqslant 1$	$1 < t \leqslant 2$	$2 < t \leqslant 3$	$3 < t \leqslant 4$	$4 < t \leqslant 5$	$5 < t \leqslant 6$
Frequency	7	17	29	24	10	3

(a) Complete the cumulative frequency curve.   [1]

(b) Use the curve to estimate

    (i)   the median,

                                            Answer .... cm [1]

    (ii) the number of leaves with a length less than 3.5 cm.

                                            Answer .... [1]

17 (a) Here are the masses, in grams, of 8 apples.

189   175   185   192   202   161   174   196

Find the median mass.   M/J17/12/10

Answer .... g [1]

(b) A bag contains 5 carrots.
The mean mass of the carrots is 60 g.

Another carrot is added to the bag.
The mean mass of the 6 carrots is 65 g.

Work out the mass of the carrot added to the bag.

Answer .... g [2]

18   80 people were each asked how much they spent on clothes last month.
     The results are summarised in the table below.   M/J17/21/11

Amount spent ($ c)	Frequency
$0 < c \leqslant 20$	3
$20 < c \leqslant 40$	8
$40 < c \leqslant 60$	14
$60 < c \leqslant 80$	21
$80 < c \leqslant 100$	18
$100 < c \leqslant 120$	9
$120 < c \leqslant 140$	5
$140 < c \leqslant 160$	2

(a)   Calculate an estimate of the mean amount spent on clothes last month.

Answer $ .... [3]

(b)   Complete the cumulative frequency table below.

Amount spent ($ c)	$c \leqslant 20$	$c \leqslant 40$	$c \leqslant 60$	$c \leqslant 80$	$c \leqslant 100$	$c \leqslant 120$	$c \leqslant 140$	$c \leqslant 160$
Cumulative frequency	3	11						80

[1]

(c)   On the grid opposite, draw a cumulative frequency curve to represent this data.   [2]

(d)   (i)   Use your graph to estimate the median.

Answer $ .... [1]

      (ii)   Use your graph to estimate the interquartile range.

Answer $ .... [2]

Amount spent ($c)
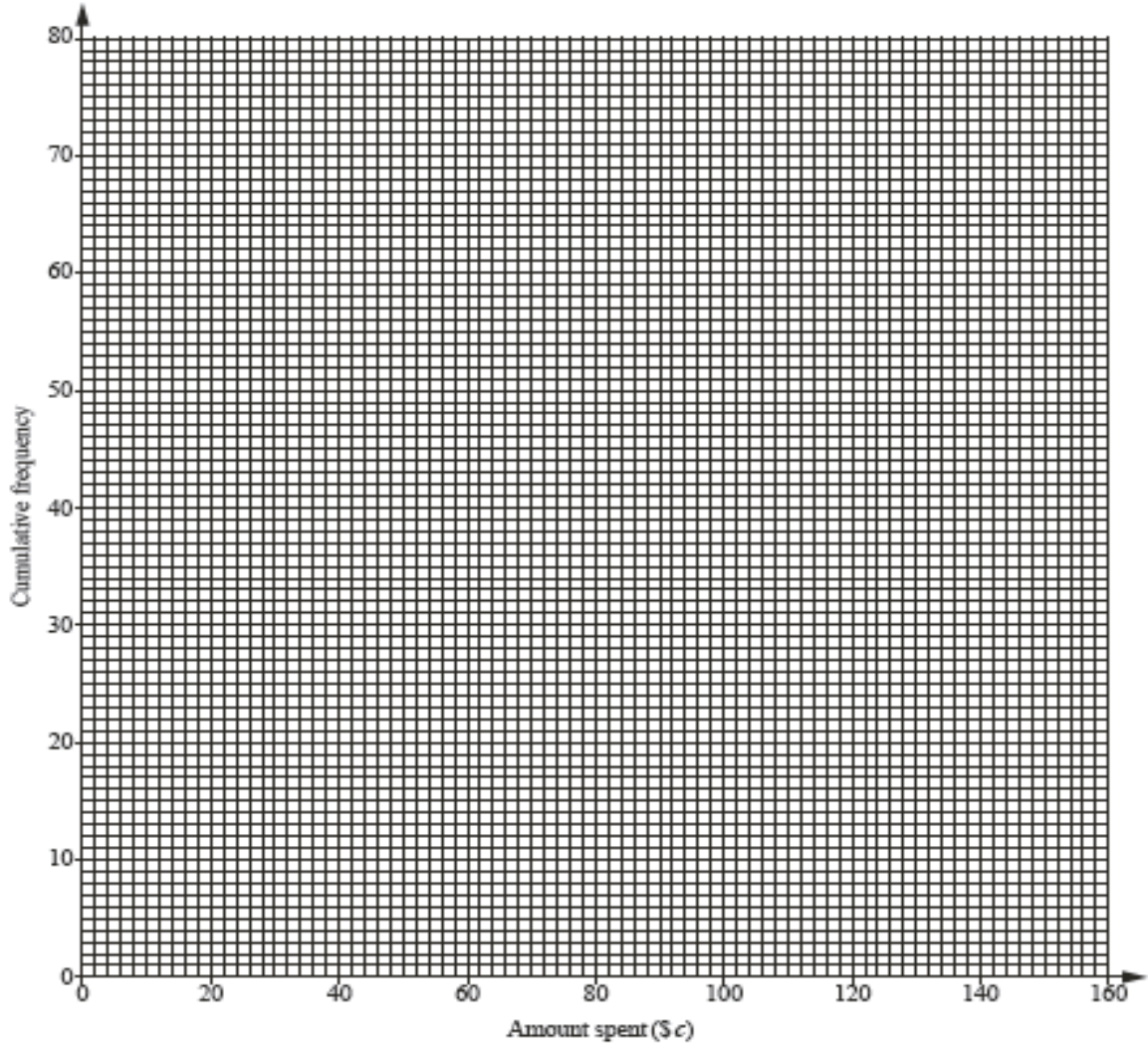
(e) The number of people who spent more than $85 last month is the same as the number of people who spent between $k and $85.

Given that k is less than 85, use your graph to estimate the value of k.

Answer $k =$ .... [3]

19 (a) The table below summarises the times taken by 50 athletes to run 400 m.

Time (t seconds)	$50 \leqslant t < 55$	$55 \leqslant t < 60$	$60 \leqslant t < 65$	$65 \leqslant t < 70$	$70 \leqslant t < 75$
Frequency	7	16	15	11	1

M/J17/22/11

(i) State the modal class.

Answer .... [1]

(ii) Calculate an estimate of the mean time taken by these athletes.

Answer .... s [3]

(iii) Calculate the probability that an athlete chosen at random took less than 60 seconds to run the 400 m.

Answer .... [2]

(b) The cumulative frequency curve summarises the times taken by 80 boys to run 200 m.
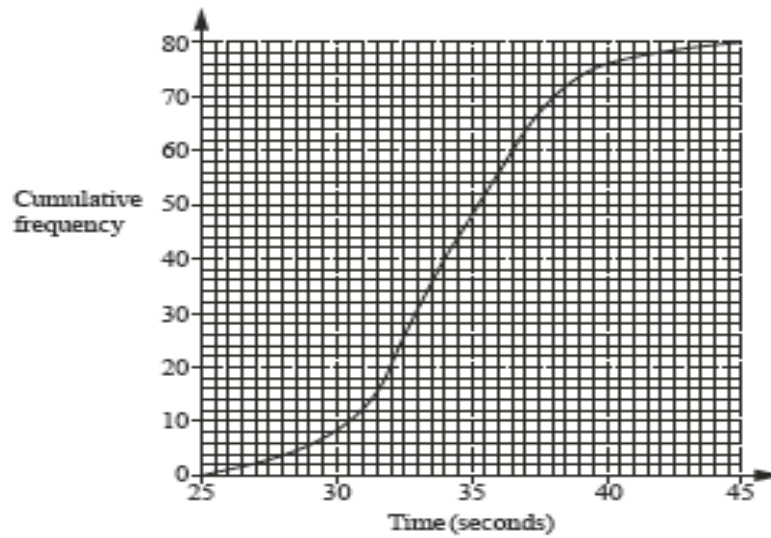
(i) Find the median time.

Answer .... s [1]

(ii) Find the interquartile range.

Answer .... s [2]

(iii) 60 girls also ran 200 m.

The girl who took the longest time ran 200 m in 40 seconds.
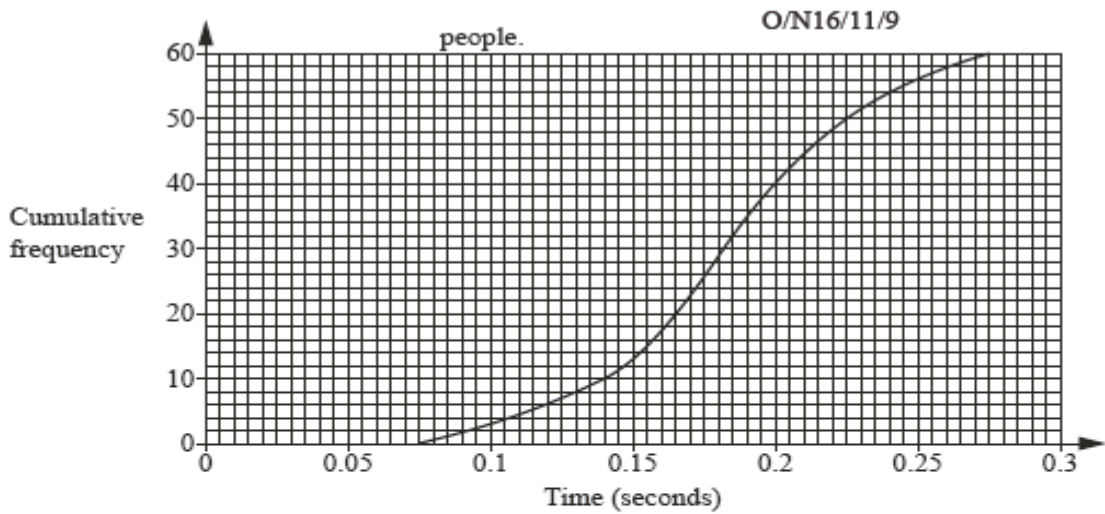The girl who took the shortest time ran 200 m in 28 seconds.

The lower quartile for the boys and the girls is the same.
The interquartile range for the girls is 4 seconds.

Draw the cumulative frequency curve on the grid above. [3]

20   The cumulative frequency graph shows information about the reaction times of 60 people.   O/N16/11/9

Use the graph to estimate

(a) the lower quartile,

Answer .... [1]

(b) the number of people who have a reaction time of more than 0.2 seconds.

Answer .... [1]

21   O/N17/12/9
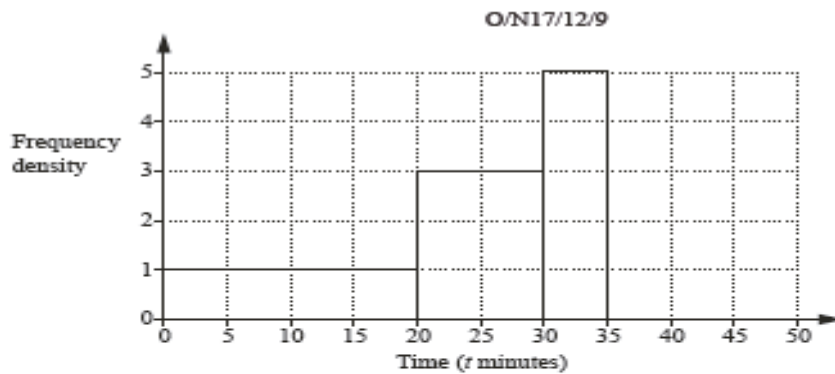
The diagram shows part of the histogram which represents the distribution of times taken by some people to travel to work.

(a) Complete the table.

Time (t minutes)	$0 < t \leqslant 20$	$20 < t \leqslant 30$	$30 < t \leqslant 35$	$35 < t \leqslant 50$
Frequency		30		30

[2]

(b) Complete the histogram.   [1]

22 Henri did a survey of the lengths of the leaves on a plant.
The results are summarised in the table.   M/J16/11/18

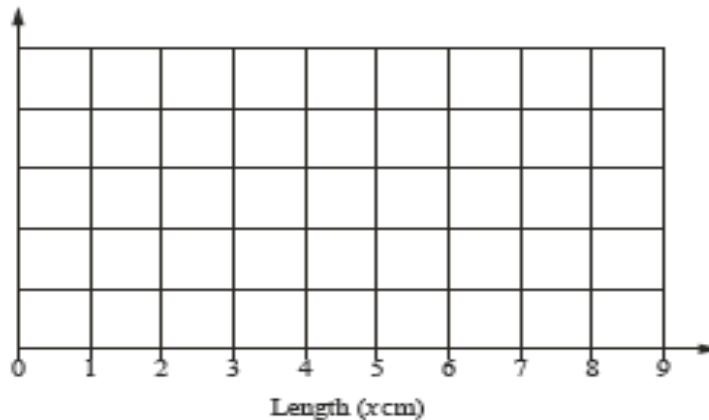| Length (x cm)	$1 < x \leqslant 3$	$3 < x \leqslant 4$	$4 < x \leqslant 5$	$5 < x \leqslant 8$
Frequency	6	8	10	6

(a) When asked to draw a histogram to illustrate the results, Henri drew the following diagram.
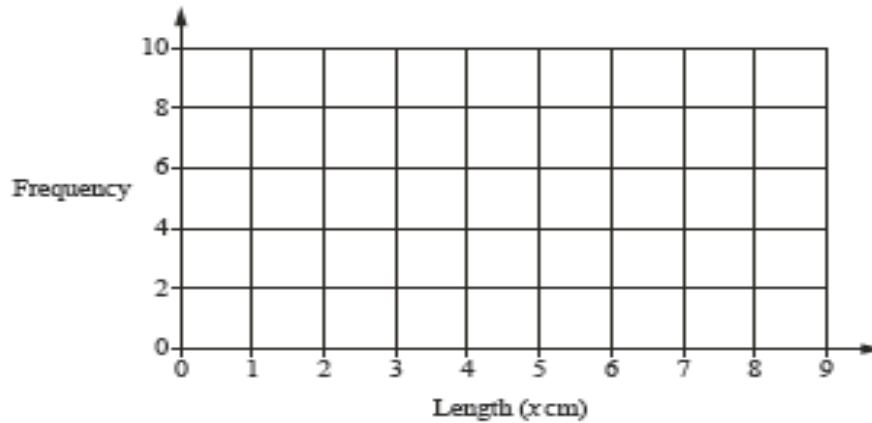
Explain why this diagram is incorrect.

....

.... [1]

(b) On the grid below, draw a correct histogram for Henri's results.

[3]

23 The number of goals scored in each of 50 football matches was recorded.
The results are given in the table.

Number of goals scored	0	1	2	3	4	5	6
Frequency	16	11	9	7	6	0	1

M/J16/11/20

For these results, find

(a) the mode,

Answer .... [1]

(b) the median,

Answer .... [1]

(c) the mean.

Answer .... [2]

24 The table and histogram show some information about the times taken by a group of students to travel to school one day.

Time (t minutes)	$0 < t \leqslant 10$	$10 < t \leqslant 20$	$20 < t \leqslant 30$	$30 < t \leqslant 60$	$60 < t \leqslant 120$
Frequency	28	40	52	18	m

M/J16/12/23
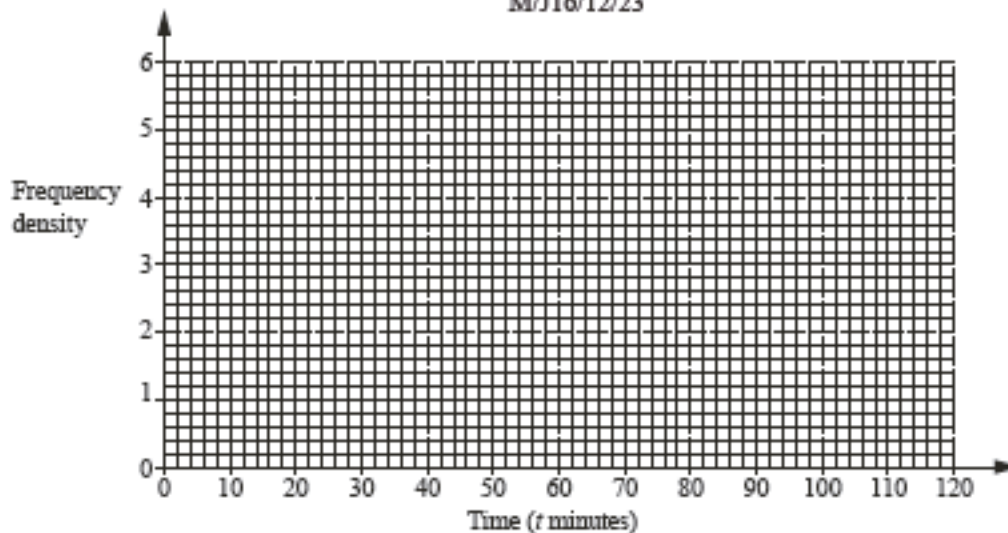
(a) Complete the histogram. [2]

(b) Find the value of m.

Answer $m =$ .... [1]

(c) Work out the fraction of students who took more than half an hour to travel to school.

Answer .... [2]

25 100 electric light bulbs of Brand A were tested to find how long each bulb lasted.
The results are summarised in the table below.   M/J16/21/10

Time (t hours)	t ≤ 50	50 < t ≤ 100	100 < t ≤ 150	150 < t ≤ 200	200 < t ≤ 250	250 < t ≤ 300	300 < t ≤ 350
Number of bulbs	2	2	10	40	30	14	2

(a) Complete the cumulative frequency table.

Time (t hours)	t ≤ 50	t ≤ 100	t ≤ 150	t ≤ 200	t ≤ 250	t ≤ 300	t ≤ 350
Cumulative frequency	2	4					100

[1]

(b) On the grid, draw a smooth cumulative frequency curve to represent this information.
Label this curve Brand A.

[2]
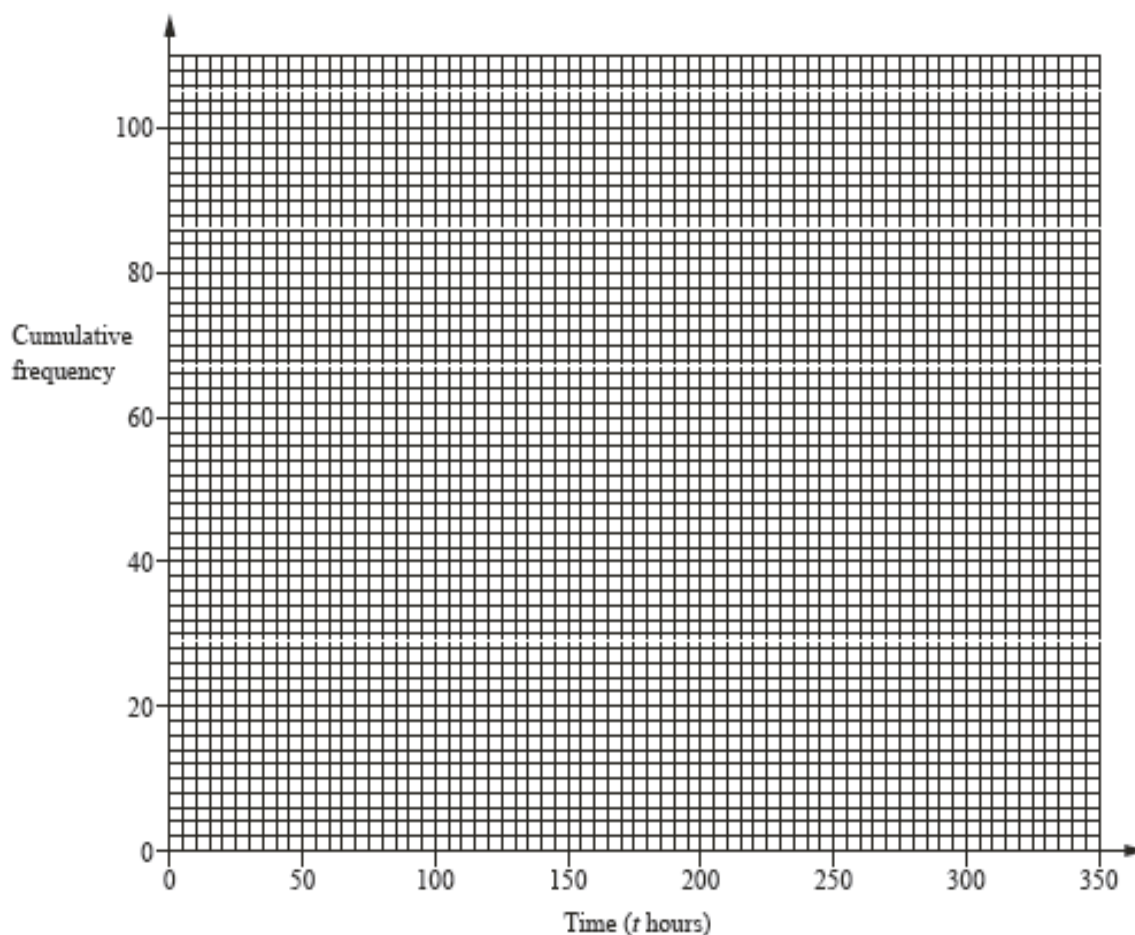
(c) (i) Use your graph to estimate the median.

Answer .... hours [1]

(ii) Use your graph to estimate the interquartile range.

Answer .... hours [2]

(d) 100 Brand B bulbs gave the following results.
4 bulbs lasted 50 hours or less.
The longest time any bulb lasted was 300 hours.
The median is 250 hours.
The upper quartile is 275 hours.
The interquartile range is 75 hours.

On the grid, draw and label the cumulative frequency curve for the Brand B bulbs.   [4]

(e) Using your graph, estimate the number of Brand A bulbs that lasted 275 hours or less.

Answer .... [1]

(f) Complete the statement below.

Brand .... had .... more bulbs that lasted longer than 275 hours than Brand .... [1]

26 Steven asked 25 women how many children they have. M/J16/22/2
The results are summarised in the table below.

Number of children	Frequency
0	7
1	5
2	6
3	4
4	3

(a) Find

(i) the mean,

Answer .... [2]

(ii) the median,

Answer .... [1]

(iii) the mode.

Answer .... [1]

(b) Steven says that the mode is the average that best represents the data.

Explain why Steven is wrong.

Answer .... [1]

(c) Steven chooses two women at random from the group.

Calculate the probability that **both** of them have just one child.
Give your answer as a fraction in its simplest form.

Answer .... [2]

Marking Scheme

1(a)	4 16 30 52 70 80	1	
1(b)	Correct **ft** curve	2	**B1** for at least 5 correct **ft** points
1(c)	16 to 19 **ft**	2	**B1** for *their* cumulative frequency (CF) at $m = 45$ **ft** After 0, allow **B1** for $80 - their$ CF at $m = 44$

2(a)(i)	6 correct plots	2	**P1** for 4 or 5 correct plots
2(a)(ii)	Positive	1	
2(a)(iii)	Line of best fit	1	**Ruled** line at least from $x = 5$ to $x = 48$, with at least 3 points on each side and drawn so would cut axis between (5, 0) and (0, 20)
2(a)(iv)	Physics (integer) value on line at M = 22	1	**Strict ft** from *their* single ruled line $5 \leqslant x \leqslant 48$.
2(b)	45	1	
2(c)	$(26 + 39 + 35 + 28 + 9 + 37 + 45 + 33 + 16 + 12) \div 10$	1	
2(d)	46 **cao**	3	**M2** for $(31 \times 12 - 28 \times 10) \div 2$ **soi** by $92 \div 2$ Or **M1** for 31×12 **soi** by 372 or 93
2(e)	$\dfrac{1}{15}$ **oe**	2	**M1** for $\dfrac{3}{10} \times \dfrac{2}{9}$ or for $\dfrac{k}{10} \times \dfrac{k-1}{9}$ with $k \geqslant 2$

3(a)(i)	Apple Banana Orange 10 8 12	1	

3(a)(ii)	Correct bar chart	2	**FT** *their* frequency table **B1** for bars labelled correctly and of equal width or bars drawn to correct height.
3(b)(i)	$\dfrac{7}{30}$ **oe**	1	
3(b)(ii)	48	1	

4(a)	$\dfrac{22}{80}$ or $\dfrac{11}{40}$ oe	1	
4(b)	33.875 or 33.9 nfww	3	**B1** for correct use of midpoints **M1** for $\dfrac{\sum fx}{80}$
4(c)	Correct cumulative frequency diagram	3	**B2** for at least 5 correct points plotted **or** **B1** for 6, 22, 47, 65, 78, 80 soi

4(d)(i)	32.5 to 34	1	FT *their* cumulative frequency diagram
4(d)(ii)	8 to 9.5	2	FT *their* cumulative frequency diagram **M1** for [LQ] 29 to 30 nfww **or** [UQ] 38 to 39 nfww
4(e)	6, 7 or 8	2	**M1** for attempt to read at 43

5(a)	(16, 23) (18,18) (20, 12) (21, 8) plotted	2	**B1** for 2 or 3 correct plots
5(b)	negative	1	
5(c)	Ruled line of best fit	1	
5(d)	Strict FT *their* ruled line of best fit read at Sydney = 15°C	1	

6(a)	$\dfrac{23}{75}$ oe	1	
6(b)	192 or 192.4 to 192.5 nfww	3	**M1** for 4×165, 10×175, 15×185, 23×195, 18×205 and 5×215 **M1** for (4×165 + 10×175 + 15×185 + 23×195 + 18×205 + 5×215) ÷ 75 oe After 0 scored, SC1 for $\sum fx \div 75$
6(c)	Correct cumulative frequency curve	3	**B2** for at least 5 points plotted correctly or **B1** for 4, 14, 29, 52, 70, 75 soi
6(d)	193 to 195	1	Dep on no clear indication of incorrect CF
6(e)	202 to 205	2	**FT** *their* increasing curve **B1** for 60 nfww
6(f)	11 or 12 cao	1	

7	Correct histogram with Freq. densities 1, 1.5, 2, 2.4, 0.8	3	**B1** for 4 or more rectangles on correct bases. **B1** for 4 or more correct Freq. densities, soi.

8(a)	5 correct plots	2	**B1** for 3 or 4 correct plots
8(b)	Negative	1	
8(c)	Line of best fit	1	Acceptable **ruled** straight line.
8(d)	Reading their value for $x = 7.6$	1	Strict **FT**.

Question	Answer	Marks	Partial Marks
9(a)(i)	6 points plotted correctly	2	**B1** for 4 or 5 points plotted correctly
9(a)(ii)	Positive	1	
9(a)(iii)	$\dfrac{3}{5}$ **cao**	2	**B1** for $\dfrac{6}{10}$ **oe** seen
9(a)(iv)	Ruled line of best fit	**B1**	
9(a)(v)	54 to 58	**B1**	**FT** reading from *their* straight line of best fit at 48 km
9(b)(i)	$20 < t \leqslant 40$	1	
9(b)(ii)	39.5	3	**B1** for correct use of midpoints **soi** **M1** for $(10 \times 29 + 30 \times 38 + 50 \times 26 + 70 \times 21 + 90 \times 6) \div 120$ **oe**
9(b)(iii)	22.5	2	**B1** for $21 + 6$ or 27 seen

Question	Answer	Marks	Partial Marks
10(a)	Correct cumulative frequency curve	2	**B1** for at least 5 correct plots
10(b)(i)	118 to 120	1	
10(b)(ii)	14 to 18 nfww	2	**M1** for reading at CF 150 or 50
10(c)	On average Lim's tomatoes had lower masses oe	**B1**	**Strict FT** their median
	Masses of Ravi's tomatoes were more consistent oe	**B1**	**Strict FT** their IQR
10(d)(i)	46, 26	1	
10(d)(ii)	$110 < m \leqslant 120$	1	**FT** their frequency table

Question	Answer	Marks	Partial Marks
10(d)(iii)	118.8	3	**B1** for correct use of midpoints soi **M1** for $(20 \times 90 + 28 \times 105 + 64 \times 115 + their\ 46 \times 125 + their\ 26 \times 135 + 16 \times 150) \div 200$ oe

Question	Answer	Marks	Partial Marks
11(a)	4 points correctly plotted	1	
11(b)	positive	1	
11(c)	Ruled line of best fit drawn	1	
11(d)	$4.35 - 4.55$	1	Dependent on a line of best fit or FT *their* straight line of best fit with +ve gradient

12(a)	1.1 0.5 0.2 0.1 oe	2	**B1** for 2 or 3 correct
12(b)	70 cao	1	

13(a)	Correct completion of the curve	1	
13(b)(i)	1.7	1	
13(b)(ii)	1.3	1	
13(b)(iii)	75	2	**B1** for 125 seen or **SC1** for answer 74 or 76

14(a)	Correct frequency polygon (ruled lines)	2	**B1** for 4 or 5 heights correct soi
14(b)	$q = 9$	B2	**M1** for $[0 \times p] + 1 \times 14 + 2 \times 15 + 3 \times 7 + 4 \times q + 5 \times 5 + 6 \times 2$ oe
	$p = 17 - \text{their } q$	B1	**Strict FT** provided q integer with $0 \leqslant q \leqslant 17$
14(c)(i)	Correct labelled pie chart: C[omedy] , D[rama] , H[orror]	3	**B2** for correct sectors without labels or incorrect labels or **B1** for one correct sector or 90, 54 and 72 seen
14(c)(ii)	$\dfrac{21}{60}, \dfrac{7}{20}, \dfrac{126}{360}$, 0.35 or 35%	1	
14(c)(iii)	$\dfrac{210}{3540}$ oe	2	**M1** for $\dfrac{15}{60} \times \dfrac{14}{59}[\times 2]$ or **SC1** for $\left(\dfrac{15}{60}\right)^2$ or answer $\dfrac{1}{16}$ oe

15(a)	−2	1	
15(b)	−1	1	
15(c)	0	1	

16(a)	Correct curve from (4, 77) to (6, 90) via (5, 87)	1	
16(b)(i)	2.8	1	
16(b)(ii)	67 or 68	1	

17(a)	187	1	
17(b)	90	2	**M1** for 65×6 **and** 60×5 soi

18(a)	75 nfww	3	M2 for $\dfrac{\sum \text{frequency} \times \text{midvalue}}{80}$ oe or M1 for $\sum fc$
18(b)	25, 46, 64, 73, 78	1	
18(c)	8 points correctly plotted and joined	2	FT increasing curve B1 for at least 6 points correctly plotted
18(d)(i)	74 to 76	1	
18(d)(ii)	36 to 44	2	B1 for 52 to 56 <u>and</u> 92 to 96 seen
18(e)	54 to 62	3	B1 for 27 to 29 M1 for attempt to read at $(80 - 2 \times \textit{their } 28)$

19(a)(i)	$55 \leqslant t < 60$	1	
19(a)(ii)	60.8 nfww	3	M2 for $\dfrac{\sum \text{frequency} \times \text{midvalue}}{50}$ oe or M1 for $\sum ft$

19(a)(iii)	$\dfrac{23}{50}$ or 0.46 or 46%	2	B1 for 23 seen or 16 + 7 seen
19(b)(i)	34	1	
19(b)(ii)	4.5	2	B1 for 31.5 to 32.5 <u>and</u> 36 to 37 seen
19(b)(iii)	(28, 0) (32, 15) (36, 45) (40, 60) plotted and points joined	3	B2 for at least 3 correct points plotted or B1 for 2 correct points plotted or (28, 0) (32, 15) (36, 45) and (40, 60) seen

20	(a)	0.155 cao	1	
	(b)	20 WWW	1 *	

21	(a)	20 25	1 1	
	(b)	Rectangle with base 35 to 50 and height 2	1	

22	(a)	Vertical axis label should be 'Frequency density' or heights should be 3, 8, 10, 2.	1	
	(b)	Rectangles with same bases as in (a), with heights 3, 8, 10, 2. Vertical label 'Frequency density' and a suitable scale.	3 *	**C2** for 4 bars correct, with no label or incorrect scale on vertical axis or for 3 bars correct with 'Frequency density' label and numbered linear scale. **C1** for numbers 3, 8, 10, 2; or 'Frequency density' label or for 3 bars correct

23	(a)	0	1	
	(b)	1	1	
	(c)	1.6 oe	2*	**M1** for $(11 \times 1 + 9 \times 2 + 7 \times 3 + 6 \times 4 + 1 \times 6)/50$

24	(a)	Two correct bars drawn	2	**C1** for rectangle base 0 to 10 height 2.8 or for rectangle base 30 to 60 height 0.6
	(b)	12	1	
	(c)	$\frac{30}{150}$ oe or $\frac{18+m}{138+m}$ oe evaluated	2ft*	**B1 FT** for fraction with numerator or denominator correct or for answer 20% or 0.2

25	(a)	(2) (4) 14 54 84 98 (100)	1	
	(b)	Correct curve	2	**P1** for at least 5 correct plots
	(c) (i)	195 ft 190 $\leqslant$ and < 200	1	
	(ii)	50 –75	2	**B1** for one quartile correct in ranges 225 to 235 or 160 to 175
	(d)	Correct curve	4	**P3** for at least 4 correct plots or **B1 + B1** for any two correct points soi.
	(e)	92 ft	1	
	(f)	B 15 ft A	1ft	Their 90 – 75

26	(a)	(i)	1.64 or $1\frac{16}{25}$	2	M1 for $\dfrac{0 \times 7 + 1 \times 5 + 2 \times 6 + 3 \times 4 + 4 \times 3}{7 + 5 + 6 + 4 + 3}$
		(ii)	2	1	
		(iii)	0	1	
	(b)		appropriate reason	1	
	(c)		$\dfrac{1}{30}$ cao	2	M1 for $\dfrac{5}{25} \times \dfrac{4}{24}$ oe
	(d)		Correct bar chart with axes labelled	2	B1 if only one error (eg incorrect height, scales missing / incorrect, inconsistent bar widths, or 4 correct bars)
	(e)		0 0 1 3 4	1	