

Statistics

Difficulty: Medium

Model Answers 1

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Medium
Booklet	Model Answers 1

Time allowed: **114 minutes**

Score: **/99**

Percentage: **/100**

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980) *Assembled by AS*

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

120 students take a mathematics examination.

- (a) The time taken, m minutes, for each student to answer question 1 is shown in this table.

Time (m minutes)	$0 < m \leq 1$	$1 < m \leq 2$	$2 < m \leq 3$	$3 < m \leq 4$	$4 < m \leq 5$	$5 < m \leq 6$	[4]
Frequency	72	21	9	11	5	2	

Calculate an estimate of the mean time taken.

The mean value is estimated by assuming that all students in a given group have the mid-value of the time for that group (i.e. 0.5 min, 1.5 min, 2.5 min, 3.5 min, 4.5 min, 5.5 min).

We sum the products of the mid-time and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

$$\text{mean} = \frac{0.5 \times 72 + 1.5 \times 21 + 2.5 \times 9 + 3.5 \times 11 + 4.5 \times 5 + 5.5 \times 2}{120}$$

$$\text{mean} = 1.35 \text{ minutes}$$

(b) (i) Using the table in **part (a)**, complete this cumulative frequency table.

[2]

Time (m minutes)	$m \leq 1$	$m \leq 2$	$m \leq 3$	$m \leq 4$	$m \leq 5$	$m \leq 6$
Cumulative frequency	72					120

For example, to get the cumulative frequency for $m \leq 3$, we sum the corresponding

frequencies (up to $2 < m \leq 3$) from the previous graph.

$$m \leq 3 = 72 + 21 + 9 = 102$$

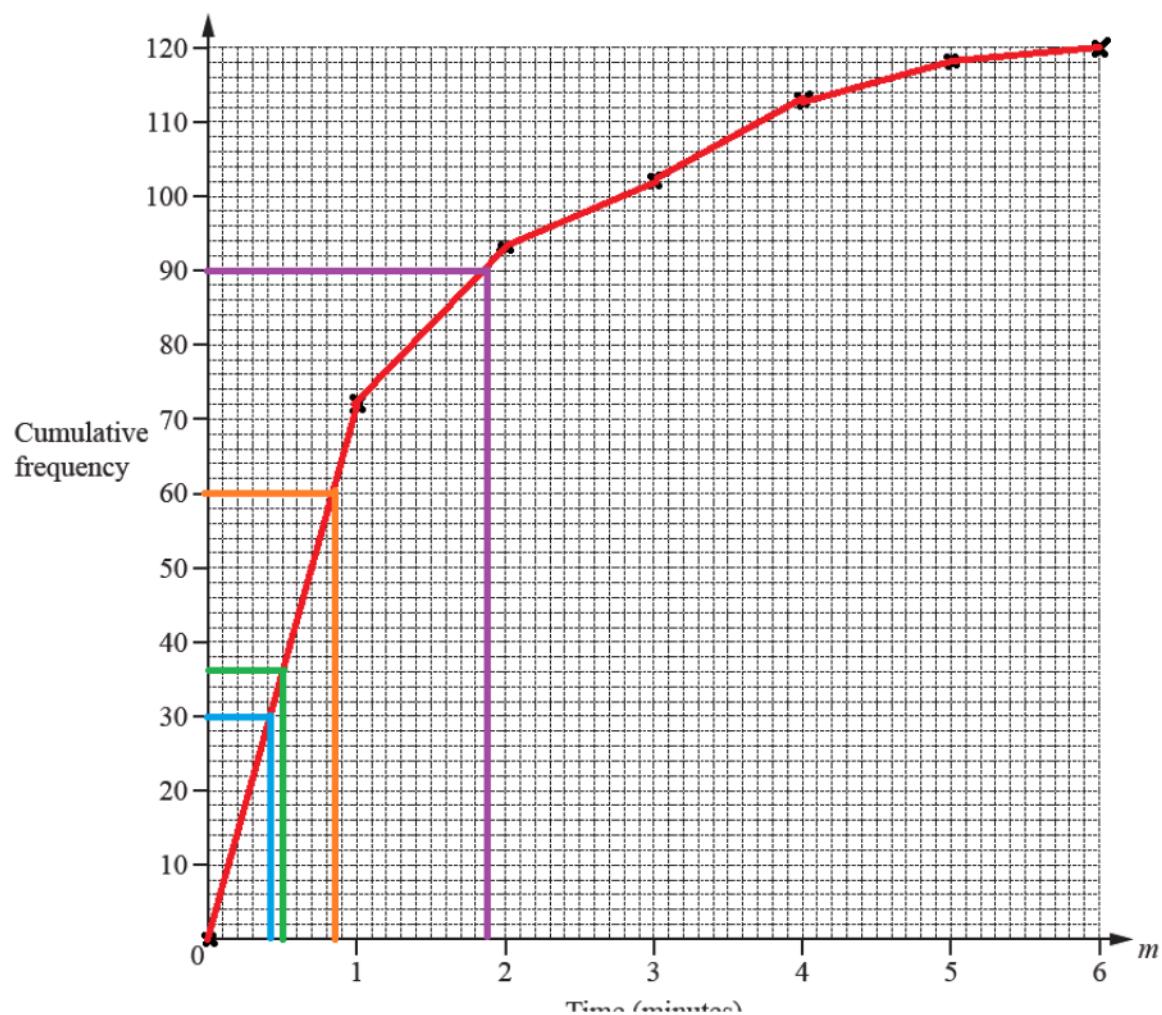
The cumulative frequency for $m \leq 3$ is 102.

Time (m minutes)	$m \leq 1$	$m \leq 2$	$m \leq 3$	$m \leq 4$	$m \leq 5$	$m \leq 6$
Cumulative frequency	72	93	102	113	118	120

(ii) Draw a cumulative frequency diagram to show the time taken.

[3]

Plot the points from the table on the grid and connect them with a curve (red).



- (iii) Use your cumulative frequency diagram to find

- (a) the median, [1]

The median is the 50th percentile. As there are 120 students, it is the time taken of 60th student (orange lines). From the graph, this is **0.85 minutes**.

- (b) the inter-quartile range, [2]

The inter-quartile range is the difference between the upper and the lower quartile.

The lower quartile is the value of student who is 25% of 120, which is 30. This is 0.4 min. (blue line)

The upper quartile is the value of student who is 75% of 120, which is 60. This is 1.9 min. (purple)

The inter-quartile range is therefore **1.5 minutes**.

- (c) the 35th percentile. [2]

We calculate 35% of 120 (total number of students). This is 36.

Therefore we are looking for the value of 36th reading (green lines).

From the graph, the 30th percentile is **0.5 minutes**.

- (c) A new frequency table is made from the table shown in part (a). [2]

- (i) Complete the table above.

Time (m minutes)	$0 < m \leq 1$	$1 < m \leq 3$	$3 < m \leq 6$
Frequency	72		

We use the same principle as in part b)i), but now only sum the values, which are within our range.

$$1 < m \leq 3 = 1 < m \leq 2 + 2 < m \leq 3$$

$$1 < m \leq 3 = 21 + 9 = \mathbf{30}$$

$$3 < m \leq 6 = 3 < m \leq 4 + 4 < m \leq 5 + 5 < m \leq 6$$

$$3 < m \leq 6 = 11 + 5 + 2 = 18$$

- (ii) A histogram was drawn and the height of the first block representing the time $0 < m \leq 1$ was 3.6 cm.
Calculate the heights of the other two blocks. [3]

The width of the first block is 1 (m between 0 and 1). For the second block and the third blocks, the widths are 2 and 3, respectively (they contain more values of m).

Before calculating the height, we divide the frequency of each group by the width to get the frequency per width:

- Frequency per width of $1 < m \leq 3$: 15
- Frequency per width of $3 < m \leq 6$: 6

The ratio of the height and the frequency per width must be the same for all blocks, so we use the ratio from the first block to calculate the rest.

$$\frac{3.6 \text{ cm}}{72} = \frac{\text{height}}{\text{frequency per width}}$$

$1 < m \leq 3$:

$$\text{height} = \frac{3.6}{72} \times 15 = 0.75 \text{ cm}$$

$3 < m \leq 6$:

$$\text{height} = \frac{3.6}{72} \times 6 = 0.3 \text{ cm}$$

Question 2

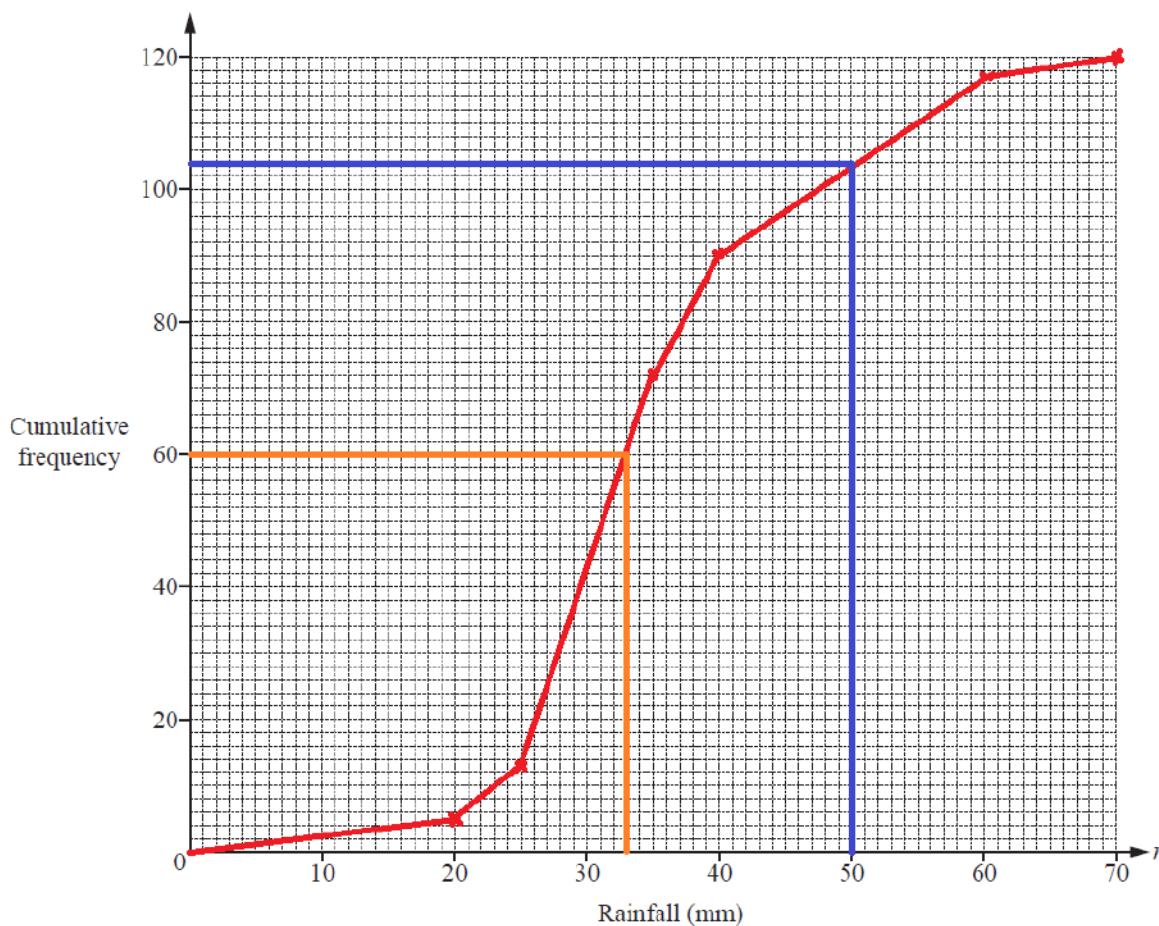
Leo measured the rainfall each day, in millimetres, for 120 days.

The cumulative frequency table shows the results.

Rainfall (r mm)	$r \leq 20$	$r \leq 25$	$r \leq 35$	$r \leq 40$	$r \leq 60$	$r \leq 70$
Cumulative frequency	5	13	72	90	117	120

- (a) On the grid below, draw a cumulative frequency diagram to show these results. [3]

Plot the points from the table on the grid and connect them with lines (red).



- (b) (i) Find the median. [1]

The median is the 50th percentile. As there are 120 days, it is the rainfall of 60th day (orange lines).

From the graph, this is 33 mm.

- (ii) Use your diagram to find the number of days when the rainfall was more than 50mm. [2]

From the diagram, we can see that there were 104 days with less rainfall than 50mm.

Subtract this number from the total number of days that were recorded.

$$120 \text{ days} - 104 \text{ days} = 16 \text{ days}$$

Therefore there were 16 days when the rainfall was more than 50 mm.

- (c) Use the information in the cumulative frequency table to complete the frequency table below. [2]

Rainfall (r mm)	$0 < r \leq 20$	$20 < r \leq 25$	$25 < r \leq 35$	$35 < r \leq 40$	$40 < r \leq 60$	$60 < r \leq 70$
Frequency	5		59			3

We subtract the boarder values of our regions.

$$20 < r \leq 25 = r \leq 25 - r \leq 20$$

$$20 < r \leq 25 = 13 - 5 = 8$$

$$35 < r \leq 40 = r \leq 40 - r \leq 35$$

$$35 < r \leq 40 = 90 - 72 = 18$$

$$40 < r \leq 60 = r \leq 60 - r \leq 40$$

$$40 < r \leq 60 = 117 - 90 = 27$$

- (d) Use your frequency table to calculate an estimate of the mean.

You must show all your working.

[4]

The mean value is estimated by assuming that during all days within a given group the rainfall was equal to the rainfall mid-value of that group (i.e. 10mm, 22.5mm, 30mm, 37.5mm, 50mm, 65mm).

We sum the products of the mid-time and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

mean

$$= \frac{10\text{mm} \times 5 + 22.5\text{mm} \times 8 + 30\text{mm} \times 59 + 37.5\text{mm} \times 18 + 50\text{mm} \times 27 + 65\text{mm} \times 3}{120 \text{ days}}$$

$$\text{mean} = 35.2 \text{ mm}$$

- (e) In a histogram drawn to show the information in the table in **part (c)**, the frequency density for the interval

$$25 < r \leq 35 \text{ is } 5.9.$$

Calculate the frequency density for the intervals $20 < r \leq 25$, $40 < r \leq 60$ and $60 < r \leq 70$. [4]

The interval $25 < r \leq 35$ has width 10 (=35-25) and frequency 59.

Calculate the density times frequency per width.

$$\frac{\text{density}}{\text{frequency}} \times \text{width} = \frac{5.9}{59} \times 10 = 1 \text{ per width}$$

To get the frequency density for each interval multiply by its frequency and divide by the width of a given interval (the difference between the interval boundaries) and the.

Frequency density of $20 < r \leq 25$:

$$\frac{1 \times 8}{5} = 1.6$$

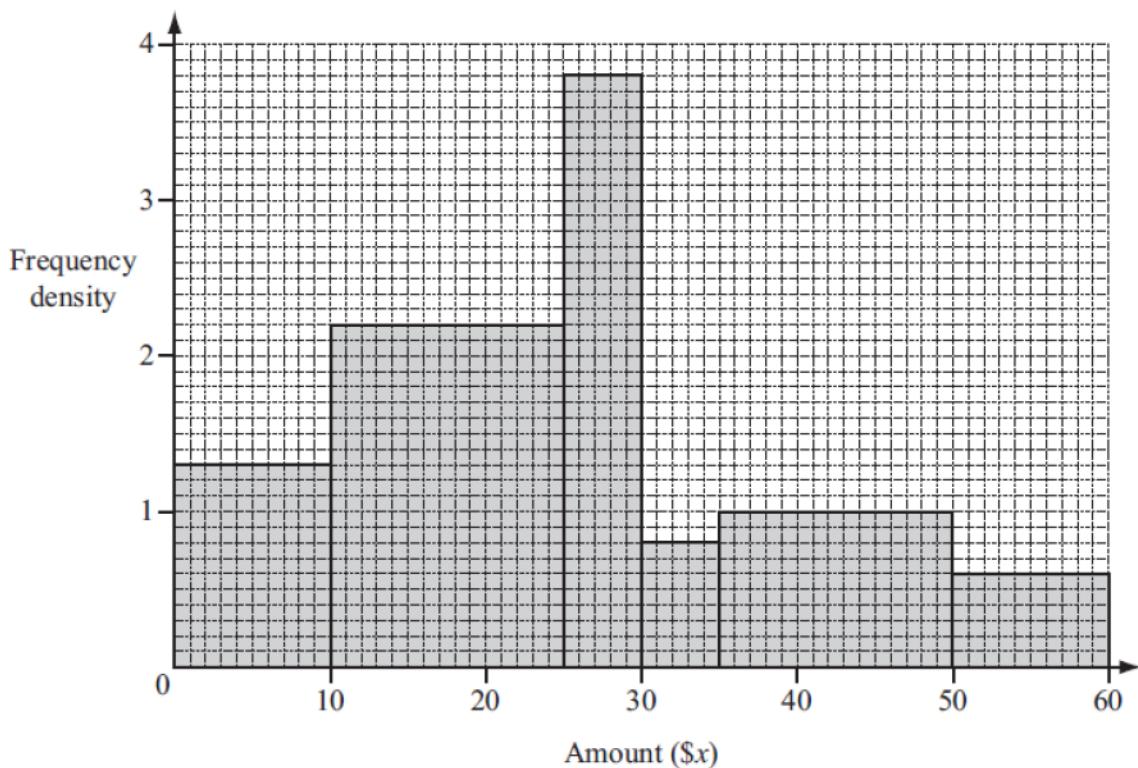
Frequency density of $40 < r \leq 60$:

$$\frac{1 \times 27}{20} = 1.35$$

Frequency density of $60 < r \leq 70$:

$$\frac{1 \times 3}{10} = 0.3$$

Question 3



A survey asked 90 people how much money they gave to charity in one month. The histogram shows the results of the survey.

- (a) Complete the frequency table for the six columns in the histogram. [5]

Amount (\$x)	$0 < x \leq 10$					
Frequency				4		

Amount (\$x)	Frequency, f	Midpoint, m	$f \times m$
$0 < x \leq 10$	13	5	65
$10 < x \leq 25$	33	17.5	577.5
$25 < x \leq 30$	19	27.5	522.5
$30 < x \leq 35$	4	32.5	130
$35 < x \leq 50$	15	42.5	637.5
$50 < x \leq 60$	6	55	330
Sum	90		2262.5

- (b) Use your frequency table to calculate an estimate of the mean amount these 90 people gave to charity.

[4]

Estimate of the mean is

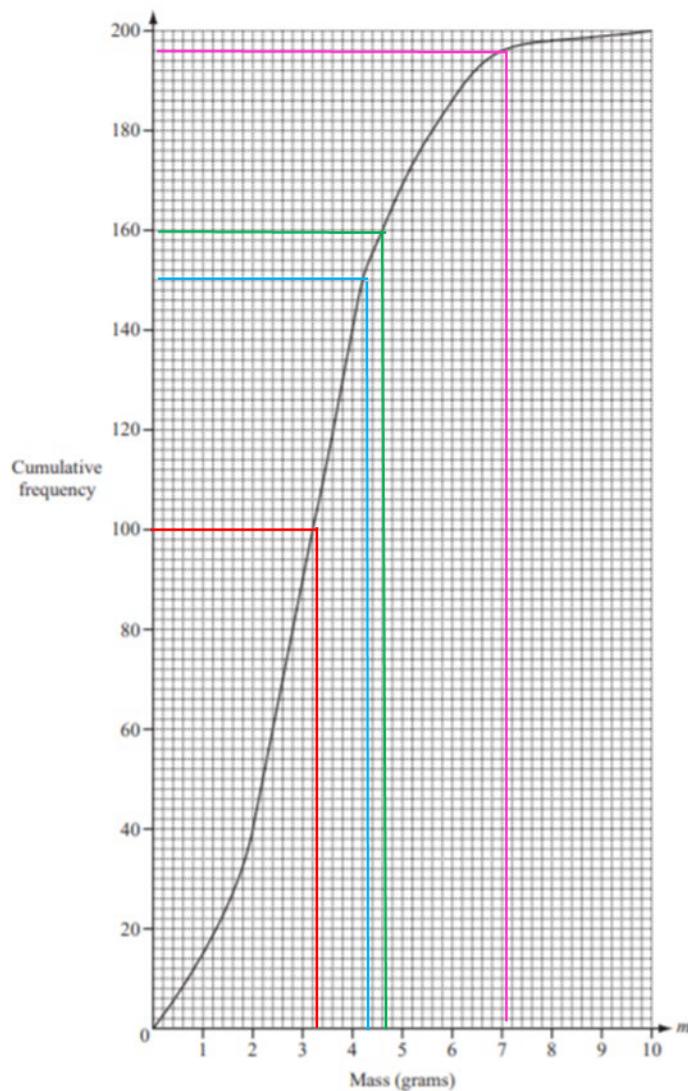
$$\mu = \frac{\sum f \times m}{\sum f}$$

$$= \frac{2262.5}{90}$$

$$= 25.14$$

Question 4

200 students estimate the mass (m grams) of a coin.
The cumulative frequency diagram shows the results.



(a) Find

- (i) the median,

[1]

To find the median we find the mass where half of the students thought it was lighter than. i.e. Cumulative frequency = 100

Mass = 3.2g

- (ii) the upper quartile,

[1]

To find the upper quartile we find the mass where three quarters of the students

thought it was lighter than. ie. Cumulative frequency = 150

Mass = 4.2g

- (iii) the 80th percentile,

[1]

To find the 80th percentile we find the mass where 80% of the students thought

it was lighter than. ie. Cumulative frequency = 160

Mass = 4.6g

- (iv) the number of students whose estimate is 7 g or less.

[1]

To find the number of students that guessed less than 7g, we find the point on

the curve that corresponds to this mass and draw a line across to find the

cumulative number of students.

Number of students = 196

- (b) (i) Use the cumulative frequency diagram to complete the frequency table.

[2]

To find the number of students the guessed in each interval we must find the

difference in the cumulative frequencies at the boundary of each interval.

Mass (m grams)	$0 < m \leq 2$	$2 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 8$	$8 < m \leq 10$
Frequency	40	100	46	12	2

- (ii) A student is chosen at random.

The probability that the student estimates that the mass is greater than M grams is 0.3.

Find the value of M .

[2]

Remember that Cumulative Frequency graphs give "less than" so it will

be quicker to say:

$$P(m < M) = 1 - 0.3 = 0.7$$

So we take the value 70% (0.7) of the way up the CF axis (ie 140) and read

across and down to get:

$$M = 4 \text{ g}$$

Question 5

- (a) A farmer takes a sample of 158 potatoes from his crop. He records the mass of each potato and the results are shown in the table.

Mass (m grams)	Frequency
$0 < m \leq 40$	6
$40 < m \leq 80$	10
$80 < m \leq 120$	28
$120 < m \leq 160$	76
$160 < m \leq 200$	22
$200 < m \leq 240$	16

Calculate an estimate of the mean mass.

Show all your working.

[4]

Mass (m grams)	Frequency (f)	Midpoint (p)	$f \times p$
$0 < m \leq 40$	6	20	120
$40 < m \leq 80$	10	60	600
$80 < m \leq 120$	28	100	2800
$120 < m \leq 160$	76	140	10640
$160 < m \leq 200$	22	180	3960
$200 < m \leq 240$	16	220	3520

Sum 158 21640

$$\text{Estimated mean} = \frac{\sum(f \times p)}{\sum f}$$

$$= \frac{21640}{158} = 136.96$$

(b) A new frequency table is made from the results shown in the table in part (a).

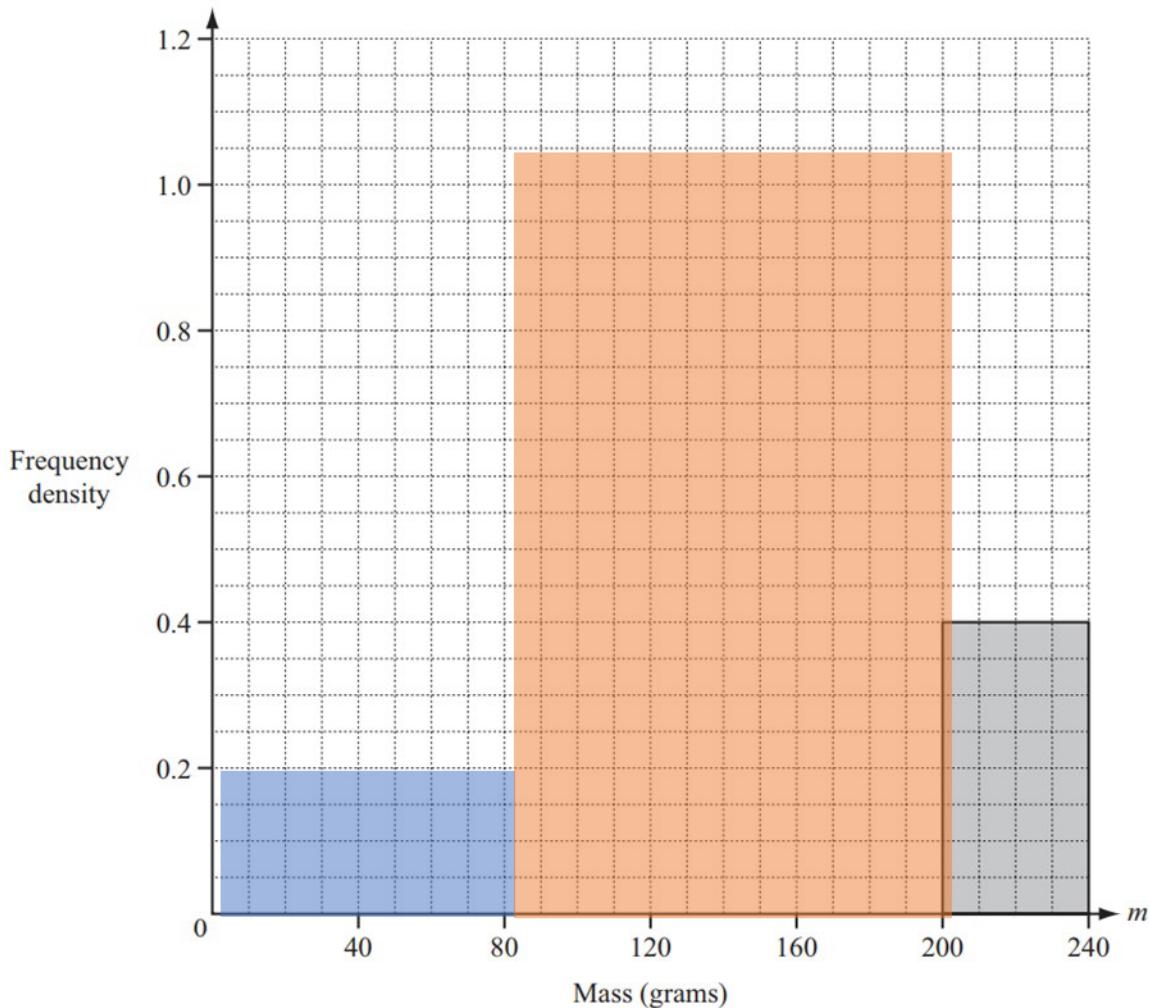
Mass (m grams)	Frequency
$0 < m \leq 80$	16
$80 < m \leq 200$	126
$200 < m \leq 240$	16

(i) Complete the table above.

[2]

(ii) On the grid opposite, complete the histogram to show the information in this new table.

[3]



(c) A bag contains 15 potatoes which have a mean mass of 136 g.

The farmer puts 3 potatoes which have a mean mass of 130 g into the bag.

[3]

Calculate the mean mass of all the potatoes in the bag.

We need the sum of all the potatoes masses divided by the number of potatoes.

The new mean is

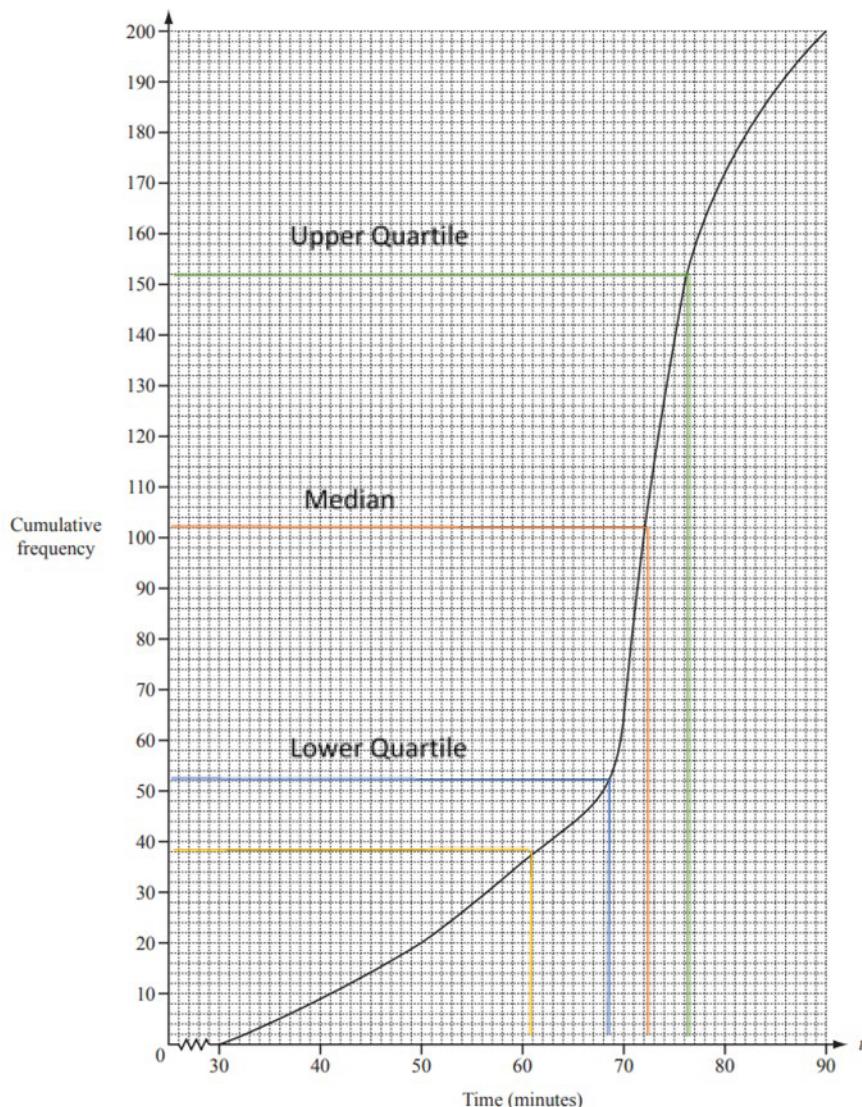
$$\frac{15 \times 136 + 3 \times 130}{15 + 3}$$

$$= 135$$

Question 6

200 students take a Mathematics examination.

The cumulative frequency diagram shows information about the times taken, t minutes, to complete the examination.



(a) Find

(i) the median,

[1]

The orange line

$$= 72$$

(ii) the lower quartile,

[1]

The blue line

$$= 68$$

- (iii) the inter-quartile range,

[1]

The upper quartile (green line) minus the lower quartile (blue line)

$$76 - 68$$

$$= 8$$

- (iv) the number of students who took more than 1 hour.

[2]

At the 60m mark the cumulative frequency is 36 (yellow line). Hence the number of students that took more than an hour is

$$200 - 36$$

$$= 164$$

- (b) (i) Use the cumulative frequency diagram to complete the grouped frequency table.

Time, <i>t</i> minutes	$30 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$	$60 < t \leq 70$	$70 < t \leq 80$	$80 < t \leq 90$
Frequency	9	11	16	28	108	28

[1]

- (ii) Calculate an estimate of the mean time taken by the 200 students to complete the examination.
 Show all your working.

[4]

Time, t minutes	Frequency, F	Midpoint, m	$f \times m$
$30 < t \leq 40$	9	35	315
$40 < t \leq 50$	11	45	495
$50 < t \leq 60$	16	55	880
$60 < t \leq 70$	28	65	1820
$70 < t \leq 80$	108	75	8100
$80 < t \leq 90$	28	85	2380
Sum	200		13990

Estimate of mean time is

$$\mu = \frac{\sum(f \times m)}{\sum f}$$

$$= \frac{13990}{200}$$

$$= 69.95$$

Question 7

- (a) In a football league a team is given 3 points for a win, 1 point for a draw and 0 points for a loss.

The table shows the 20 results for Athletico Cambridge.

Points	3	1	0
Frequency	10	3	7

- (i) Find the median and the mode.

[3]

Median is the average (mean) of the middle two values (the 10th and 11th entry in this case).

Mode is the most commonly occurring value.

$$\text{Median} = \frac{1 + 3}{2} = 2$$

$$\text{Mode} = 3$$

- (ii) Thomas wants to draw a pie chart using the information in the table.

Calculate the angle of the sector which shows the number of times Athletico Cambridge were given 1 point. [2]

We have that Athletico Cambridge were given 1 point 3 times out of 20.

Hence, as an angle, we would have

$$\frac{3}{20} \times 360$$

$$= 54$$

(b) Athletico Cambridge has 20 players.

The table shows information about the heights (h centimetres) of the players.

Height (h cm)	$170 < h \leq 180$	$180 < h \leq 190$	$190 < h \leq 200$
Frequency	5	12	3

Calculate an estimate of the mean height of the players.

[4]

Height (h cm)	Frequency (f)	Midpoint (M)	$f \times M$
$170 < h \leq 180$	5	175	875
$180 < h \leq 190$	12	185	2220
$190 < h \leq 200$	3	195	585
Sum	20		3680

Estimate of the mean is

$$\mu = \frac{\sum(f \times M)}{\sum f}$$

$$= \frac{3680}{20} = 184$$

Question 8

(a) The times, t seconds, for 200 people to solve a problem are shown in the table.

Time (t seconds)	Frequency
$0 < t \leq 20$	6
$20 < t \leq 40$	12
$40 < t \leq 50$	20
$50 < t \leq 60$	37
$60 < t \leq 70$	42
$70 < t \leq 80$	50
$80 < t \leq 90$	28
$90 < t \leq 100$	5

Calculate an estimate of the mean time.

[4]

To work out an estimate of the mean we initially need to assume that each frequency represents the middle value of its corresponding interval.

Therefore, we need to work out the middle values:

Time (t seconds)	Frequency	Middle value	Frequency x Middle value
$0 < t \leq 20$	6	10	60
$20 < t \leq 40$	12	30	360
$40 < t \leq 50$	20	45	900
$50 < t \leq 60$	37	55	2035
$60 < t \leq 70$	42	65	2730
$70 < t \leq 80$	50	75	3750
$80 < t \leq 90$	28	85	2380
$90 < t \leq 100$	5	95	475
Total	200	-	12960

Then, for each interval, we multiply the frequency by the middle value to work out the amount of time represented by that interval.

We add up all the values to obtain the total amount of time, 12960.

To work out the estimate we need to divide the total amount of time by the total number of people.

$$\text{Mean} = \frac{12960}{200}$$

$$\text{Mean} = 63.45$$

(b) (i) Complete the cumulative frequency table for this data. [2]

Time (t seconds)	$t \leq 20$	$t \leq 40$	$t \leq 50$	$t \leq 60$	$t \leq 70$	$t \leq 80$	$t \leq 90$	$t \leq 100$
Cumulative Frequency	6	18	38			167		

To work out the cumulative frequency for each inequality we need to add up the frequency for multiple intervals which in the end will represent the inequality.

For example:

$t \leq 20$ is represented by the interval $0 < t \leq 20$, with the frequency 6.

$t \leq 40$ is represented by the interval $0 < t \leq 20$ and the interval $20 < t \leq 40$, with the frequency $6 + 12 = 18$.

...

$t \leq 60$ is represented by the inequality $t \leq 50$ and the interval $50 < t \leq 60$, with the frequency $38 + 37 = 75$.

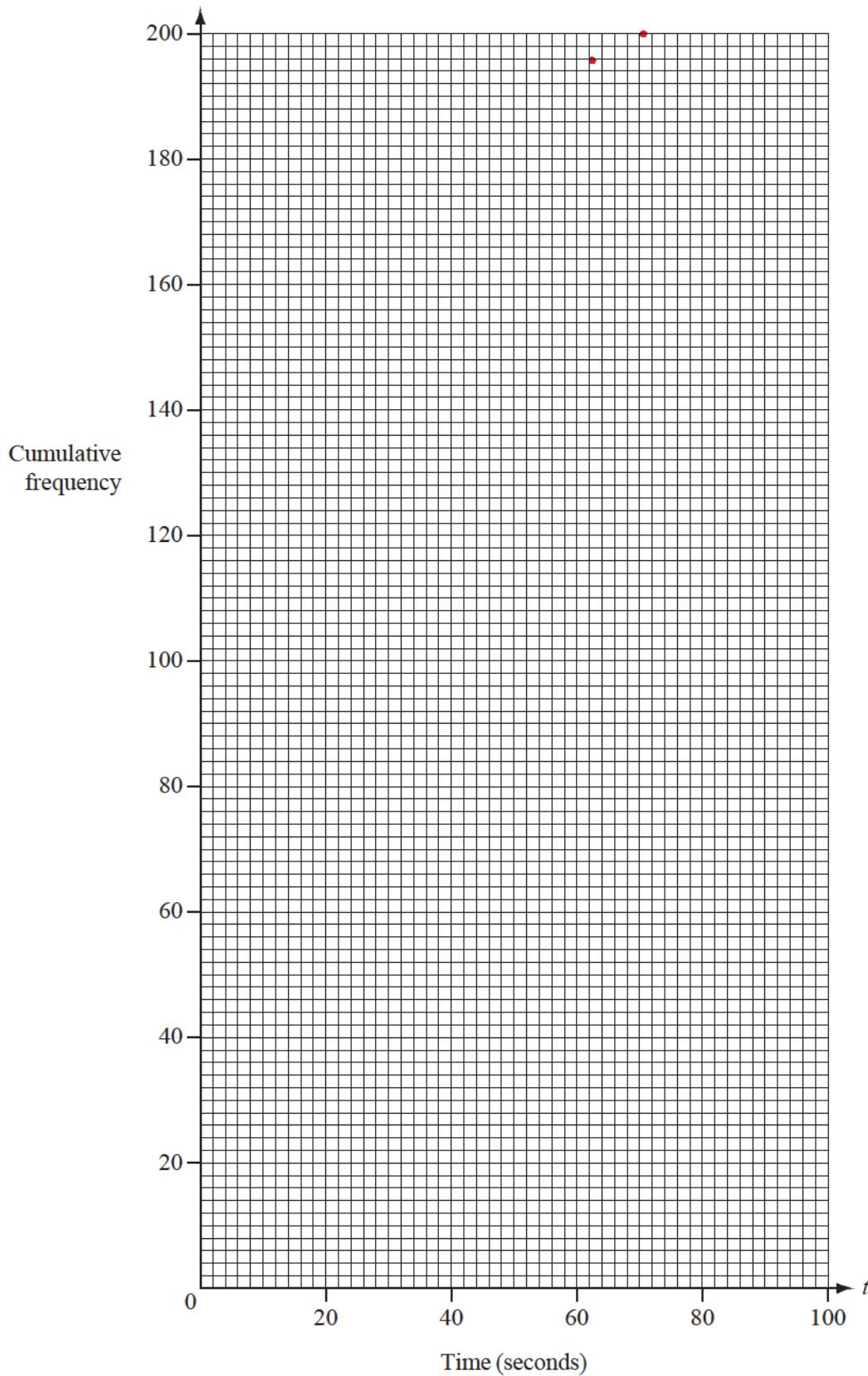
$t \leq 70$ is represented by the inequality $t \leq 60$ and the interval $60 < t \leq 70$, with the frequency $75 + 42 = 117$.

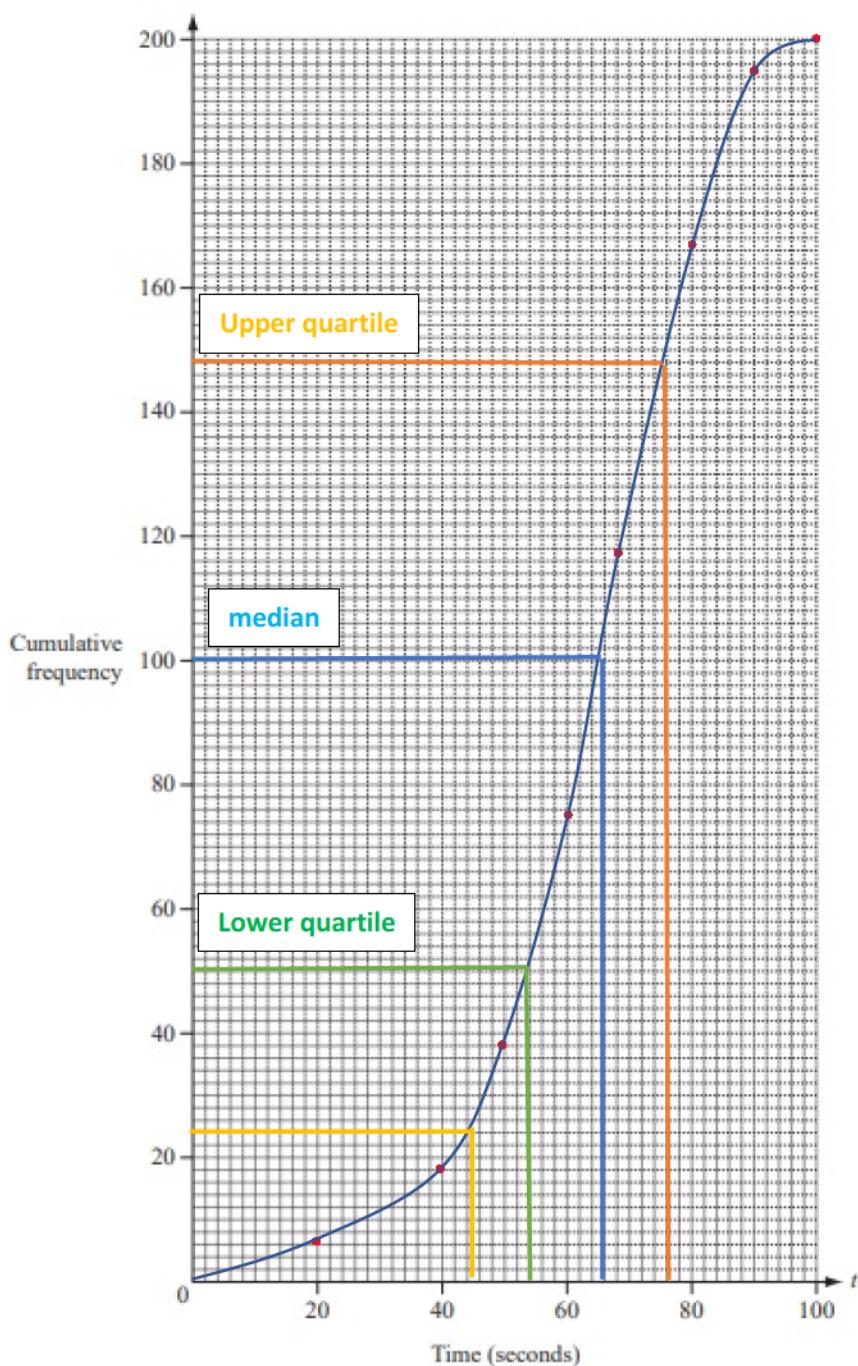
$t \leq 90$ is represented by the inequality $t \leq 80$ and the interval $80 < t \leq 90$, with the frequency $167 + 28 = 195$.

$t \leq 100$ is represented by the inequality $t \leq 90$ and the interval $90 < t \leq 100$, with the frequency $195 + 5 = 200$.

- (ii) Draw the cumulative frequency graph on the grid opposite to show this data.

[4]





- (c) Use your cumulative frequency graph to find

- (i) the median time,

[1]

The median will be the time corresponding to half of the number of people, 100.

This median is 65 s.

- (ii) the lower quartile,

[1]

The lower quartile will be the time corresponding to 25% of the number of people, 50.

This lower quartile is 52 s.

The upper quartile will be the time corresponding to 75% of the number of people, 150.

This upper quartile is 73 s.

- (iii) the inter-quartile range,

[1]

The interquartile range will be the difference between the upper and the lower quartile.

$$\text{IQR} = 73 \text{ s} - 52 \text{ s} = 21 \text{ s}$$

- (iv) how many people took between 65 and 75 seconds to solve the problem,

[1]

Similarly, the number of people who took between 65 and 75 seconds will be the number of people who took 75 seconds or less minus the number of people who took 65 second or less.

$$148 - 100$$

$$= 48 \text{ people who took between 65 s and 75s}$$

- (v) how many people took longer than 45 seconds to solve the problem. [2]

Similarly, the number of people who took longer than 45 seconds will be the total frequency, 200, minus the number of people who took less than 45s. Using the graph, we can see that the number of people who took 45 seconds or less is 24.

$$200 - 24$$

= 176 people who took longer

Statistics

Difficulty: Medium

Model Answers 2

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Medium
Booklet	Model Answers 2

Time allowed: 117 minutes

Score: /102

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

The times, t minutes, taken for 200 students to cycle one kilometre are shown in the table.

Time (t minutes)	$0 < t \leq 2$	$2 < t \leq 3$	$3 < t \leq 4$	$4 < t \leq 8$
Frequency	24	68	72	36

- (a) Write down the class interval that contains the median. [1]

The middle value of the number of students is 100.

By adding up the frequencies, we can observe that the frequency of 100 will be reached for the time interval

$3 < t \leq 4$.

$$(24 + 68 < 100, 24 + 68 + 72 > 100)$$

- (b) Calculate an estimate of the mean.

Show all your working.

[4]

To work out an estimate of the mean we initially need to assume that each

frequency represents the middle value of its corresponding interval.

Therefore, we need to work out the middle values:

Time (t minutes)	$0 < t \leq 2$	$2 < t \leq 3$	$3 < t \leq 4$	$4 < t \leq 8$	Total
Frequency	24	68	72	36	200
Mid-point	1	2.5	3.5	6	-
Mid-point x Frequency	24	170	252	216	662

Then, for each interval, we multiply the frequency by the middle value to work out the amount of time represented by that interval.

We add up all the values to obtain the total time, 662 minutes.

To work out the estimate we need to divide the total amount of time by the total number of students.

$$\text{Mean} = \frac{662}{200}$$

$$\text{Mean} = 3.31$$

- (c) (i) Use the information in the table opposite to complete the cumulative frequency table. [1]

Time (t minutes)	$t \leq 2$	$t \leq 3$	$t \leq 4$	$t \leq 8$
Cumulative frequency	24			200

To work out the cumulative frequency for each inequality we need to add up the frequency for multiple intervals which in the end will represent the inequality.

$t \leq 2$ is represented by the interval $0 < t \leq 2$, with the frequency 24.

$t \leq 3$ is represented by the interval $0 < t \leq 2$ and the interval $2 < t \leq 3$, with the frequency $24 + 68$

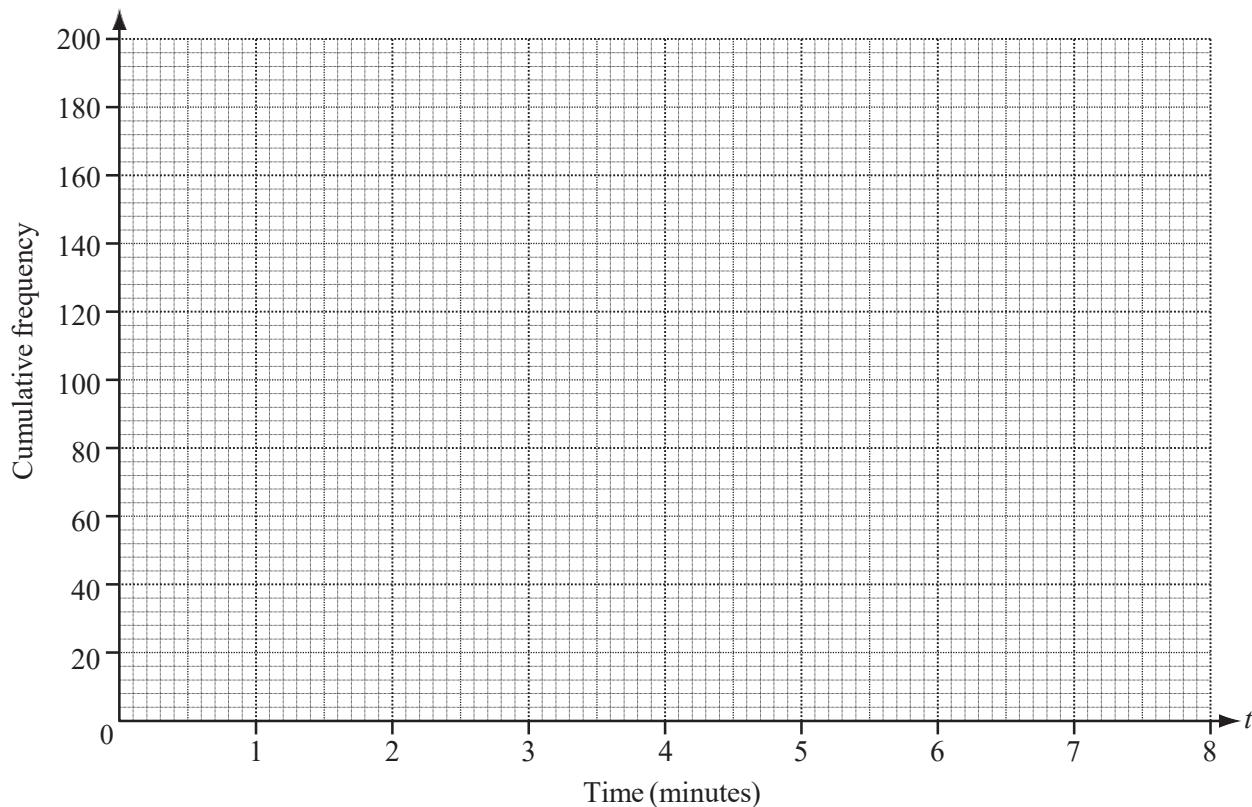
= 92.

$t \leq 4$ is represented by the inequality $t \leq 3$ and the interval $3 < t \leq 4$, with the frequency $92 + 72$

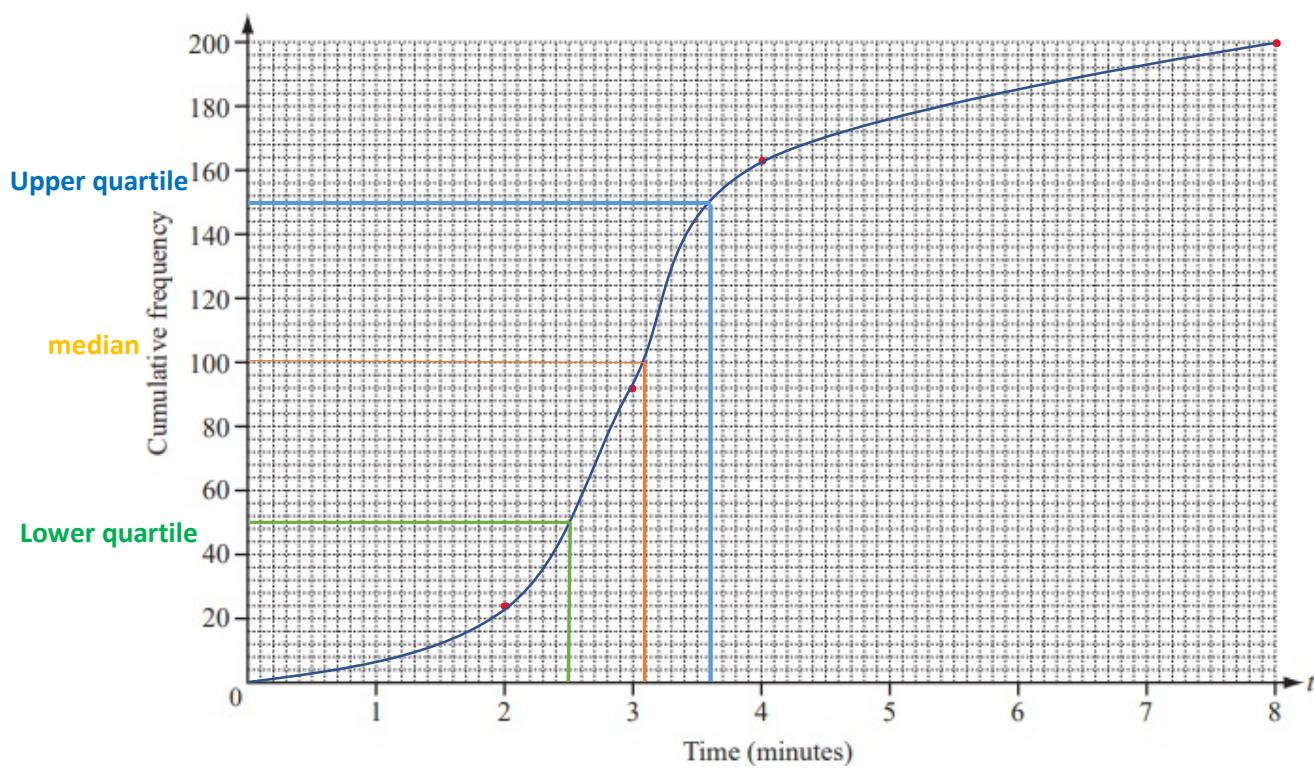
= 164.

- (ii) On the grid, draw a cumulative frequency diagram.

[3]



We plot the values from the cumulative frequency table obtained and connect them to form a curve.



- (iii) Use your diagram to find the median, the lower quartile and the inter-quartile range. [3]

The median will be the time corresponding to half of the number of students, 100.

This median is 3.1 minutes.

Similarly, the lower quartile will be the time corresponding to 25% of the number of students, 50.

This lower quartile is 2.5 minutes.

Similarly, the upper quartile will be the time corresponding to 75% of the number of students, 150.

This upper quartile is 3.6 minutes.

The interquartile range is the difference between the upper and the lower quartiles.

The IQR = 3.6 minutes – 2.5 minutes

IQR = 1.1 minutes

Question 2

Time (t mins)	$0 < t \leq 20$	$20 < t \leq 35$	$35 < t \leq 45$	$45 < t \leq 55$	$55 < t \leq 70$	$70 < t \leq 80$
Frequency	6	15	19	37	53	20

The table shows the times taken, in minutes, by 150 students to complete their homework on one day.

- (a) (i) In which interval is the median time? [1]

Median is the interval with 50th percentile. There were 150 students in total, so we are summing the frequencies on the bottom of the table until we get to 75 (half the maximum).

When we include intervals up to t=45, we only have 40 students, but when we include the interval

$$45 < t \leq 55$$

the number of students we are including increases to 77, hence we are last the 50th percentile.

Therefore the interval with the median time:

$$\mathbf{45 < t \leq 55}$$

- (ii) Using the mid-interval values 10, 27.5,calculate an estimate of the mean time. [3]

The mean value is estimated by assuming that the time taken by each student in a interval is equal to the mid-interval value (i.e. 10min, 27.5min, 40min, 50min, 62.5min, 75min).

We sum the products of the mid-time and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

mean

$$= \frac{10\text{min} \times 6 + 27.5\text{min} \times 15 + 40\text{min} \times 19 + 50\text{min} \times 37 + 62.5\text{min} \times 53 + 75\text{min} \times 20}{150 \text{ students}}$$

$$\mathbf{mean = 52.6\text{min}}$$

(b) (i) Complete the table of cumulative frequencies. [2]

Time (t mins)	$t \leq 20$	$t \leq 35$	$t \leq 45$	$t \leq 55$	$t \leq 70$	$t \leq 80$
Cumulative frequency	6	21				

For example, to get the cumulative frequency for $t \leq 45$, we sum the corresponding frequencies (up to $35 < t \leq 45$) from the previous table.

$$t \leq 45 = 6 + 15 + 19 = 40$$

The cumulative frequency for $t \leq 45$ is 40.

We use the same method to complete the cumulative frequency table.

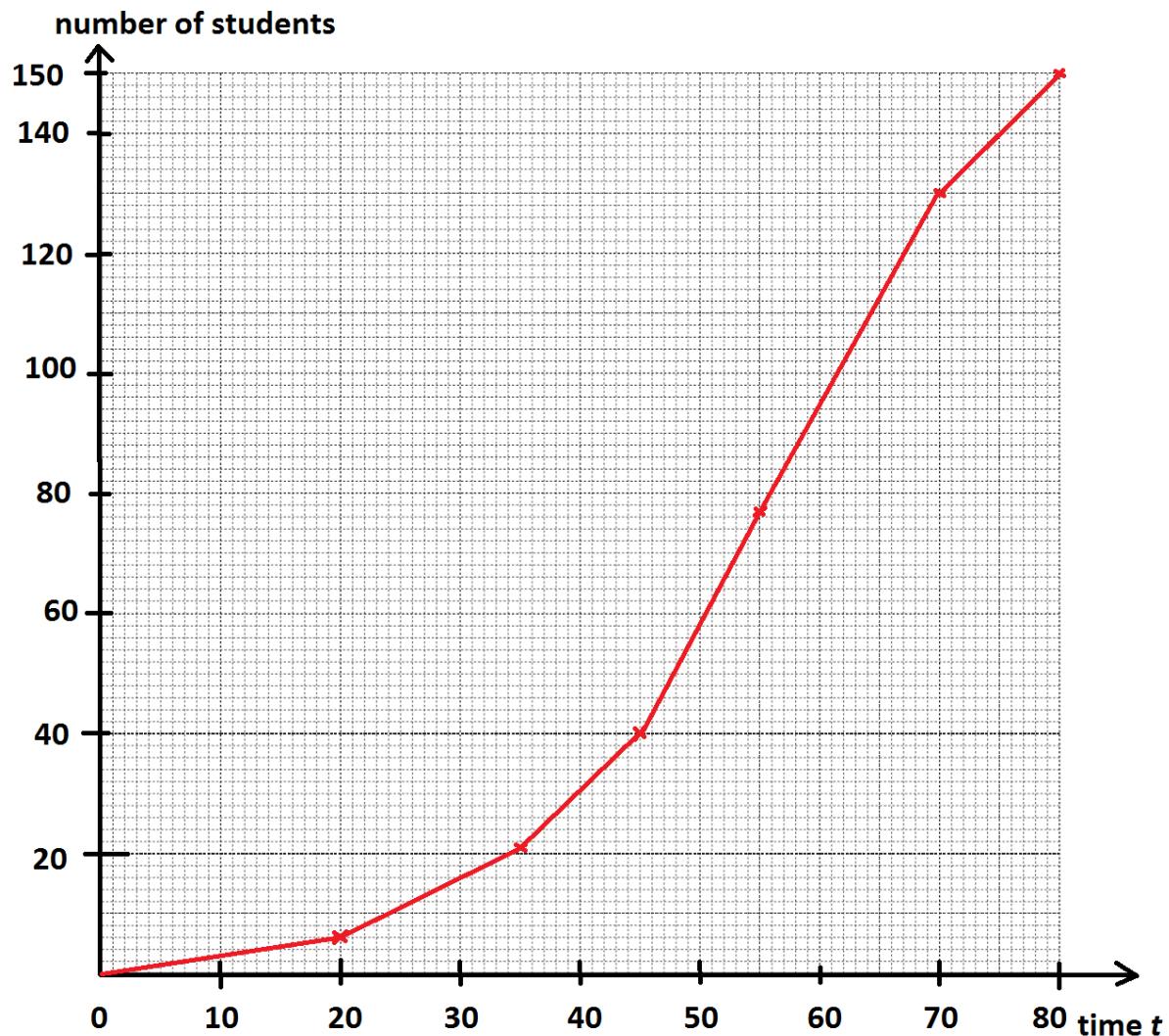
$$t \leq 45 = 40; \quad t \leq 55 = 77; \quad t \leq 70 = 130; \quad t \leq 80 = 150$$

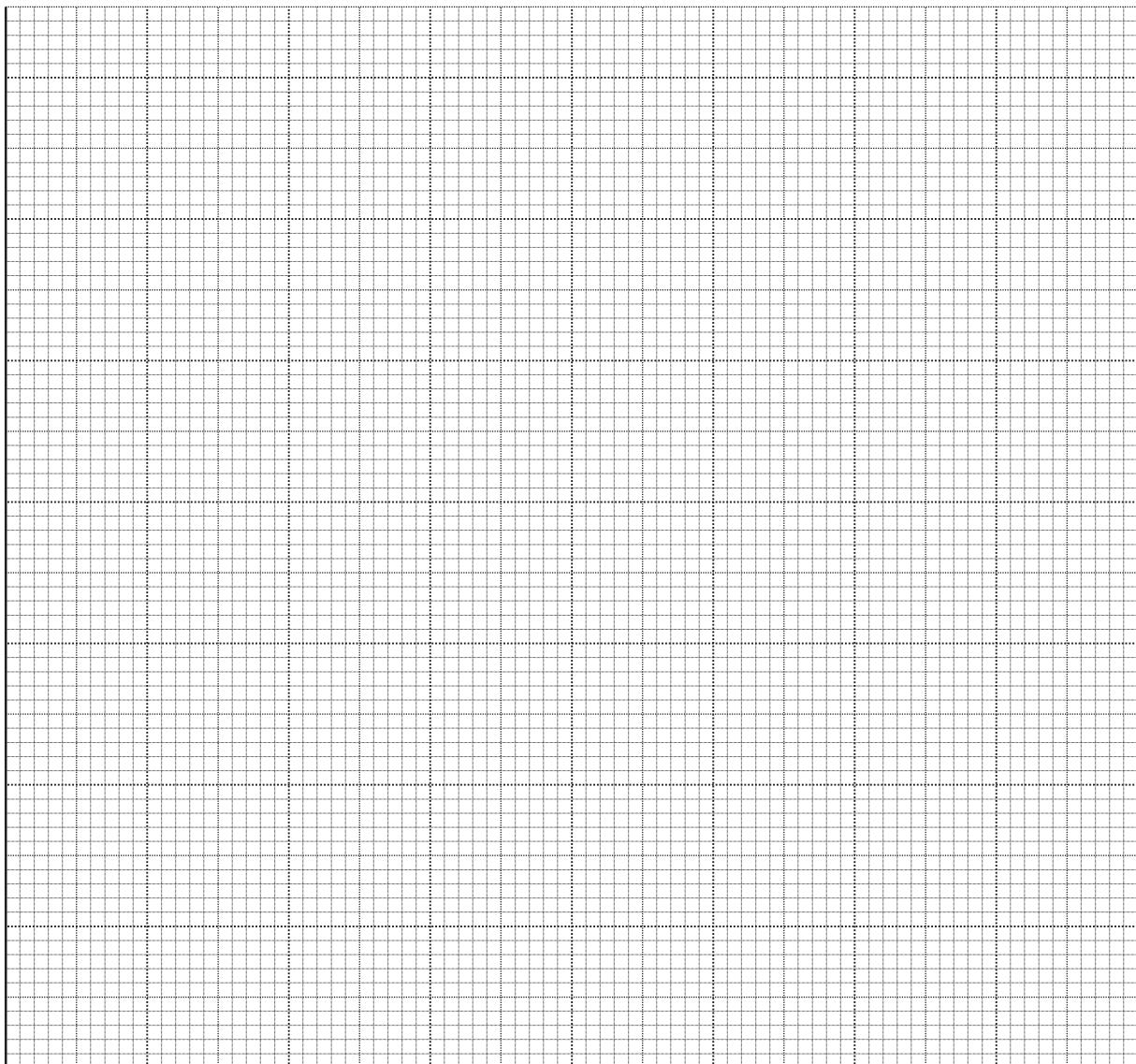
Time (t mins)	$t \leq 20$	$t \leq 35$	$t \leq 45$	$t \leq 55$	$t \leq 70$	$t \leq 80$
Cumulative frequency	6	21	40	77	130	150

- (ii) On the grid, label the horizontal axis from 0 to 80, using the scale 1 cm represents 5 minutes and the vertical axis from 0 to 150, using the scale 1 cm represents 10 students.

Draw a cumulative frequency diagram to show this information. [5]

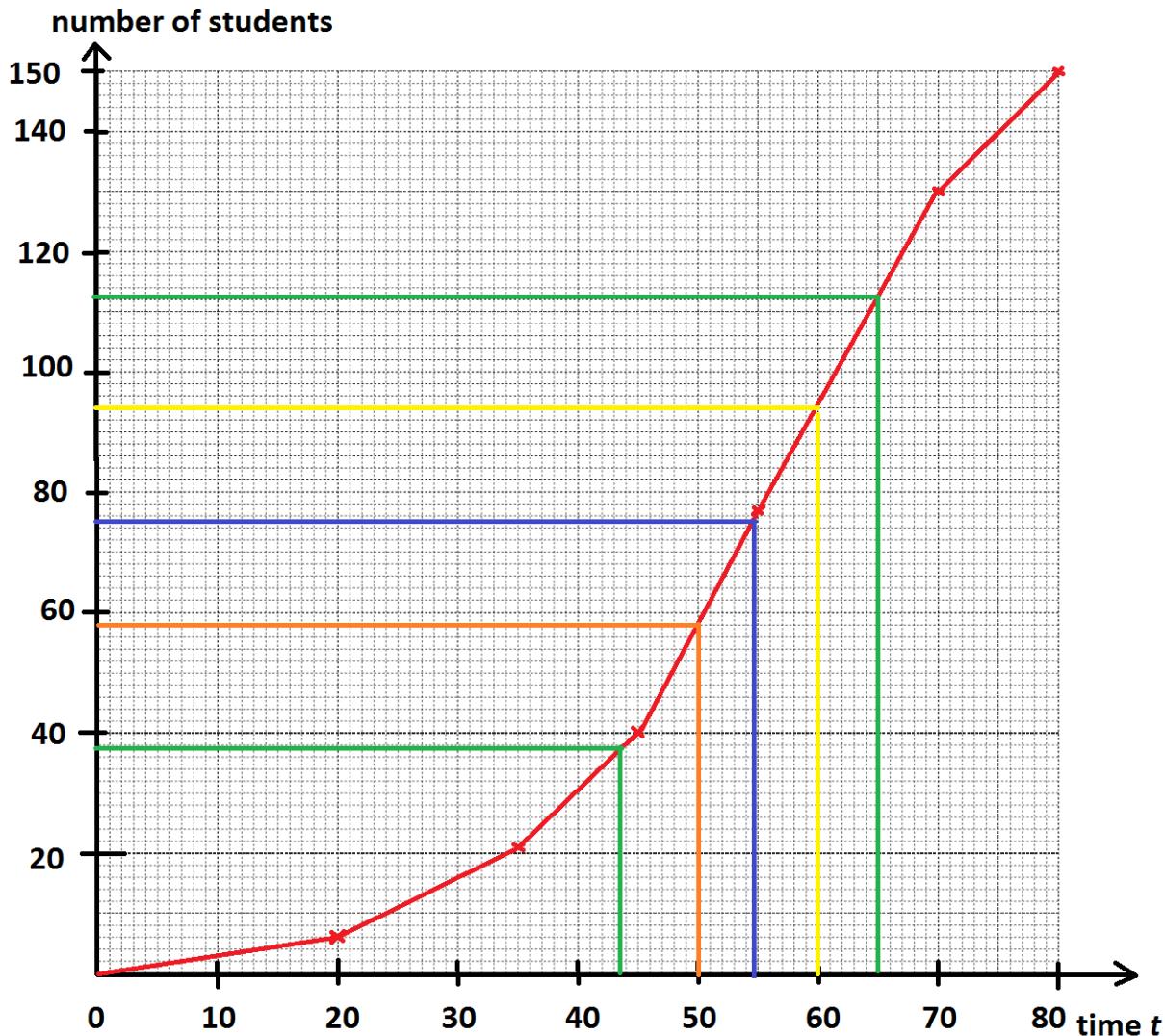
We set up a grid and plot the points from the table above on the grid and connect them with lines.





- (c) Use your graph to estimate

We will refer to the graph below throughout part c)



- (i) the median time, [1]

The median is the 50th percentile. As there are 150 students in total, it is the time taken of 75th student (blue lines). From the graph, this is

54.5 minutes.

- (ii) the inter-quartile range, [2]

The inter-quartile range is the difference between the upper and the lower quartile.

The lines representing these values are green.

The lower quartile is the value of student who is 25% of 150, which is 37.5.

This is 43.5 min.

The upper quartile is the value of student who is 75% of 150, which is 112.5.

This is 65 min.

The inter-quartile range is therefore

$$\mathbf{65\text{min} - 43.5\text{min} = 21.5 \text{ minutes.}}$$

- (iii) the number of students whose time was in the range $50 < t \leq 60$, [1]

To find the number of student who finished the test in 50 min or less, we follow the orange line.

$$(< 50) = 48 \text{ students}$$

Similarly, we follow the yellow line to find the number of students who took 60 min or less.

$$(< 60) = 94 \text{ students}$$

We subtract the two values to find the number of students in the range $50 < t \leq 60$

$$(50 < t \leq 60) = (< 60) - (< 50)$$

$$(50 < t \leq 60) = 94 - 48$$

$$(50 < t \leq 60)$$

$$= \mathbf{46 \text{ students}}$$

- (iv) the probability, as a fraction, that a student, chosen at random, took longer than 50 minutes, [2]

We know that there were 48 students who took less than 50 minutes to complete the test. Subtract this number from the total number of students (150) to get the number of students who took longer than 50 minutes.

$$(> 50) = 150 - (< 50)$$

$$(> 50) = 150 - 48$$

$$(> 50) = 102 \text{ students}$$

There are 150 students in total, so the probability that a random student took longer than 50 minutes to complete the test is:

$$\text{probability } (> 50) = \frac{(> 50)}{\text{total students}}$$

$$\text{probability } (> 50) = \frac{102}{150}$$

- (v) the probability, as a fraction, that two students, chosen at random, both took longer than 50 minutes.

[2]

We already know the probability that one of them is chosen. Now we are choosing the second one.

There are only 149 students left as one was already chosen and also there are only 101 students who took 50 minutes or less because one of them was chosen already as well.

Hence the probability that the second is chosen given we already have the first one:

$$\text{probability second } (> 50) = \frac{101}{149}$$

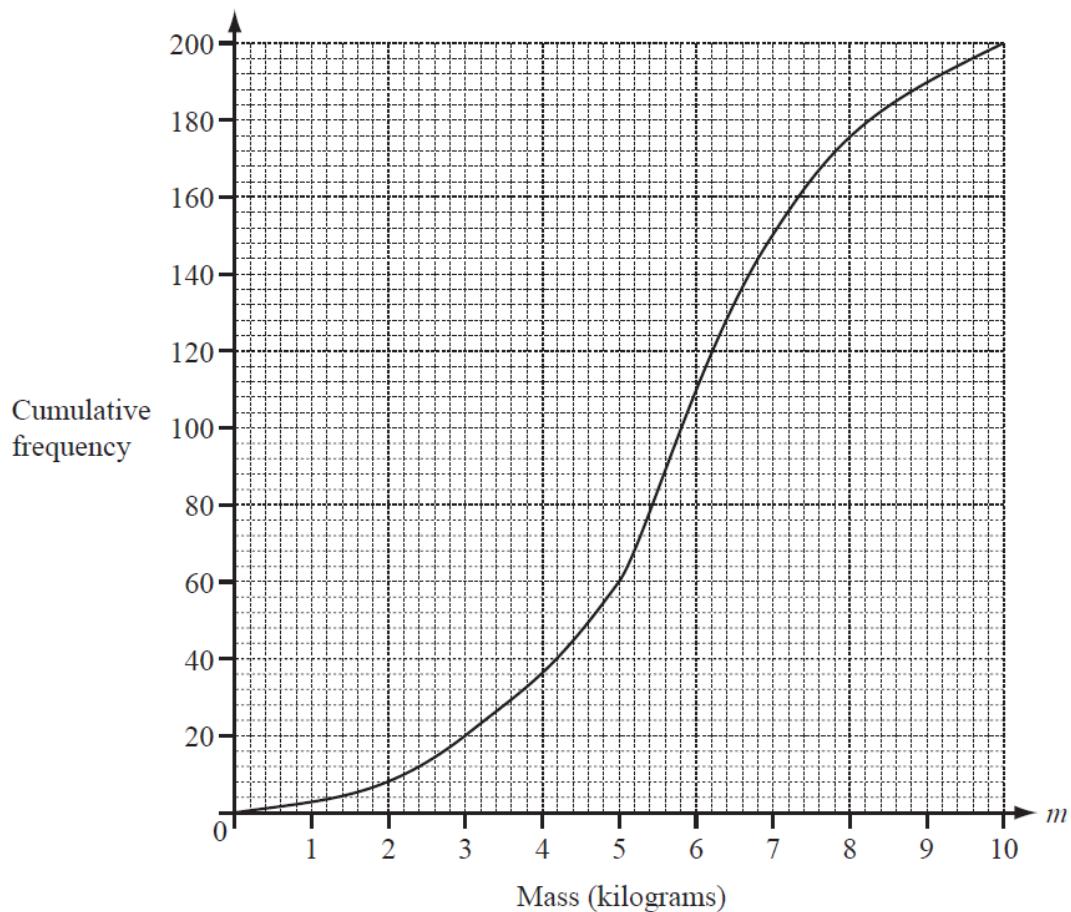
Probabilities multiply to get the total probability that two students are chosen.

$$\text{probability: two students } (> 50) = \frac{102}{150} \times \frac{101}{149}$$

Multiply and simplify the fractions.

$$\text{probability: two students } (> 50) = \frac{1717}{3725}$$

Question 3

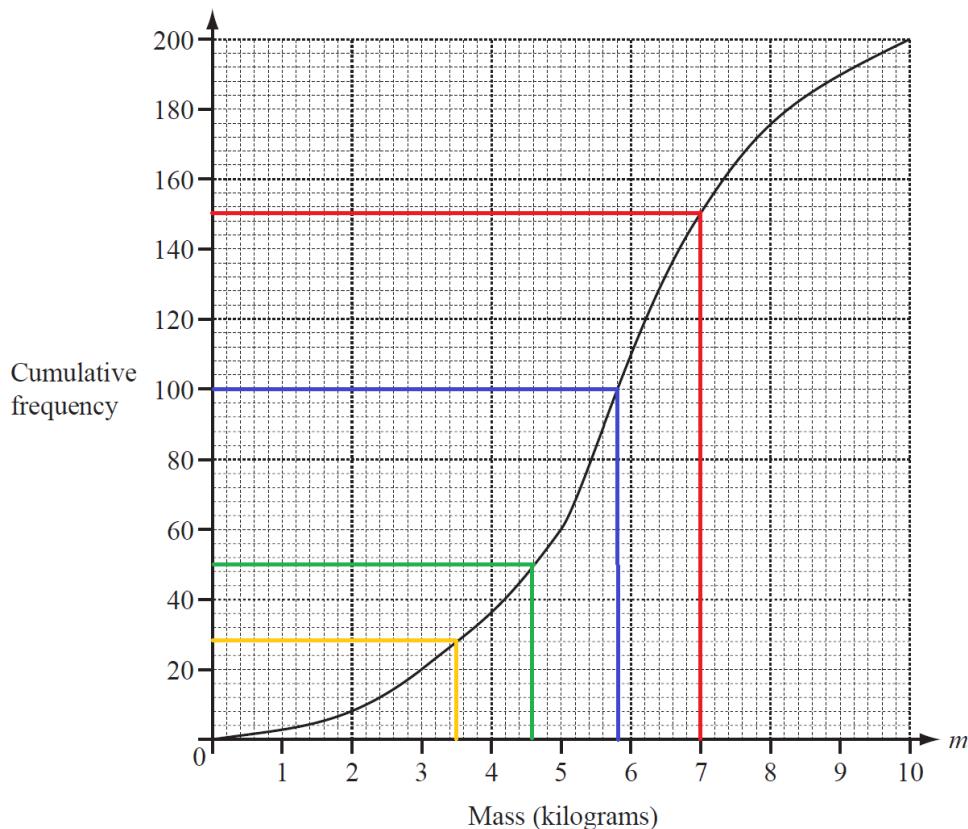


The masses of 200 parcels are recorded.

The results are shown in the cumulative frequency diagram above.

(a) Find

We will refer to the graph below throughout part a)



(i) the median,

[1]

The median is the 50th percentile. As there are 200 parcels in total, it is the mass of 100th parcel (blue lines). From the graph, this is

5.8 kg.

(ii) the lower quartile,

[1]

The lower quartile is the mass of parcel which is 25% of 200, which is 50 (marked green).

This is 4.6kg.

- (iii) the inter-quartile range, [1]

The inter-quartile range is the difference between the upper and the lower quartile. We already have the lower quartile from part ii), so we only need to find the upper quartile.

The upper quartile is the mass of parcel which is 75% of 200, which is 150 (marked red).

This is 7kg.

The inter-quartile range is therefore

$$7\text{kg} - 4.6\text{kg} = 2.4\text{kg}.$$

- (iv) the number of parcels with a mass greater than 3.5 kg. [2]

To find the number of parcels which mass is 3.5 kg or less, we follow the orange line.

$$(\leq 3.5\text{kg}) = 28 \text{ parcels}$$

We subtract the value from the total number of parcels (200) to get the number of parcel with mass greater than 3.5 kg.

$$(> 3.5\text{kg}) = \text{total} - (\leq 3.5\text{kg})$$

$$(> 3.5\text{kg}) = 200 \text{ parcels} - 28 \text{ parcels}$$

$$(> 3.5\text{kg})$$

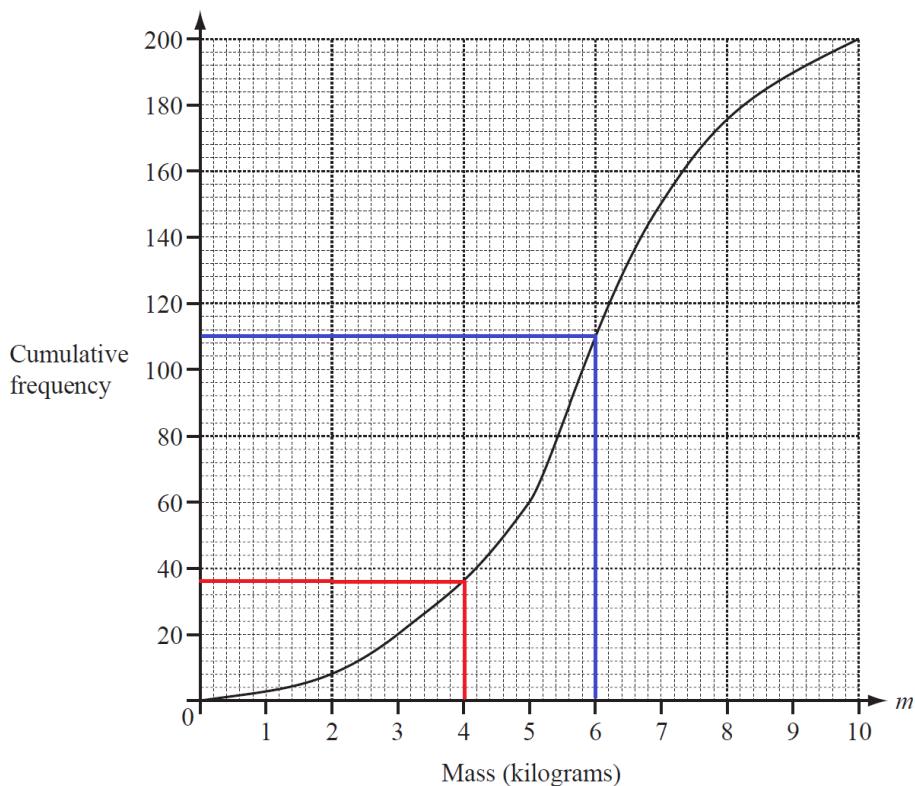
$$= 172 \text{ parcels}$$

- (b) (i) Use the information from the cumulative frequency diagram to complete the grouped frequency table. [2]

Mass (m) kg	$0 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 7$	$7 < m \leq 10$
Frequency	36			50

To complete the table, we need to find the number of parcels with masses below 4kg and 6kg (follow red, blue lines on the graph below respectively).

In principle, we could require knowing this value for 7kg as well to determine the third data point in the table, but since we know that there are 200 parcels in total, we can use this constrain instead.



$$(m \leq 4\text{kg}) = 36 \text{ parcels}$$

$$(m \leq 6\text{kg}) = 110 \text{ parcels}$$

By subtracting these values, we find the number of parcels with masses in the given interval:

$$(4\text{kg} < m \leq 6\text{kg}) = (m \leq 6\text{kg}) - (m \leq 4\text{kg})$$

$$(4\text{kg} < m \leq 6\text{kg}) = 110 \text{ parcels} - 36 \text{ parcels}$$

$$\mathbf{(4kg < m \leq 6kg) = 74 parcels}$$

As mentioned before, the fourth data point in the table can be calculated using the fact that there must be 200 parcels in total.

$$(6\text{kg} < m \leq 7\text{kg}) = \text{total} - \text{already known}$$

$$(6\text{kg} < m \leq 7\text{kg}) = 200 - (36 + 74 + 50)$$

$$\mathbf{(6kg < m \leq 7kg) = 40 parcels}$$

Mass (m) kg	$0 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 7$	$7 < m \leq 10$
Frequency	36	74	40	50

- (ii) Use the grouped frequency table to calculate an estimate of the mean. [4]

The mean value is estimated by assuming that the mass of each parcel in a given interval is equal to the mid-interval value (i.e. 2kg, 5kg, 6.5kg, 8.5kg).

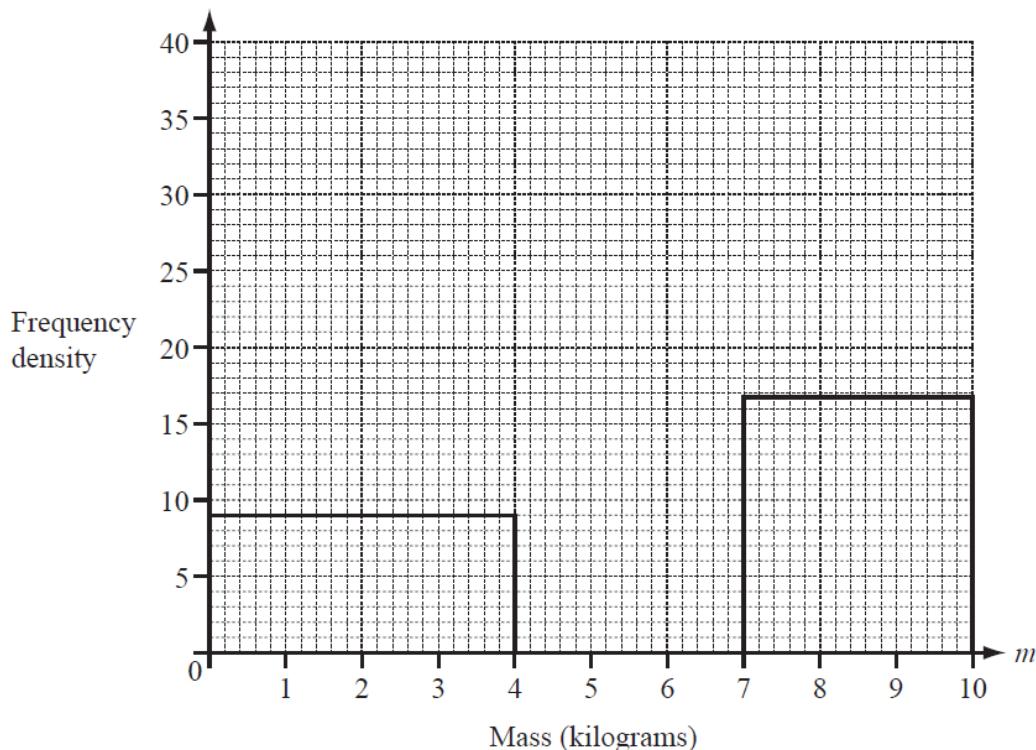
We sum the products of the mid-mass and the frequency of given group and then divide by the total number of parcels to get the mean estimate. The mean is therefore:

$$\text{mean} = \frac{2\text{kg} \times 36 + 5\text{kg} \times 74 + 6.5\text{kg} \times 40 + 8.5\text{kg} \times 50}{200 \text{ parcels}}$$

$$\text{mean} = 5.635 \text{ kg}$$

- (iii) Complete the frequency density table and use it to complete the histogram. [4]

Mass (m) kg	$0 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 7$	$7 < m \leq 10$
Frequency density	9			16.7



The frequency density is calculated using the previous table as the ratio of the frequency and the range of the corresponding interval (difference between the upper and the lower bounds).

$$\text{frequency density} = \frac{\text{frequency}}{\text{range of values}}$$

For the second interval, the frequency is 70 and the range of values is 2 kg (allowing values from 4kg to 6kg so range is the difference of these two values). The frequency density:

$$\text{frequency density } (4\text{kg} < m \leq 6\text{kg}) = \frac{74}{2}$$

$$\text{frequency density } (4\text{kg} < m \leq 6\text{kg}) = 37$$

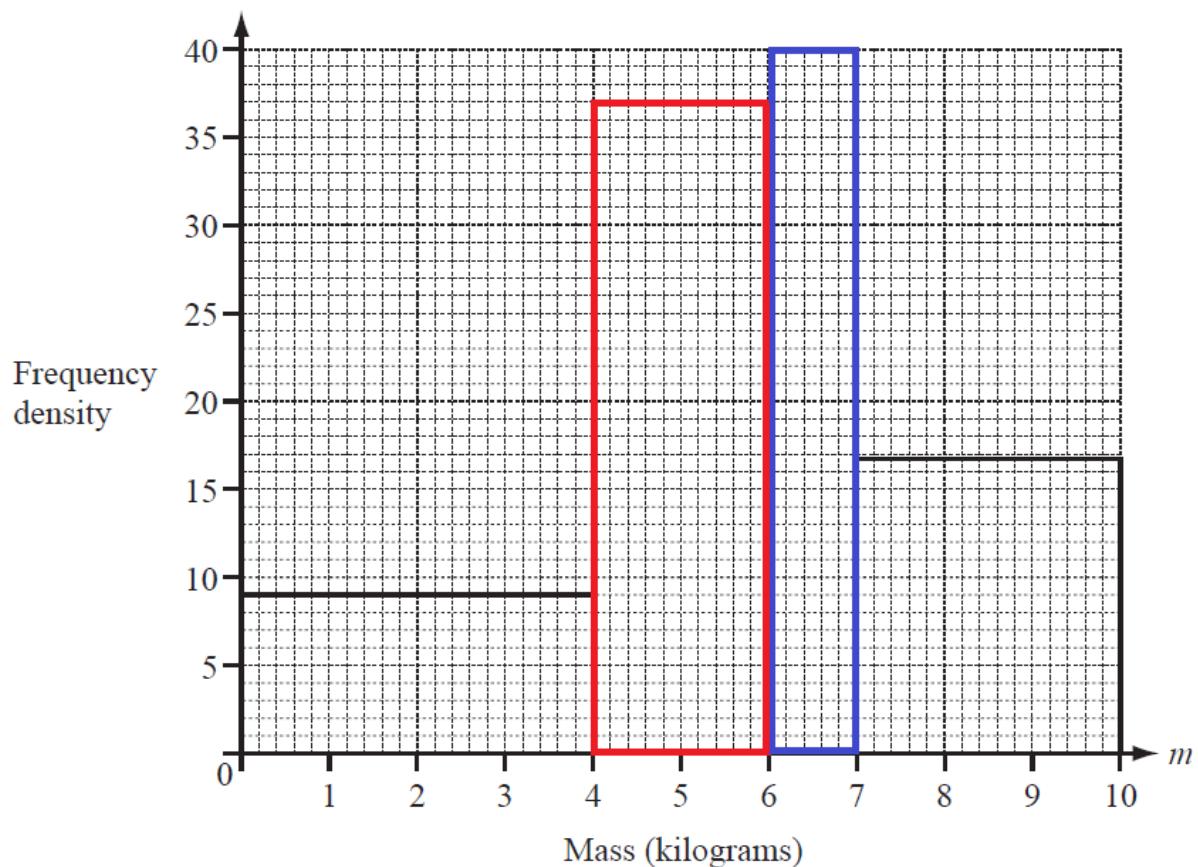
Similarly for the third interval, the frequency is 40 and the range of accepted values if 1 kg.

$$\text{frequency density } (6\text{kg} < m \leq 7\text{kg}) = \frac{40}{1}$$

$$\text{frequency density } (6\text{kg} < m \leq 7\text{kg}) = 40$$

Mass (m) kg	$0 < m \leq 4$	$4 < m \leq 6$	$6 < m \leq 7$	$7 < m \leq 10$
Frequency density	9	37	40	16.7

We plot the frequency density into the histogram, making sure that the correct value is assigned to the corresponding mass. [Height of the histogram changes when the interval changes!](#)

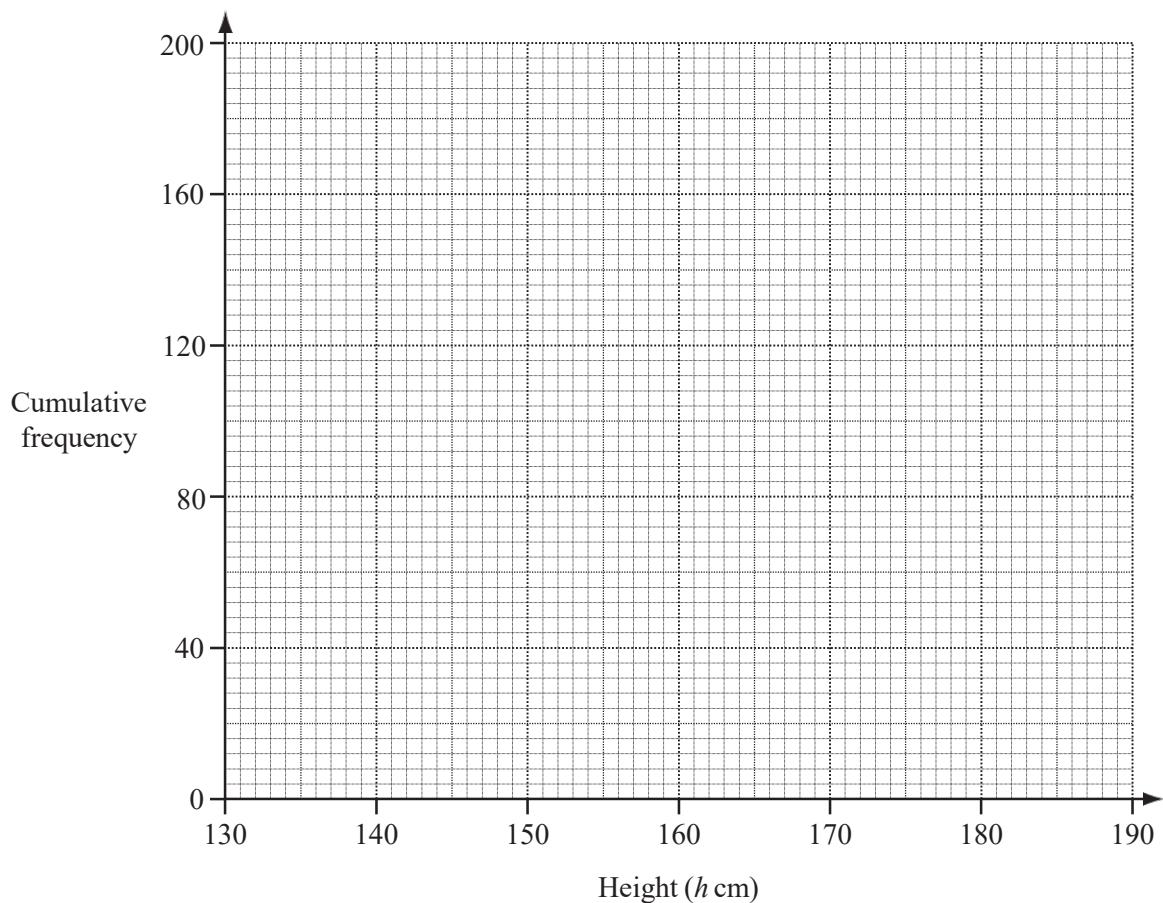


Question 4

The cumulative frequency table shows the distribution of heights, h centimetres, of 200 students.

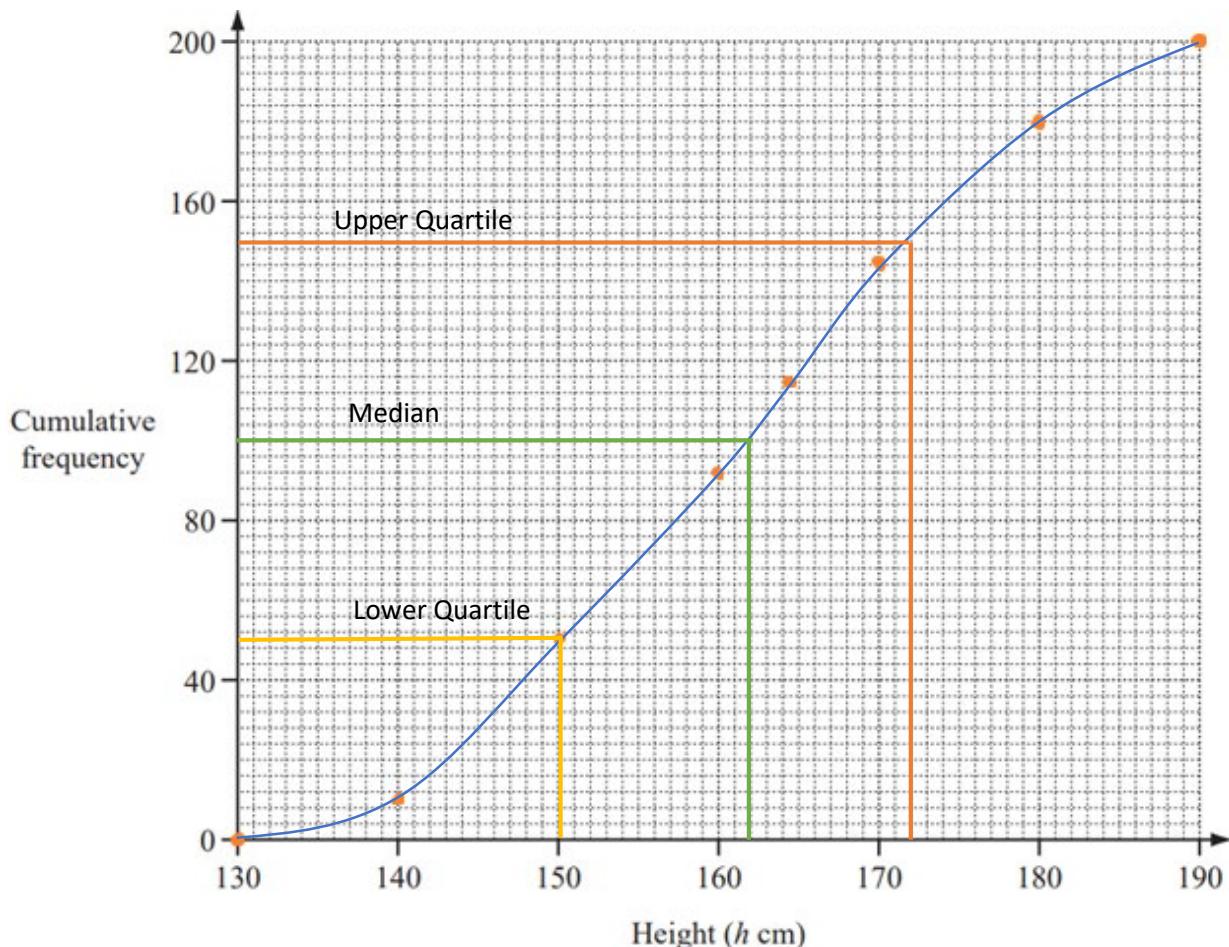
Height (h cm)	≤ 130	≤ 140	≤ 150	≤ 160	≤ 165	≤ 170	≤ 180	≤ 190
Cumulative frequency	0	10	50	95	115	145	180	200

(a) Draw a cumulative frequency diagram to show the information in the table.



[4]

The blue curve (with orange plots).



(b) Use your diagram to find

- (i) the median,

[1]

The green line

162

- (ii) the upper quartile,

[1]

The orange line

172

(iii) the interquartile range.

[1]

The upper quartile (orange line) minus the lower quartile (yellow line)

$$172 - 150$$

$$= 22$$

(c) (i) One of the 200 students is chosen at random.

Use the table to find the probability that the height of this student is greater than 170 cm.

Give your answer as a fraction.

[1]

145 students are less than or equal to 170 cm.**Therefore $200 - 145 = 55$ students are greater than 170 cm.**

$$P(h > 170 \text{ cm}) = \frac{55}{200}$$

$$= \frac{11}{40}$$

(ii) One of the 200 students is chosen at random and then a second student is chosen at random from the remaining students.

Calculate the probability that one has a height greater than 170 cm and the other has a height of 140 cm or less.

Give your answer as a fraction.

[3]

The probability of the second student being less than 140 cm**without replacement is**

$$\frac{10}{199}$$

We multiply the two probabilities together and multiply by 2 (because we can pick either student first) to get

$$2 \times \frac{11}{40} \times \frac{10}{199}$$

$$= \frac{11}{398}$$

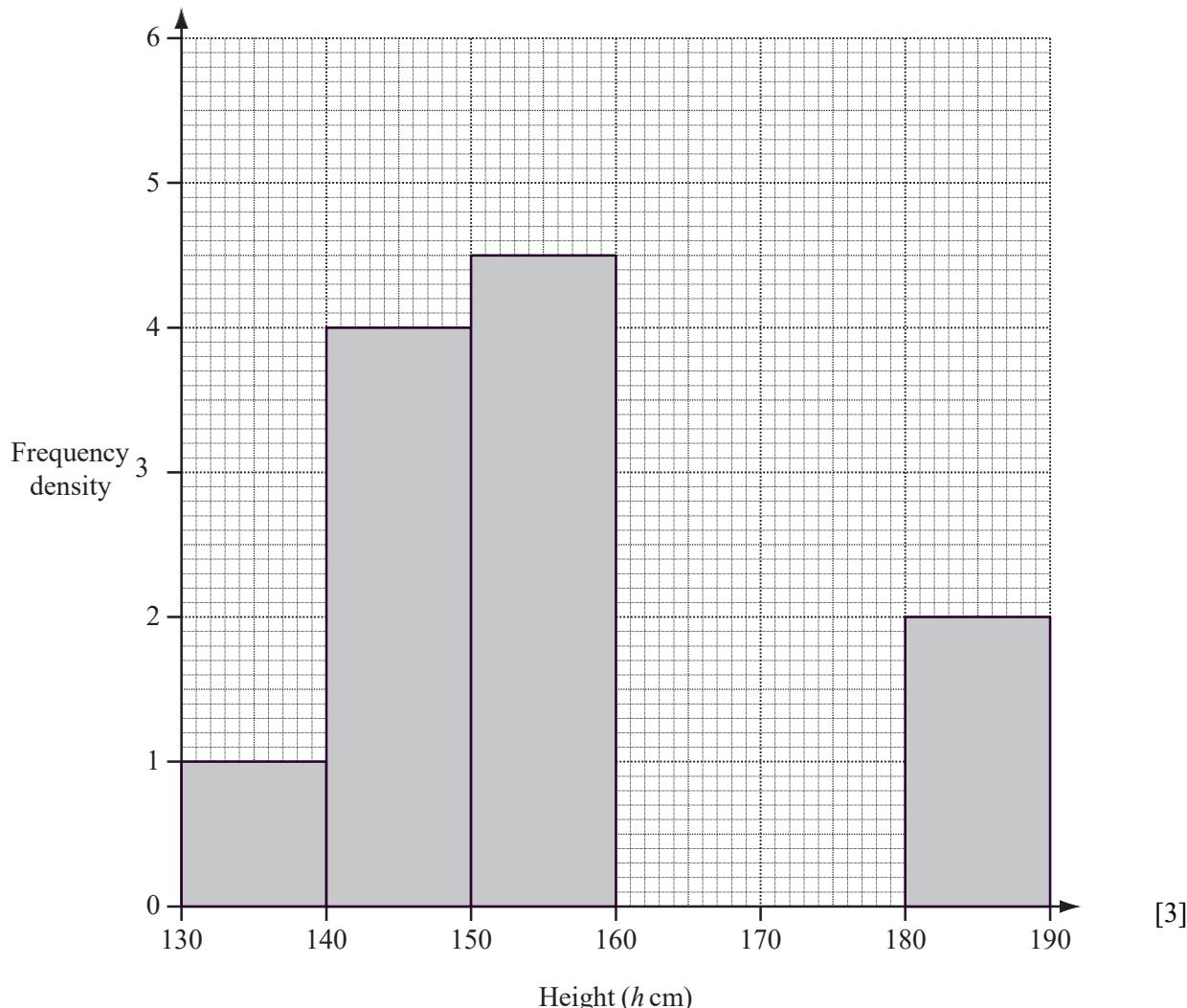
- (d) (i) Complete this frequency table which shows the distribution of the heights of the 200 students.

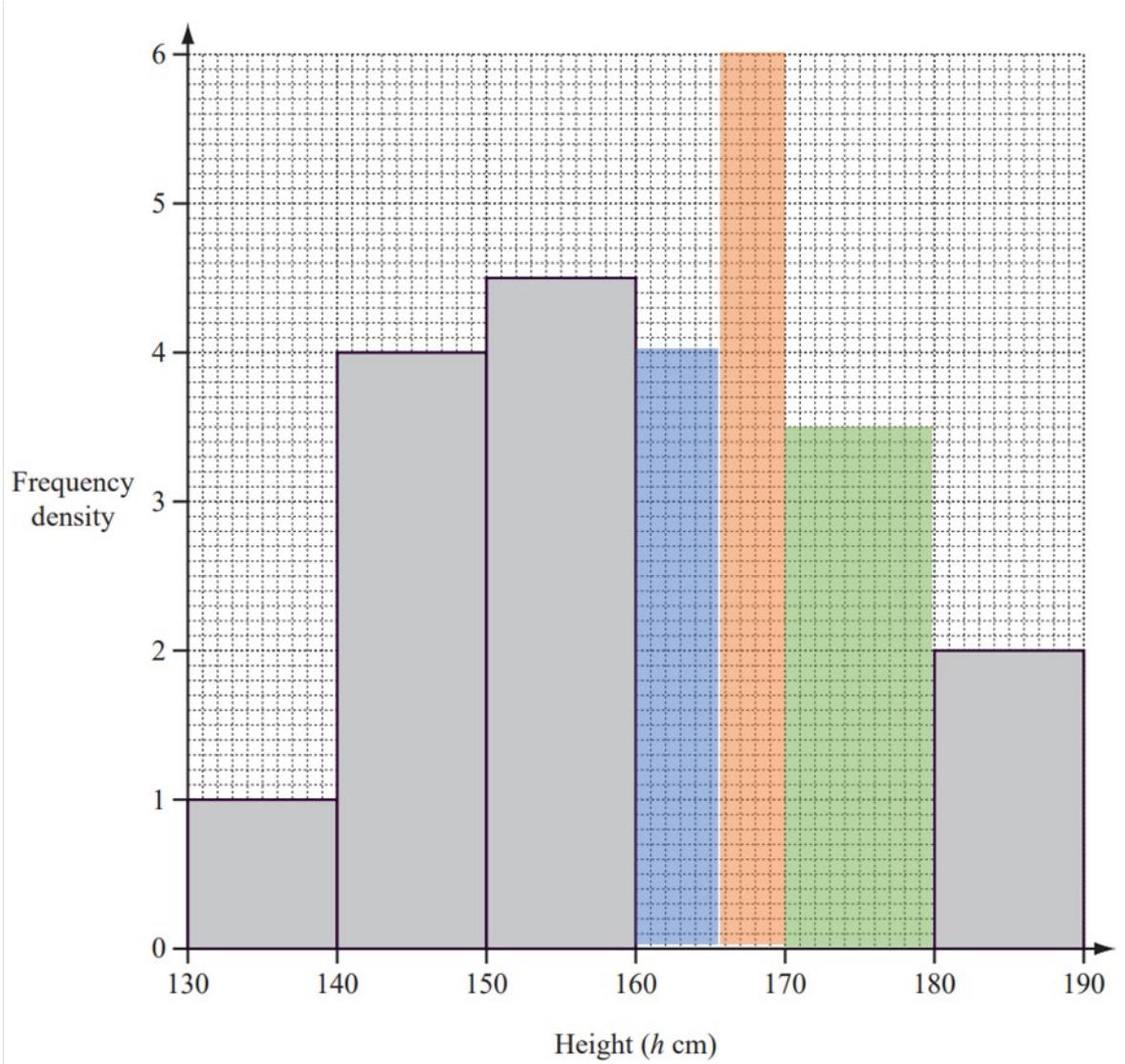
Height (h cm)	$130 < h \leq 140$	$140 < h \leq 150$	$150 < h \leq 160$	$160 < h \leq 165$	$165 < h \leq 170$	$170 < h \leq 180$	$180 < h \leq 190$
Frequency	10	40	45	20			

[2]

Height (h cm)	Frequency
$130 < h \leq 140$	10
$140 < h \leq 150$	40
$150 < h \leq 160$	45
$160 < h \leq 165$	20
$165 < h \leq 170$	30
$170 < h \leq 180$	35
$180 < h \leq 190$	20

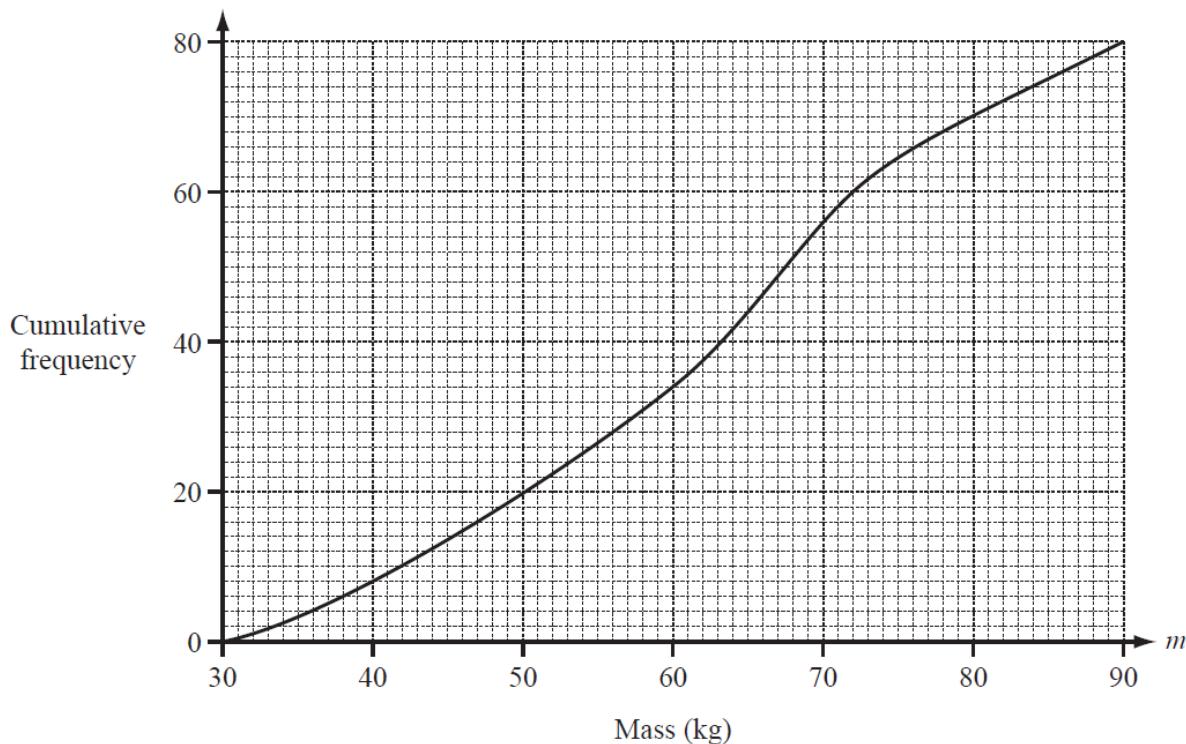
- (ii) Complete this histogram to show the distribution of the heights of the 200 students.





Question 5

80 boys each had their mass, m kilograms, recorded.
The cumulative frequency diagram shows the results.



(a) Find

(i) the median,

[1]

The orange line

63

(ii) the lower quartile,

[1]

The blue line

50

(iii) the interquartile range.

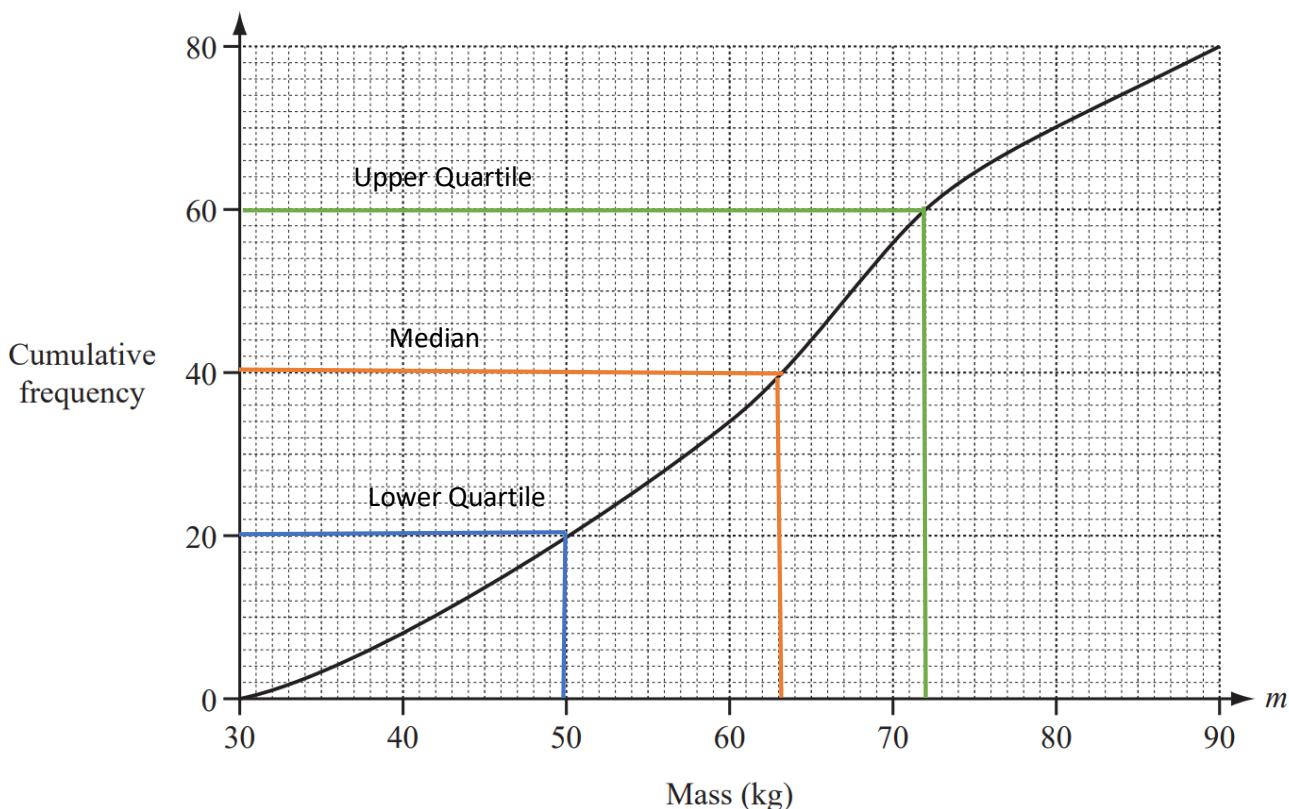
[1]

The upper quartile (green line) minus the lower quartile

(blue line)

$$72 - 50$$

$$= 22$$



(b) How many boys had a mass greater than 60kg?

[2]

We read off from the graph that there are 34 boys with

mass less than or equal to 60, hence there are

$$80 - 34$$

$$= 46$$

boys with mass greater than 60.

- (c) (i) Use the cumulative frequency graph to complete this frequency table.

Mass, m	Frequency
$30 < m \leq 40$	8
$40 < m \leq 50$	
$50 < m \leq 60$	14
$60 < m \leq 70$	22
$70 < m \leq 80$	
$80 < m \leq 90$	10

[2]

Mass, m	Frequency, f	Midpoint, M	$M \times f$
$30 < m \leq 40$	8	35	280
$40 < m \leq 50$	12	45	540
$50 < m \leq 60$	14	55	770
$60 < m \leq 70$	22	65	1430
$70 < m \leq 80$	14	75	1050
$80 < m \leq 90$	10	85	850
SUM	80		4920

- (ii) Calculate an estimate of the mean mass.

[4]

Estimate of the mean mass is

$$\mu = \frac{\sum(f \times m)}{\sum f}$$

$$= \frac{4920}{80}$$

$$= 61.5$$

Question 6

40 students are asked about the number of people in their families.

The table shows the results.

Number of people in family	2	3	4	5	6	7
Frequency	1	1	17	12	6	3

(a) Find

(i) the mode,

[1]

4

(ii) the median,

[1]

5

(iii) the mean.

[3]

Multiply the family number by the frequency and sum those numbers

$$2 \times 1 + 3 \times 1 + 4 \times 17 + 5 \times 12 + 6 \times 6 + 7 \times 3$$

$$= 190$$

Now divide by number of students

$$190 \div 40$$

$$= 4.75$$

(b) Another n students are asked about the number of people in their families.

The mean for these n students is 3.

Find, in terms of n , an expression for the mean number for all $(40 + n)$ students.

[2]

Let this new mean be μ .

We add the frequency multiplied by the mean to the top and add the frequency to the bottom.

$$\mu = \frac{190 + 3n}{40 + n}$$

Question 7

The masses of 60 potatoes are measured.

The table shows the results.

Mass (m grams)	$10 < m \leq 20$	$20 < m \leq 40$	$40 < m \leq 50$
Frequency	10	30	20

(a) Calculate an estimate of the mean.

[4]

Mass	Frequency	Midpoint	$f \times midpoint$
$10 < m \leq 20$	10	15	150
$20 < m \leq 40$	30	30	900
$40 < m \leq 50$	20	45	900
Sum	60		1950

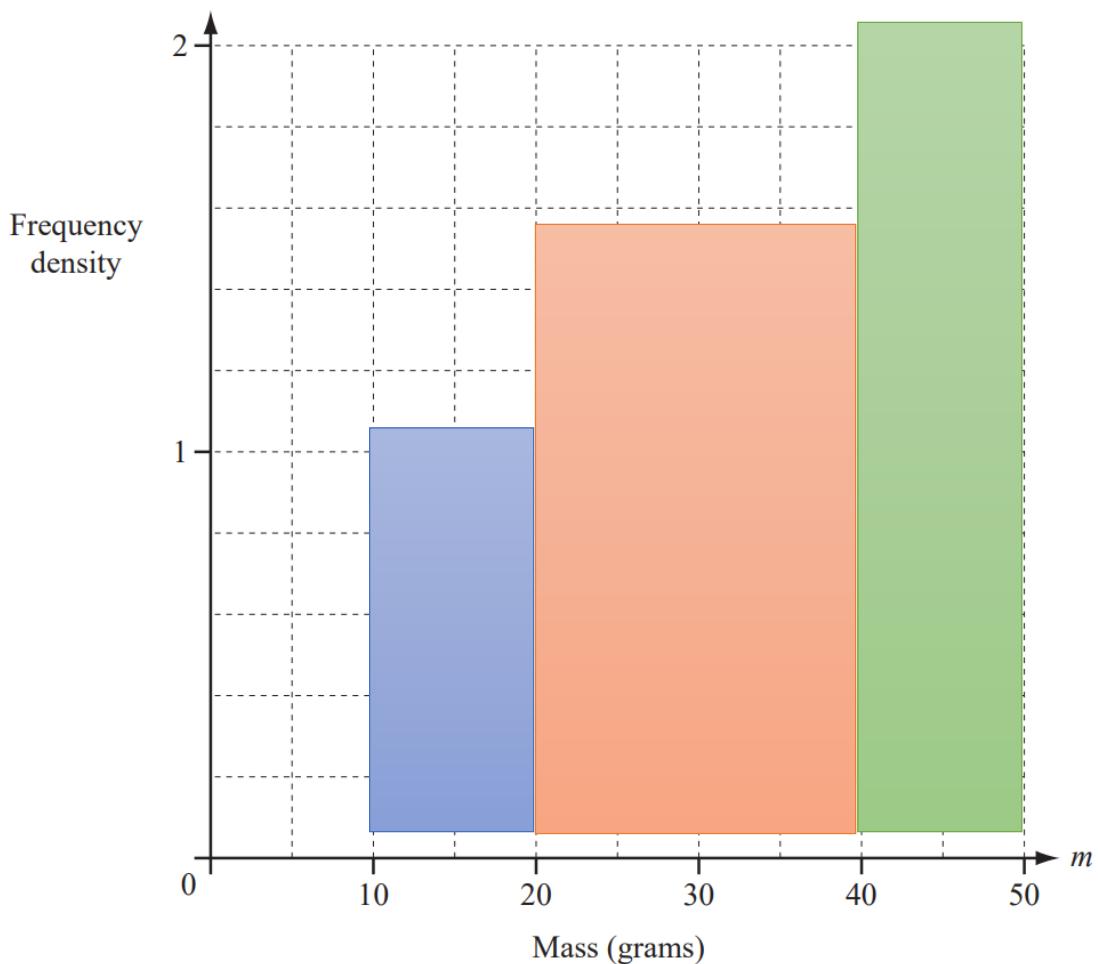
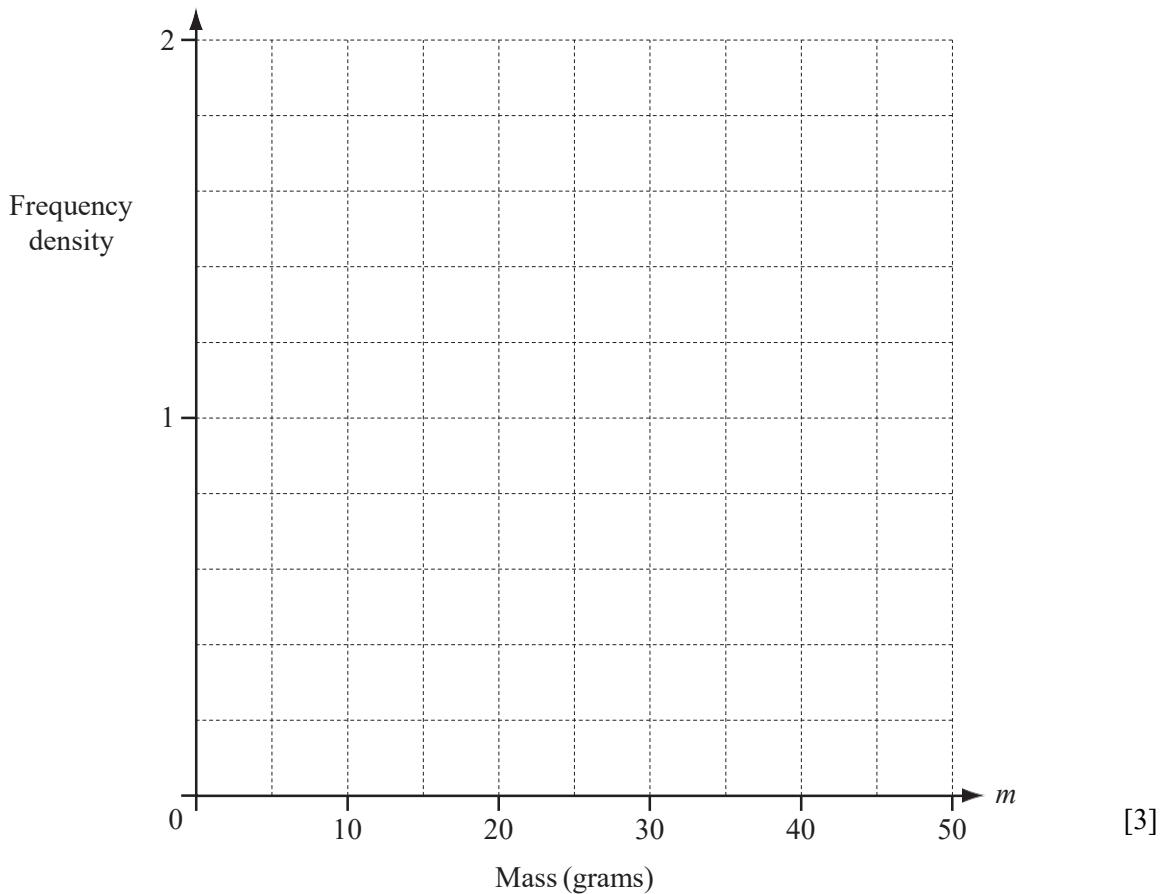
Estimate of the mean is

$$\mu = \frac{\sum(f \times midpoint)}{\sum f}$$

$$= \frac{1950}{60}$$

$$= 32.5$$

(b) On the grid, draw an accurate histogram to show the information in the table.



Question 8

200 students were asked how many hours they exercise each week.

The table shows the results.

Time (t hours)	$0 < t \leq 5$	$5 < t \leq 10$	$10 < t \leq 15$	$15 < t \leq 20$	$20 < t \leq 25$	$25 < t \leq 30$	$30 < t \leq 35$	$35 < t \leq 40$
Number of students	12	15	23	30	40	35	25	20

(a) Calculate an estimate of the mean.

[4]

Time (t hours)	Number of students	Midpoint	$f \times midpoint$
$0 < t \leq 5$	12	2.5	30
$5 < t \leq 10$	15	7.5	112.5
$10 < t \leq 15$	23	12.5	287.5
$15 < t \leq 20$	30	17.5	525
$20 < t \leq 25$	40	22.5	900
$25 < t \leq 30$	35	27.5	962.5
$30 < t \leq 35$	25	32.5	812.5
$35 < t \leq 40$	20	37.5	750
Sum		200	4380

Estimate of the mean is

$$\mu = \frac{\sum(f \times m)}{\sum f}$$

$$= \frac{4380}{200}$$

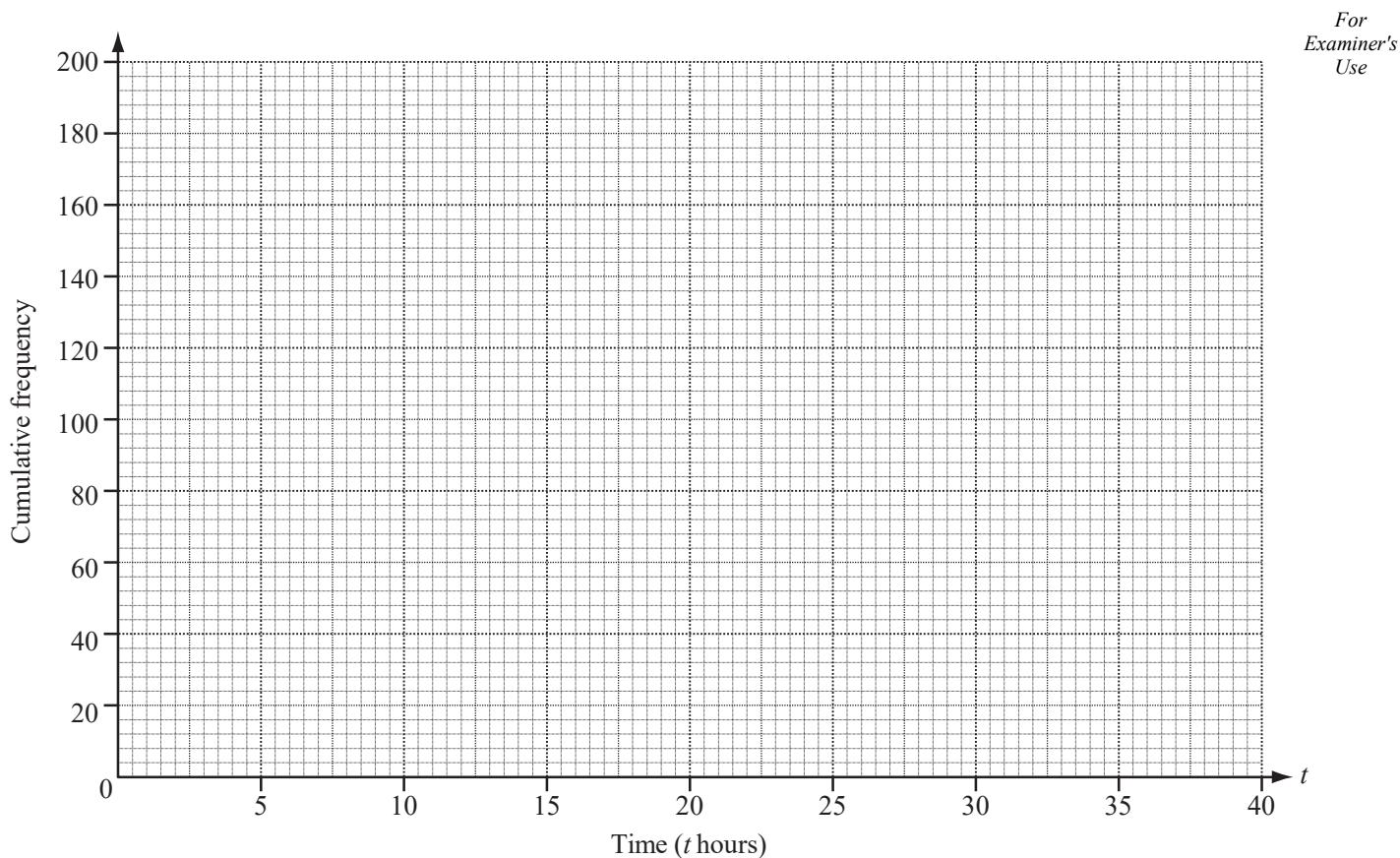
$$= 21.9$$

(b) Use the information in the table above to complete the cumulative frequency table.

Time (t hours)	$t \leq 5$	$t \leq 10$	$t \leq 15$	$t \leq 20$	$t \leq 25$	$t \leq 30$	$t \leq 35$	$t \leq 40$
Cumulative frequency	12	27	50	80	120			200

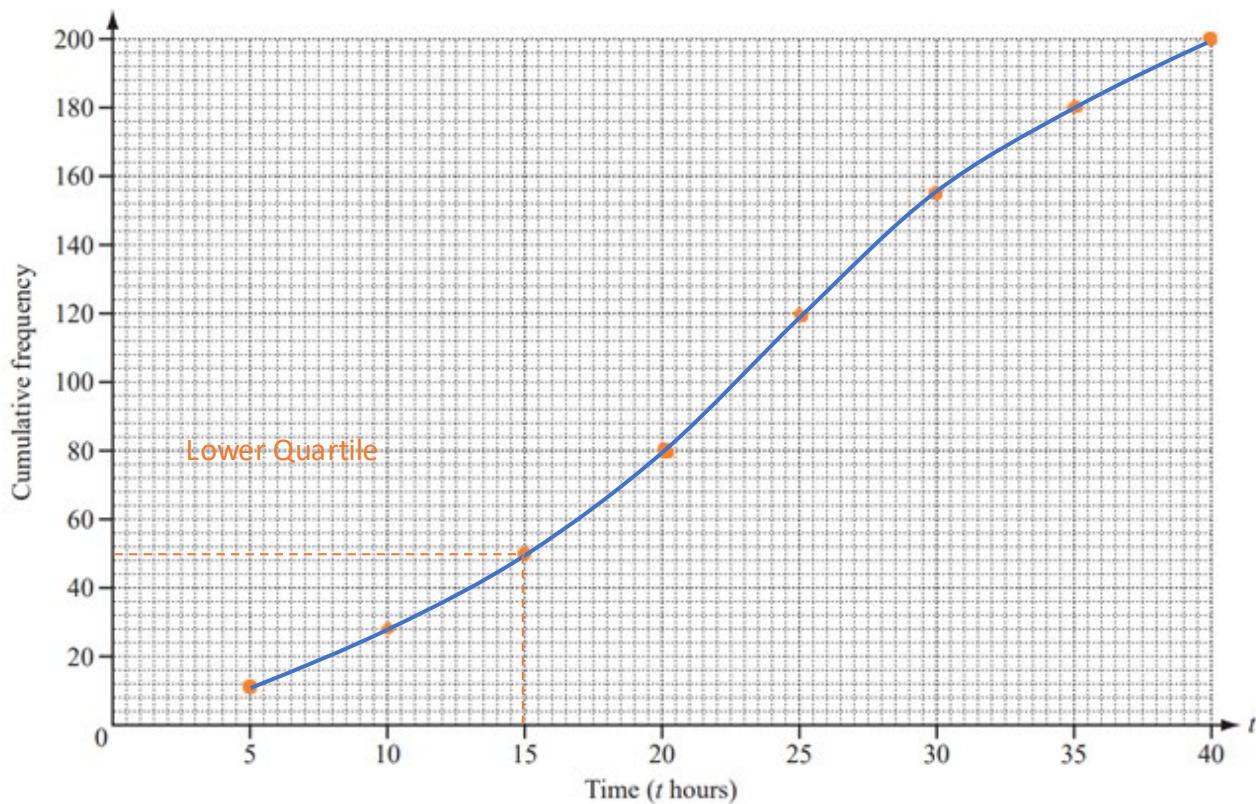
[1]

Time	$t \leq 5$	$t \leq 10$	$t \leq 15$	$t \leq 20$	$t \leq 25$	$t \leq 30$	$t \leq 35$	$t \leq 40$
Cumulative frequency	12	27	50	80	120	155	180	200



(c) On the grid, draw a cumulative frequency diagram to show the information in the table in part (b).

[4]



- (d) On your cumulative frequency diagram show how to find the lower quartile. [1]

Lower quartile drawn on in dotted orange on graph above.

- (e) Use your cumulative frequency diagram to find

- (i) the median, [1]

22.5

- (ii) the inter-quartile range, [1]

29 – 15

= 14

- (iii) the 64th percentile, [1]

26.5

(iv) the number of students who exercise for more than 17 hours.

[2]

$$200 - 60$$

$$= 140$$

Statistics

Difficulty: Medium

Model Answers 3

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Medium
Booklet	Model Answers 3

Time allowed: 124 minutes

Score: /108

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

- (a) The table shows how many books were borrowed by the 126 members of a library group in a month.

Number of books	11	12	13	14	15	16
Number of members (frequency)	35	28	22	18	14	9

Find the mode, the median and the mean for the number of books borrowed.

[6]

Mode is the most common occurring number

$$\text{mode} = \mathbf{11}$$

Median is the average of the middle two numbers in
order (63rd and 64th)

$$\text{median} = \frac{12 + 13}{2}$$

$$= \mathbf{12.5}$$

mode is the average of the numbers

mean

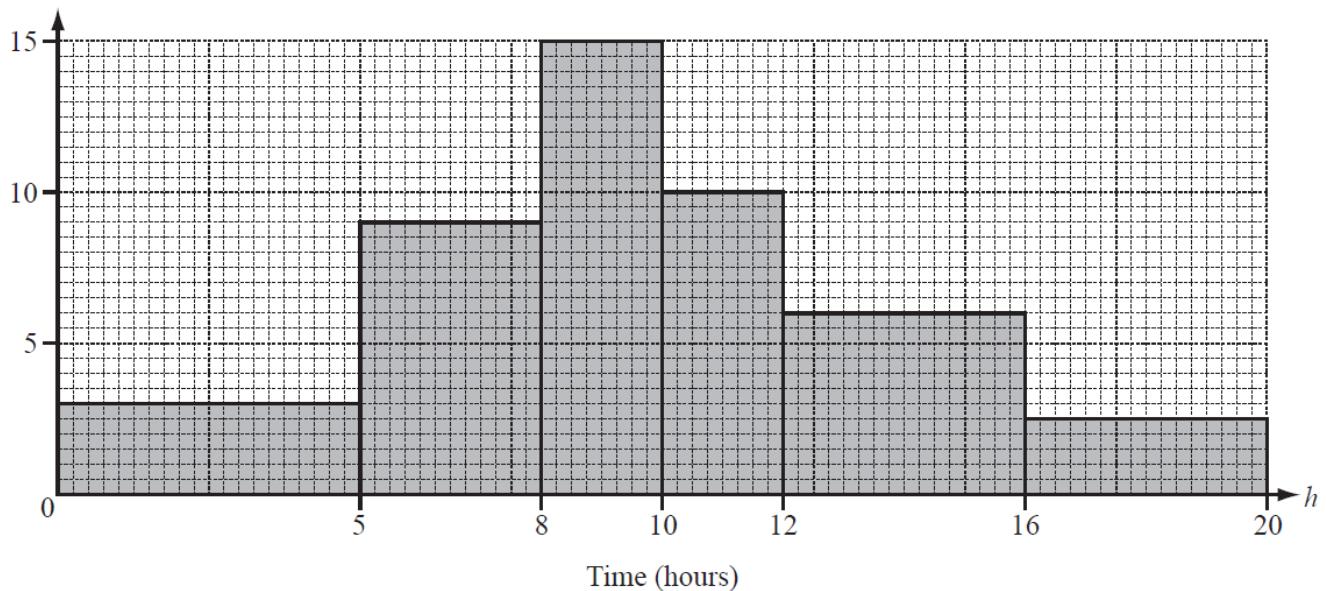
$$= \frac{11 \times 35 + 12 \times 28 + 13 \times 22 + 14 \times 18 + 15 \times 14 + 16 \times 9}{126}$$

$$= \mathbf{12.8}$$

- (b) The 126 members record the number of hours they read in one week.

The histogram shows the results.

Frequency
density



- (i) Use the information from the histogram to complete the frequency table.

Number of hours (h)	$0 < h \leq 5$	$5 < h \leq 8$	$8 < h \leq 10$	$10 < h \leq 12$	$12 < h \leq 16$	$16 < h \leq 20$
Frequency				20	24	10

[3]

Number of hours (h)	Frequency
$0 < h \leq 5$	15
$5 < h \leq 8$	27
$8 < h \leq 10$	30
$10 < h \leq 12$	20
$12 < h \leq 16$	24
$16 < h \leq 20$	10

- (ii) Use the information in this table to calculate an estimate of the mean number of hours.
Show your working.

[4]

Add the following two columns

Number of hours (h)	Frequency	Midpoint	<i>midpoint × f</i>
$0 < h \leq 5$	15	2.5	37.5
$5 < h \leq 8$	27	6.5	175.5
$8 < h \leq 10$	30	9	270
$10 < h \leq 12$	20	11	220
$12 < h \leq 16$	24	14	336
$16 < h \leq 20$	10	18	180
Sum	126		1219

Then the estimate of the mean is

$$\mu = \frac{\sum f \times \text{midpoint}}{\sum f}$$

$$= \frac{1219}{126}$$

$$= 9.67$$

Question 2

Fifty students are timed when running one kilometre. The results are shown in the table.

Time (t minutes)	$4.0 < t \leq 4.5$	$4.5 < t \leq 5.0$	$5.0 < t \leq 5.5$	$5.5 < t \leq 6.0$	$6.0 < t \leq 6.5$	$6.5 < t \leq 7.0$
Frequency	2	7	8	18	10	5

- (a) Write down the modal time interval. [1]

5.5 < $t \leq 6.0$

- (b) Calculate an estimate of the mean time. [4]

Find midpoints for each time interval

4.25, 4.75, 5.25, 5.75, 6.25, 6.75

$$(2 \times 4.25) + (7 \times 4.75) + (8 \times 5.25) + (18 \times 5.75) + (10 \times 6.25) + (5 \times 6.75) = 283.5$$

Divide by sum of the frequencies

$$283.5 / (2 + 7 + 8 + 18 + 10 + 5) = 283.5 / 50$$

$$= 5.67$$

- (c) A new frequency table is made from the results shown in the table above.

Time (t minutes)	$4.0 < t \leq 5.5$	$5.5 < t \leq 6.0$	$6.0 < t \leq 7.0$
Frequency		18	

- (i) Complete the table by filling in the two empty boxes. [1]

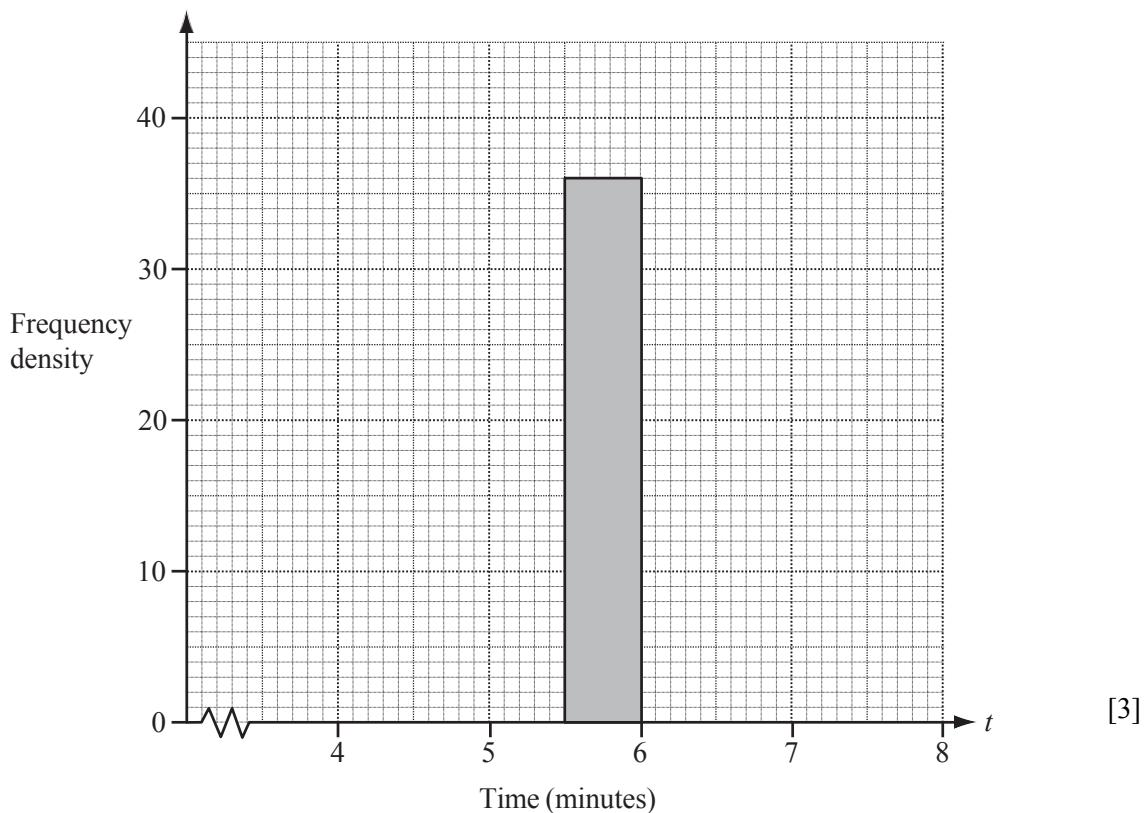
$$4.0 < t \leq 5.5 \rightarrow 2 + 7 + 8$$

$$= 17$$

$$6.0 < t \leq 7.0 \rightarrow 10 + 5$$

$$= 15$$

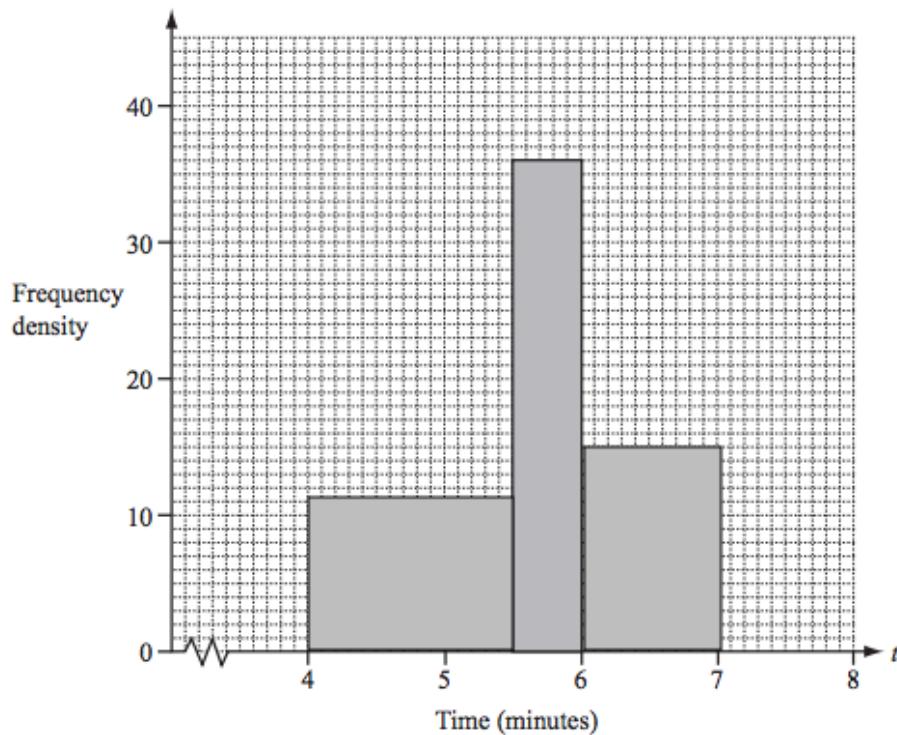
(ii) On the grid below, complete an accurate histogram to show the information in this new table.



$$fd = \text{frequency} / \text{class width}$$

$$4.0 < t \leq 5.5 \rightarrow fd = 17 / 1.5 = 11.33\dots$$

$$6.0 < t \leq 7.0 \rightarrow fd = 15 / 1 = 15$$



(iii) Find the number of students represented by 1 cm^2 on the histogram.

[1]

1cm = 5 squares

Find area in cm^2 of a bar with a known frequency

$$5.5 < t < 6.0 \rightarrow 36 \text{ squares} \times 5 \text{ squares} = 36/5\text{cm} \times 5/5\text{cm} = 7.2\text{cm}^2$$

Divide frequency by area

$$18/7.2$$

$$= 2.5$$

Question 3

A normal die, numbered 1 to 6, is rolled 50 times.



The results are shown in the frequency table.

Score	1	2	3	4	5	6
Frequency	15	10	7	5	6	7

(a) Write down the modal score.

[1]

The modal score is the one with the highest frequency, in this case,

1.

(b) Find the median score.

[1]

The median score represents the middle value from the ordered list of scores.

The total frequency is 50, therefore the middle value would be the mean of the scores 2 and 3. This is because there is an even number of values, 50.

$$\frac{2+3}{2}$$

$$= 2.5$$

(c) Calculate the mean score.

[2]

To calculate the mean score we need to multiply each score by its corresponding frequency and then sum up the values to obtain the total score. This score is divided by the total frequency to obtain the mean.

$$\text{Mean} = \frac{15 \times 1 + 10 \times 2 + 7 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6}{50}$$

$$\text{Mean} = 2.96$$

(d) The die is then rolled another 10 times.

The mean score for the 60 rolls is 2.95.

Calculate the mean score for the extra 10 rolls.

[3]

The score for the first 50 times is:

$$15 \times 1 + 10 \times 2 + 7 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6 = 148$$

The new mean for 60 rolls is:

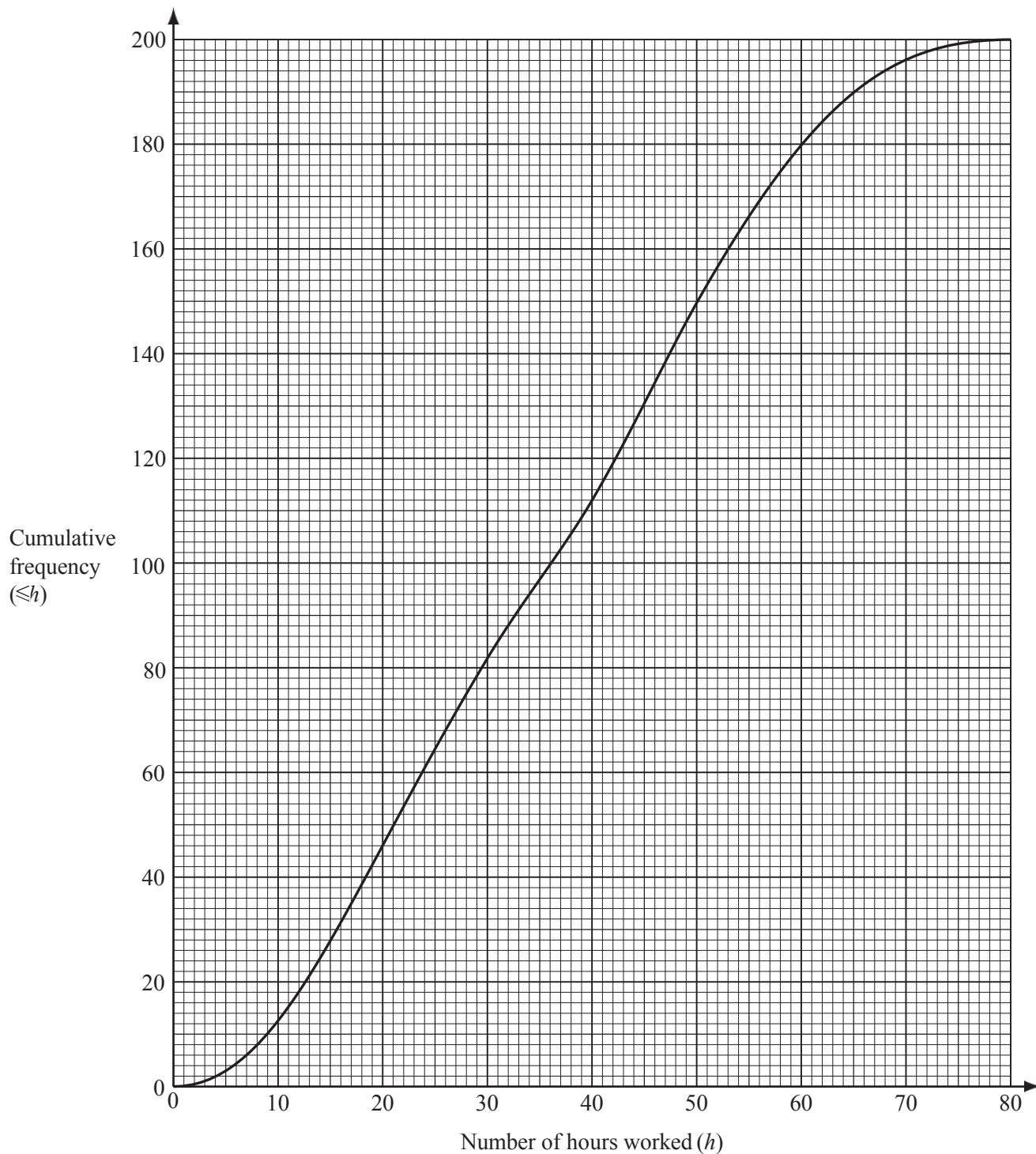
$$\text{Mean} = \frac{148 + \text{score for the extra 10 times}}{60} = 2.95$$

$$\text{The score for the extra 10 times} = 29$$

The mean only for the extra 10 times is:

$$\text{Mean} = \frac{29}{10} = 2.9$$

Question 4



200 people record the number of hours they work in a week.
The cumulative frequency graph shows this information.

Let y be the Cumulative frequency for the purposes of these answers

(a) Use the graph to find

(i) the median,

[1]

Read off the graph at $y = 100$ to get

$$\mathbf{h = 36}$$

(ii) the upperquartile,

[1]

Read off the graph at $y = 150$ to get

$$\mathbf{h = 50}$$

(iii) the inter-quartilerange,

[1]

Read off the graph at $y = 50$ to get the lower quartile

$$\mathbf{h = 21}$$

Hence the inter-quartile range is

$$\mathbf{50 - 21}$$

$$\mathbf{= 29}$$

(iv) the number of people who work more than 60 hours in a week.

[2]

Read of the graph at $h = 60$

$$\mathbf{y = 180}$$

This gives us the number of people who have worked 60 hours or less.

Hence the number of people who worked over 60 hours is

$$\mathbf{200 - 180}$$

$$\mathbf{= 20}$$

(b) Omar uses the graph to make the following frequency table.

Hours worked (h)	$0 < h \leq 10$	$10 < h \leq 20$	$20 < h \leq 30$	$30 < h \leq 40$	$40 < h \leq 50$	$50 < h \leq 60$	$60 < h \leq 70$	$70 < h \leq 80$
Frequency	12	34	36	30	38	30	p	q

(i) Use the graph to find the values of p and q .

[2]

Read off the graph at $h = 70$ and minus the y -value for $h = 60$

$$p = 196 - 180$$

$$= 16$$

Duplicate for $h = 80$ and $h = 70$

$$q = 200 - 196$$

$$= 4$$

(ii) Calculate an estimate of the mean number of hours worked in a week.

[4]

Hours worked (h)	Frequency (f)	Midpoint (M)	$f \times M$
$0 < h \leq 10$	12	5	60
$10 < h \leq 20$	34	15	510
$20 < h \leq 30$	36	25	900
$30 < h \leq 40$	30	35	1050
$40 < h \leq 50$	38	45	1710
$50 < h \leq 60$	30	55	1650
$60 < h \leq 70$	16	65	1040
$70 < h \leq 80$	4	75	300
Sum	200		7220

Estimate of the mean is

$$\mu = \frac{\sum f \times M}{\sum f}$$

$$= \frac{7220}{200}$$

$$= 36.1$$

(c) Shalini uses the graph to make a different frequency table.

Hours worked (h)	$0 < h \leq 30$	$30 < h \leq 40$	$40 < h \leq 50$	$50 < h \leq 80$
Frequency	82	30	38	50

When she draws a histogram, the height of the column for the interval $30 < h \leq 40$ is 9 cm.

Calculate the height of each of the other three columns.

[4]

The height of each column, x , is its frequency density.

Let s be the scalar for the height axis.

So, we have that, for $30 < h \leq 40$

$$(40 - 30) \times 9s = 30$$

$$\rightarrow 90s = 30$$

$$\rightarrow s = \frac{1}{3}$$

For the $0 < h \leq 30$ column

$$(30 - 0) \times \frac{1}{3}x_1 = 82$$

$$\rightarrow 10x_1 = 82$$

$$\rightarrow x_1 = 8.2$$

For the $40 < h \leq 50$ column

$$(50 - 40) \times \frac{1}{3}x_2 = 38$$

$$\rightarrow \frac{10}{3}x_2 = 38$$

$$\rightarrow x_2 = 11.4$$

For the $50 < h \leq 80$ column

$$(80 - 50) \times \frac{1}{3}x_3 = 50$$

$$\rightarrow 10x_3 = 50$$

$$\rightarrow x_3 = 5$$

Question 5

Kristina asked 200 people how much water they drink in one day.

The table shows her results.

Amount of water (x litres)	Number of people
$0 < x \leq 0.5$	8
$0.5 < x \leq 1$	27
$1 < x \leq 1.5$	45
$1.5 < x \leq 2$	50
$2 < x \leq 2.5$	39
$2.5 < x \leq 3$	21
$3 < x \leq 3.5$	7
$3.5 < x \leq 4$	3

(a) Write down the modal interval.

[1]

There are 200 people in total, therefore, the middle value is 100.

To work out the model interval, we need to find the interval which includes the 100th person, in order.

Starting from the first interval, we add up the frequencies:

$$8 < 100$$

$$8 + 27 = 35 < 100$$

$$\dots 8 + 27 + 45 + 50 = 130 > 100.$$

The modal interval is the one in which the 100th person is, in this case, $1.5 < x \leq 2$.

(b) Calculate an estimate of the mean.

[4]

To work out the estimate of the mean we need to initially work out the mid-value for each of the intervals. This mid-value is multiplied by the frequency corresponding to that interval to obtain an estimate of the total amount of water drank in one day by the people included in that interval.

This total amount of water is divided by the total frequency to obtain an estimate of the mean.

Amount of water (litres)	Frequency	Mid-value	Mid-value x frequency
$0 < x \leq 0.5$	8	0.25	2
$0.5 < x \leq 1$	27	0.75	20.25
$1 < x \leq 1.5$	45	1.25	56.25
$1.5 < x \leq 2$	50	1.75	87.5
$2 < x \leq 2.5$	39	2.25	87.75
$2.5 < x \leq 3$	21	2.75	57.75
$3 < x \leq 3.5$	7	3.25	22.75
$3.5 < x \leq 4$	3	3.75	11.25
Total	200	-	345.5

$$\text{Estimate mean} = \frac{345.5}{200}$$

$$\text{Estimate mean} = 1.7275$$

- (c) Make a cumulative frequency table for this data.

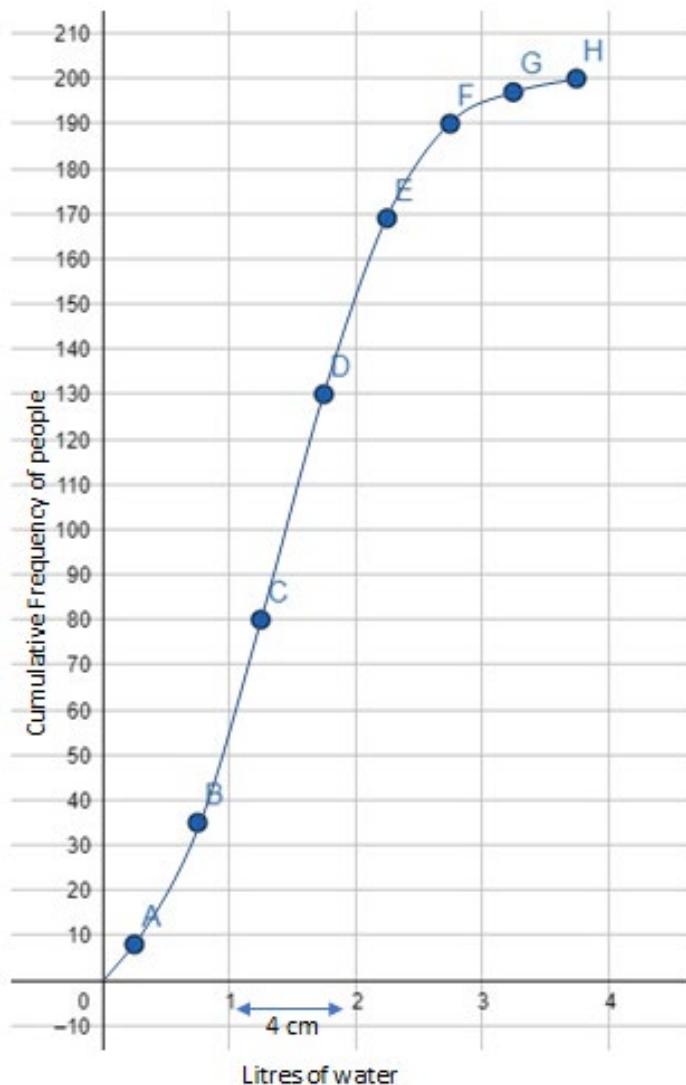
[2]

We draw a cumulative frequency table by adding up the frequencies for each interval, in order.

Amount of water (litres)		Cumulative Frequency
$0 < x \leq 0.5$	8	8
$0.5 < x \leq 1$	$27 + 8$	35
$1 < x \leq 1.5$	$35 + 45$	80
$1.5 < x \leq 2$	$80 + 50$	130
$2 < x \leq 2.5$	$130 + 139$	169
$2.5 < x \leq 3$	$169 + 21$	190
$3 < x \leq 3.5$	$190 + 7$	197
$3.5 < x \leq 4$	$197 + 3$	200

- (d) Using a scale of 4 cm to 1 litre of water on the horizontal axis and 1 cm to 10 people on the vertical axis, draw the cumulative frequency graph.

[5]



(e) Use your cumulative frequency graph to find

(i) the median,

[1]

Median = approximately 1.7 (the x value corresponding to 100 people – half the frequency)

(ii) the 40th percentile,

[1]

40th percentile = 1.5 (the x value corresponding to 40% of the people – y = 80)

(iii) the number of people who drink at least 2.6 litres of water.

[2]

The y value corresponding to 2.6 litres of water is approximately 175 people – the ones that drink less than 2.6.

The number of people who drink 2.6 litres or more is approximately = 200 – 175 = 25 people

(f) A doctor recommends that a person drinks at least 1.8 litres of water each day.
What percentage of these 200 people do not drink enough water?

[2]

There are at least 110 people which do not drink more than 1.8 litres.

Percentage = 55%

Question 6

- (a) The quiz scores of a class of n students are shown in the table.

Quiz score	6	7	8	9
Frequency (number of students)	9	3	a	5

The mean score is 7.2. Find

- (i) a ,

[3]

The mean score is worked out by multiplying each score by its frequency,

adding up the total score and then dividing it by the total frequency:

$$\text{Mean score} = \frac{6 \times 9 + 7 \times 3 + 8a + 9 \times 5}{9 + 3 + a + 5}$$

$$\text{Mean score} = \frac{120 + 8a}{17 + a} = 7.2$$

We solve the equation for a .

$$120 + 8a = 7.2(17 + a)$$

$$120 + 8a = 122.4 + 7.2a$$

$$0.8a = 2.4$$

$$\mathbf{a = 3}$$

- (ii) n ,

[1]

$$n = 9 + 3 + a + 5$$

$$\mathbf{n = 20}$$

- (iii) the median score.

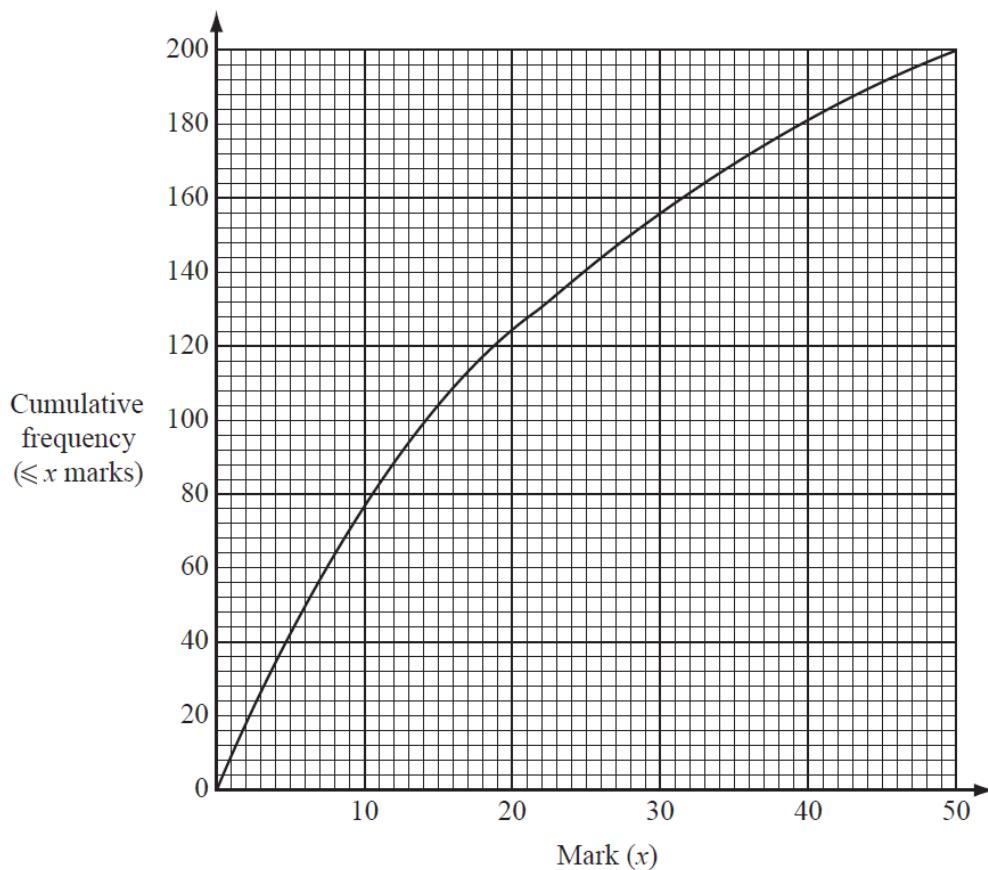
[1]

The median score is 7 since the middle frequency is 10, therefore, is included

in this score. ($9 + 3 = 12 > 10$)

(b) 200 students take a mathematics test.

The cumulative frequency diagram shows the results.



Write down

- (i) the median mark,

[1]

The median is the mark corresponding to the middle frequency.

The middle frequency value is 100, so the **median mark is 14**.

- (ii) the lower quartile,

[1]

The lower quartile is the mark corresponding to one quarter of the frequency value, $200/4 = 50$.

The lower quartile is: 6

- (iii) the upper quartile,

[1]

The upper quartile is the mark corresponding to 3 quarters of the frequency value,

$$200 \times 3/4 = 150.$$

The upper quartile is: 28

- (iv) the inter-quartile range,

[1]

The interquartile range (IQR) is the difference between the upper and lower quartiles:

$$\text{IQR} = 28 - 6 = 22$$

- (v) the lowest possible mark scored by the top 40 students,

[1]

The top 40 students are the ones between 160 and 200. The lowest possible mark from this range will be the one corresponding to the last student in this range, the 160th one.

The lowest mark is: 32

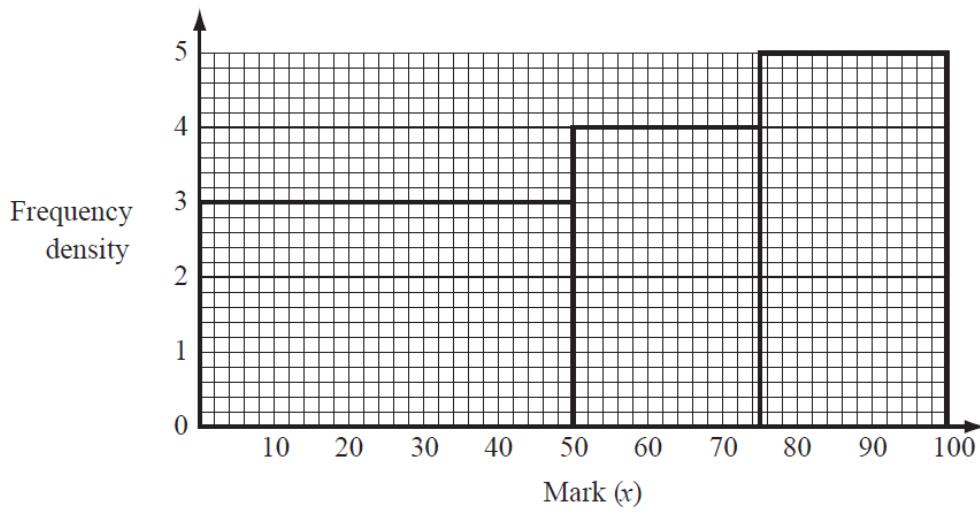
- (vi) the number of students scoring more than 25 marks.

[1]

25 marks are scored by the 140th students.

Therefore, there are $200 - 140 = 60$ students which got more than 25 marks.

- (c) Another group of students takes an English test.
 The results are shown in the histogram.



- (i) How many students score marks in the range $0 < x \leq 50$? [1]

$$\text{frequency density} = \text{frequency} / \text{group width}$$

$$\text{group width} = 50 - 0 = 50 \text{ and frequency density} = 3$$

$$\text{frequency} = 3 \times 50 = 150$$

- (ii) How many students score marks in the range $75 < x \leq 100$? [1]

$$\text{group width} = 100 - 75 = 25 \text{ and frequency density} = 5$$

$$\text{frequency} = 5 \times 25 = 125$$

- (iii) Calculate an estimate of the mean mark of this group of students. [4]

To calculate an estimate of the mean mark we need to work out the middle values

of each group and multiply it by the corresponding frequency. By adding up these

values we obtain the total marks which we can divide by the total frequency to work

out the mean.

$$\text{Mean mark} = \frac{25 \times 150 + 100 \times 62.5 + 87.5 \times 125}{150 + 100 + 125}$$

$$\text{Mean mark} = 55.8$$

Question 7

- (a) Students are given marks 0, 1, 2, 3 or 4 for a piece of work.

The table shows the number of students getting each mark.

Mark	0	1	2	3	4
Frequency	3	10	12	9	x

- (i) The mean mark is 2.125.

Find the value of x .

[4]

The mean is the sum of all the numbers divided by the amount of numbers.

$$\text{Mean} = \frac{3 \times 0 + 10 \times 1 + 12 \times 2 + 9 \times 3 + 4 \times X}{3 + 10 + 12 + 9 + X}$$

$$2.125 = \frac{61 + 4x}{34 + x}$$

$$(34 + x) \times 2.125 = 61 + 4x$$

$$72.25 + 2.125x = 61 + 4x$$

$$11.1 = 1.875x$$

$$x = 6$$

- (ii) Write down the lower quartile mark.

[1]

The lower quartile is the median of the lower half of the data.

The total frequency is:

$$3 + 10 + 12 + 9 + 6 = 40$$

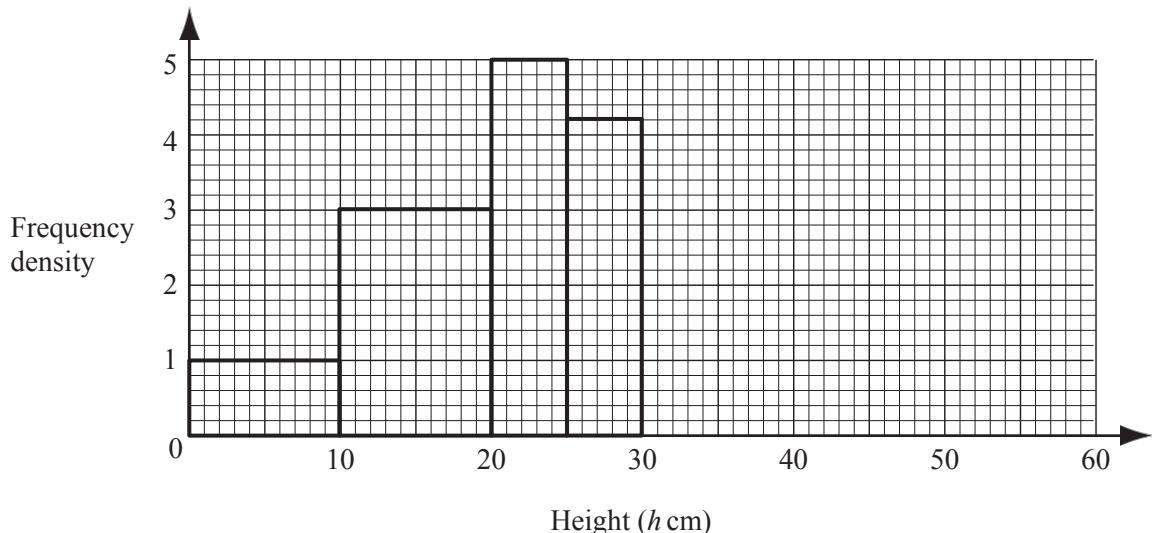
To work out the lower quartile, we divide the total by 4.

$$\text{Lower quartile} = 10$$

By looking at the table, we can affirm that the mark which has a frequency including the lower quartile is 1, since by adding up the frequencies for the first 2 marks we obtain 13, number which is greater than 10, and the first mark has a frequency of 3, lower than 10.

$$\text{Lower quartile mark} = 1$$

- (b) The heights (h centimetres) of flowers in a shop are shown in the histogram below.
 All the flowers are less than 60 cm high.
 One bar has not been drawn on the histogram.



$$\text{frequency density} = \text{frequency} / \text{group width}$$

In our case, in the interval $20 < h \leq 25$ has a frequency density of 5, according to the histogram.

The group width is:

$$25 - 20 = 5$$

$$\text{Frequency} = 5 \times 5$$

Frequency = 25 flowers.

- (i) There are 25 flowers in the interval $20 < h \leq 25$.
 How many flowers are there in the intervals

$$(a) \quad 25 < h \leq 30,$$

[1]

In the same way, the interval $25 < h \leq 30$ has the frequency density 4.2 and the group width:

$$30 - 25 = 5$$

$$\text{Frequency} = 5 \times 4.2$$

Frequency = 21 flowers.

(b) $10 < h \leq 20$

[1]

The interval $10 < h \leq 20$ has the frequency density 3 and the group width:

$$20 - 10 = 10$$

$$\text{Frequency} = 10 \times 3$$

$$\text{Frequency} = 30 \text{ flowers}$$

- (ii) There are 42 flowers in the interval $30 < h \leq 60$.
This can be shown by a single bar on the histogram.
Calculate the height of this bar.

[2]

The height of the histogram bar is represented by the frequency density.

$$\text{Frequency density} = \text{frequency} / \text{group width}$$

In this case, the group width is $60 - 30 = 30$ and the frequency is 42.

$$\text{The frequency density} = \frac{42}{30}$$

$$\text{Frequency density} = 1.4$$

- (iii) Calculate an estimate of the mean height of the flowers.

[3]

We first need to work out the frequency for each interval.

$$\text{Frequency density} = \text{frequency} / \text{group width}.$$

For $0 < h \leq 10$:

$$1 = \text{frequency} / 10$$

$$\text{Frequency} = 10$$

For $20 < h \leq 25$:

$$5 = \text{frequency}/5$$

$$\text{Frequency} = 25$$

To estimate the mean height, we know for example that there are 10 flowers with the height between 0 and 10 cm. Since we do not know an exact height, we will estimate that each student in this group is exactly 5 cm, the midpoint of the group.

Similarly, we work out the mid-point for every interval.

H	Frequency	Mid-point	Mid-point x Frequency
$0 < h \leq 10$	10	5	50
$10 < h \leq 20$	30	15	450
$20 < h \leq 25$	25	22.5	562.5
$25 < h \leq 30$	21	27.5	577.5
$30 < h \leq 60$	42	45	1890
Total	128	-	3530

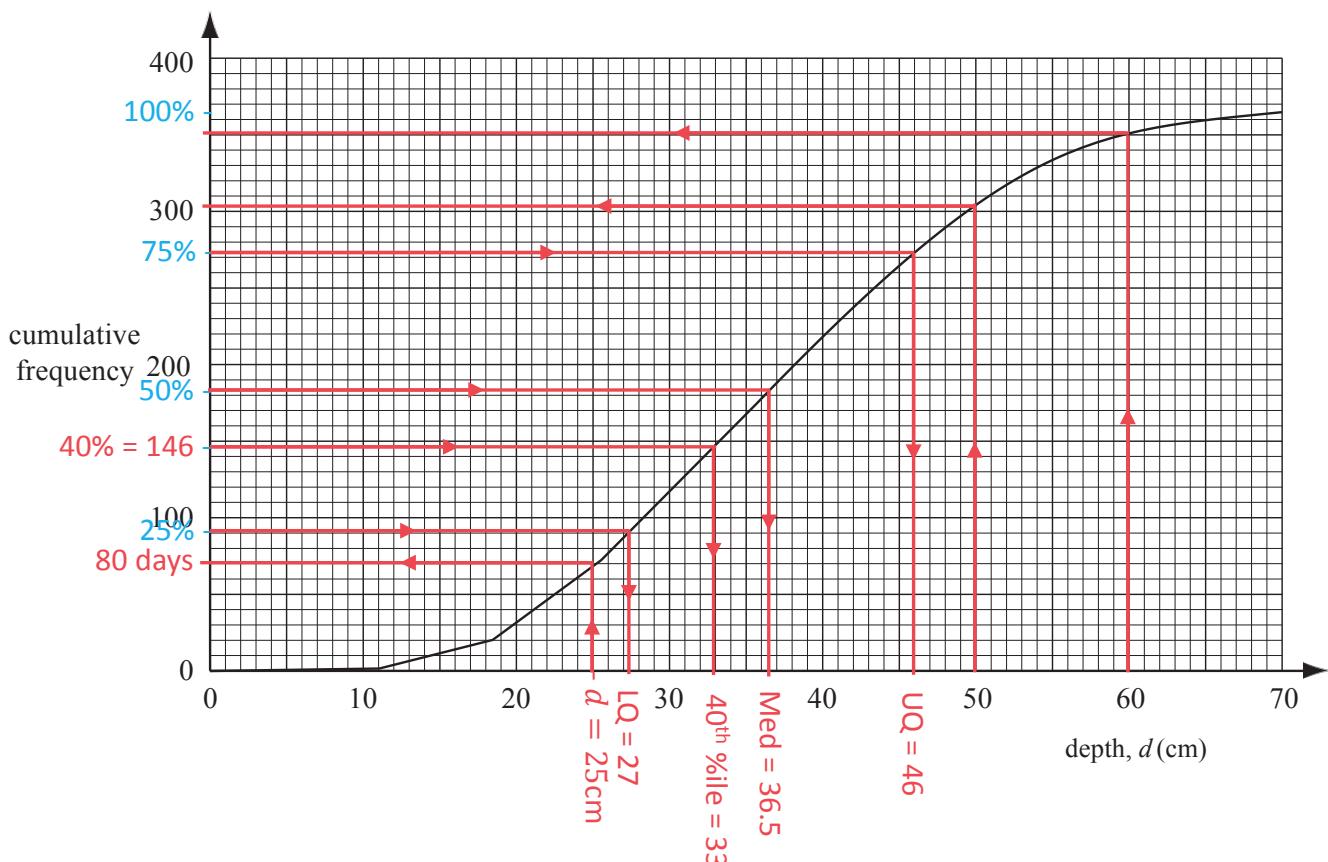
We then calculate the mid-point x frequency for each interval to work out the total height of all the flowers, estimating the mid-point for each height interval.

$$\text{Mean} = \frac{3530}{128}$$

$$\text{Mean} = 27.57$$

Question 8

The depth, d centimetres, of a river was recorded each day during a period of one year (365 days). The results are shown by the cumulative frequency curve.



(a) Use the cumulative frequency curve to find

Relabel the CF axis 0% to 100% as shown and then read across and down as appropriate

(i) the median depth, Median = 36.5 [1]

(ii) the inter-quartile range, IQR = UQ - LQ = 46 - 27 = 19 [2]

(iii) the depth at the 40th percentile, 40th percentile = 33 [2]

(iv) the number of days when the depth of the river was at least 25 cm. [2]

Go to 25 on the depth axis and read up and across.

Number of days less than 25cm = 80

Number of days more than 25cm = 365 - 80 = 285

(b)

d	$0 < d \leq 10$	$10 < d \leq 20$	$20 < d \leq 30$	$30 < d \leq 40$	$40 < d \leq 50$	$50 < d \leq 60$	$60 < d \leq 70$
Number of days	17	41	62	98	85	p	q

- (i) Show that $p = 47$ and $q = 15$.

[2]

Go to 50, 60 (and 70) on the depth axis and read up and across to get 303, 350 (and 365).

$$p = 350 - 303 = 47$$

$$q = 365 - 350 = 15$$

- (ii) Use the information in the table and the values of p and q to calculate an estimate of the mean depth of the river.

[4]

Use the Mid-Interval Values as the x values (d values) in the formula for the mean:

$$\text{Mean} = \frac{\sum fx}{n}$$

$$\text{Mean} = \frac{17 \times 5 + 41 \times 15 + 62 \times 25 + 98 \times 35 + 85 \times 45 + 47 \times 55 + 15 \times 65}{365}$$

$$\text{Mean} = \frac{13065}{365} = 35.8\text{cm}$$

- (c) The following information comes from the table in part (b).

d	$0 < d \leq 20$	$20 < d \leq 40$	$40 < d \leq 70$
Number of days	58	160	147

A histogram was drawn to show this information.

The height of the column for the interval $20 < d \leq 40$ was 8cm.

Calculate the height of each of the other two columns.

[Do not draw the histogram.]

[3]

On a histogram the Frequency is represented by the Area. So

$$\text{Frequency} = \text{Class Width} \times \text{Frequency Density}$$

For $20 < d \leq 40$:

$$\text{FD} = \frac{F}{\text{CW}} = \frac{160}{20} = 8$$

So 1cm represents 1 unit on the Frequency Density axis

For $0 < d \leq 20$:

$$\text{FD} = \frac{F}{\text{CW}} = \frac{58}{20} = 2.9$$

$$\text{Height} = 2.9\text{cm}$$

For $40 < d \leq 70$:

$$\text{FD} = \frac{F}{\text{CW}} = \frac{147}{30} = 4.9$$

$$\text{Height} = 4.9\text{cm}$$

Statistics

Difficulty: Hard

Model Answers 1

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Hard
Booklet	Model Answers 1

Time allowed: 107 minutes

Score: /93

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

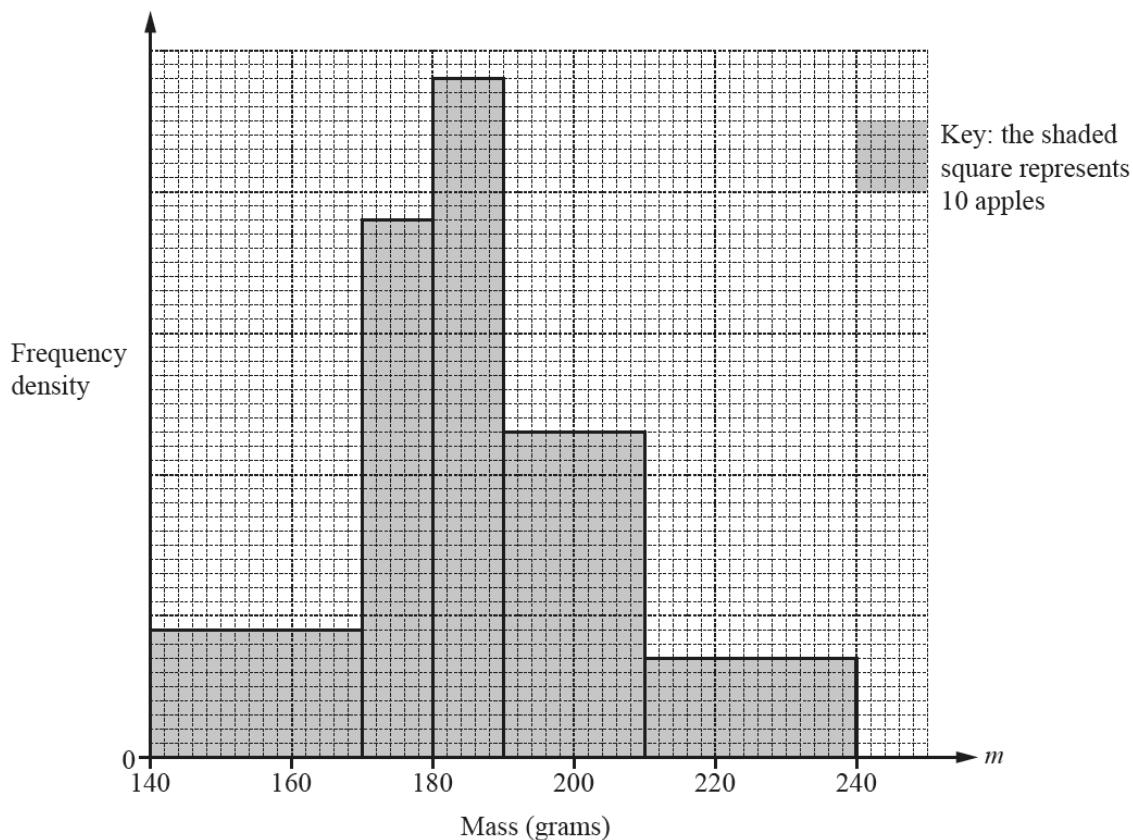
A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

The histogram shows the distribution of the masses, m grams, of 360 apples.



(a) Use the histogram to complete the frequency table.

[3]

Mass (m grams)	Number of apples	Midpoint	Midpoint × Frequency
$140 < m \leq 170$	54	155	8370
$170 < m \leq 180$	76	175	13300
$180 < m \leq 190$	96	185	17760
$190 < m \leq 210$	92	200	18400
$210 < m \leq 240$	42	225	9450
Sum	360		67280

(b) Calculate an estimate of the mean mass of the 360 apples.

[4]

Estimate of the mean is

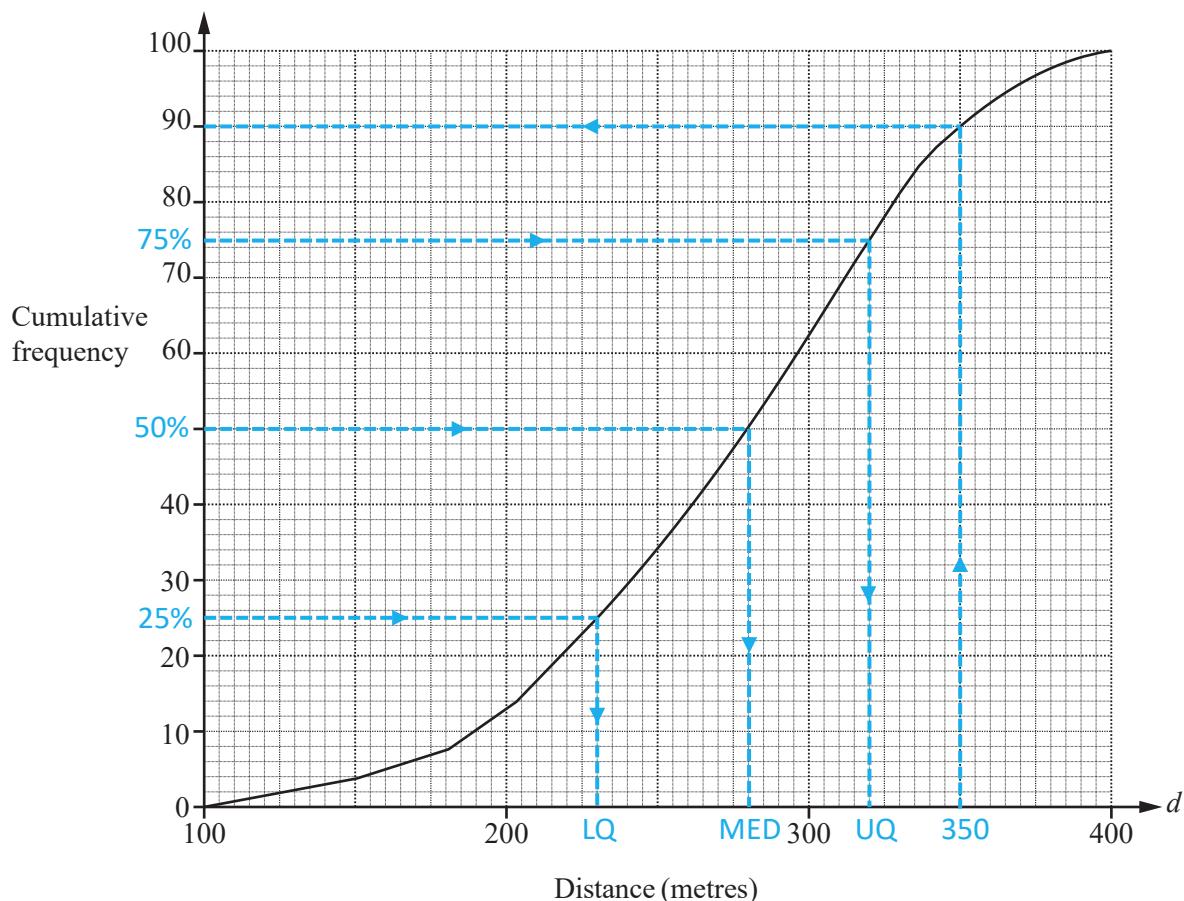
$$\begin{aligned}\mu &= \frac{\sum \text{Midpoint} \times \text{frequency}}{\sum \text{frequency}} \\ &= \frac{67280}{360} = 186.9\end{aligned}$$

Question 2

- (a) There are 100 students in group A.

The teacher records the distance, d metres, each student runs in one minute.

The results are shown in the cumulative frequency diagram.



Find

- (i) the median,

[1]

Read across and down from 50% to get **MED = 280**

- (ii) the upper quartile,

[1]

Read across and down from 75% to get **UQ = 320**

- (iii) the inter-quartile range,

[1]

Read across and down from 25% to get **LQ = 230**

And then use $IQR = UQ - LQ$ to get **IQR = $320 - 230 = 90$**

- (iv) the number of students who run more than 350m.

[2]

Read up and across from 350 to get: 90 students run less than 350m

So **100 - 90 = 10 students run more than 350m**

(b) There are 100 students in group B .

The teacher records the distance, d metres, each of these students runs in one minute.

The results are shown in the frequency table.

Distance (d metres)	$100 < d \leq 200$	$200 < d \leq 250$	$250 < d \leq 280$	$280 < d \leq 320$	$320 < d \leq 400$
Number of students	20	22	30	16	12

(i) Calculate an estimate of the mean distance for group B .

[4]

To find the mean from grouped data use the formula $\text{mean} = \frac{\sum fx}{n}$ where x is the

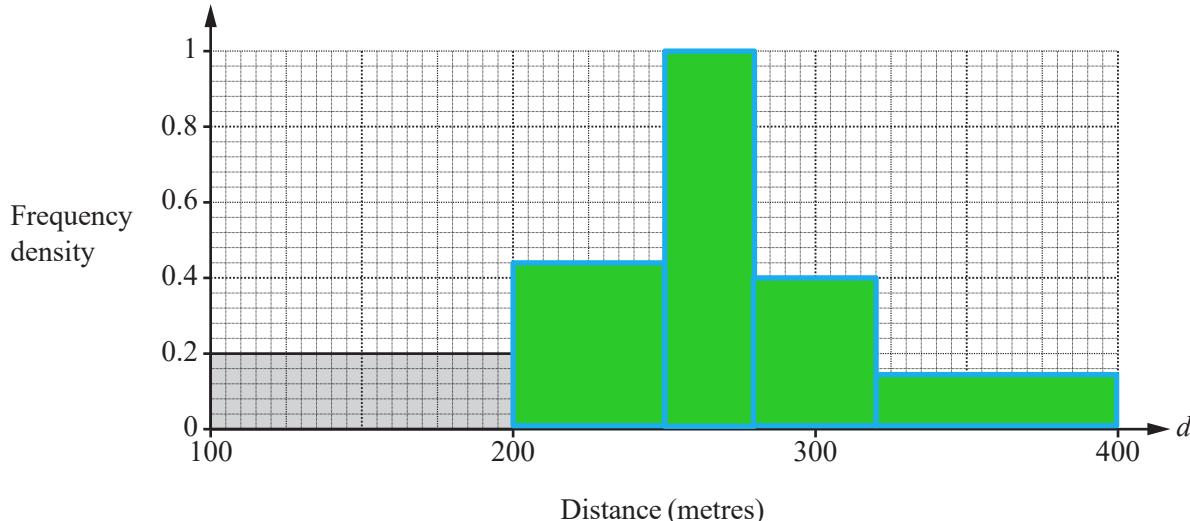
mid-interval value and f the frequency in each group, and n is the total frequency.

$$\text{mean} = \frac{20 \times 150 + 22 \times 225 + 30 \times 265 + 16 \times 300 + 12 \times 360}{100}$$

$$\text{mean} = \frac{25020}{100}$$

$$\text{mean} = 250.2$$

(ii) Complete the histogram to show the information in the frequency table.



For each group calculate the Frequency Density = $\frac{\text{Frequency}}{\text{Class Width}}$

[4]

$$\frac{20}{100} = 0.2, \quad \frac{22}{50} = 0.44, \quad \frac{30}{30} = 1, \quad \frac{16}{40} = 0.4, \quad \frac{12}{80} = 0.15$$

and draw a rectangle of height = Frequency Density and width = Class Width for each group.

The first one is already drawn.

- (c) For the 100 students in group *B*, the median is 258m.

Complete the statement.

On average, the students in group *A* run than the students in group *B*.

[1]

The Median is a type of Average.

Since the Median (=280) for Group A is bigger than the Median (=258) for Group B we can say:

On average, the students in group *A* run **FURTHER** than the students in group *B*.

Question 3

The time taken for each of 90 cars to complete one lap of a race track is shown in the table.

Time (t seconds)	$70 < t \leq 71$	$71 < t \leq 72$	$72 < t \leq 73$	$73 < t \leq 74$	$74 < t \leq 75$
Frequency	17	24	21	18	10

- (a) Write down the modal time interval.

[1]

71 < $t \leq 72$

- (b) Calculate an estimate of the mean time.

[4]

Time (t seconds)	Frequency (f)	Midpoint (m)	$f \times m$
$70 < t \leq 71$	17	70.5	1198.5
$71 < t \leq 72$	24	71.5	1716
$72 < t \leq 73$	21	72.5	1522.5
$73 < t \leq 74$	18	73.5	1323
$74 < t \leq 75$	10	74.5	745
Sum	90		6505

Estimate of the mean is

$$\begin{aligned}\mu &= \frac{\sum(f \times m)}{\sum f} \\ &= \frac{6505}{90} \\ &= 72.3\end{aligned}$$

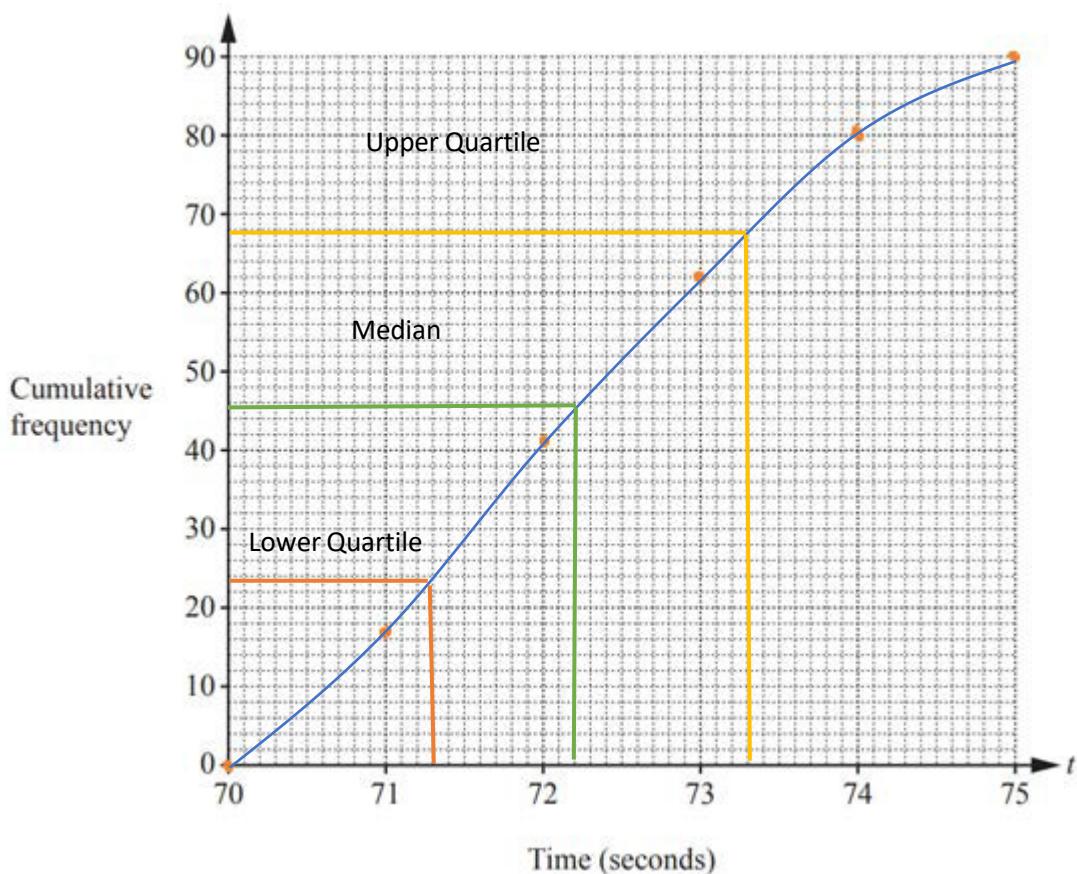
- (c) (i) Complete the cumulative frequency table.

Time (t seconds)	$t \leq 71$	$t \leq 72$	$t \leq 73$	$t \leq 74$	$t \leq 75$
Cumulative frequency	17	41	62	80	90

[2]

- (ii) On the grid, draw a cumulative frequency diagram to show this information. [3]

Plotted in orange, drawn in blue.



- (iii) Find the median time. [1]

Median found by reading off the green lines drawn in diagram above

72.2

- (iv) Find the inter-quartile range. [2]

Upper quartile line in yellow, lower quartile in orange.

The interquartile range is the difference

$$73.3 - 71.3$$

$$= 2$$

- (d) One lap of the race track measures 3720 metres, correct to the nearest 10 metres.
A car completed the lap in 75 seconds, correct to the nearest second.

Calculate the upper bound for the average speed of this car.
Give your answer in kilometres per hour.

[4]

The speed, time, distance relation is

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

For an upper bound on speed we need the distance to be as large as possible and the time to be as small as possible

$$\text{distance} = 3725$$

$$\text{time} = 74.5$$

Upper bound for speed is then

$$\frac{3725}{74.5} = 50\text{ms}^{-1}$$

We need to convert this to kilometres per hour

$$50 \times \frac{60^2}{1000}$$
$$= 180 \text{ kmh}^{-1}$$

Question 4

- (a) 200 students estimate the capacity, x millilitres, of a cup.
The results are shown in the frequency table.

Capacity (x ml)	$0 < x \leq 100$	$100 < x \leq 150$	$150 < x \leq 200$	$200 < x \leq 250$	$250 < x \leq 400$
Frequency	20	55	66	35	24

- (i) Calculate an estimate of the mean. [4]

Capacity (x ml)	Frequency (f)	Midpoint (m)	$f \times m$
$0 < x \leq 100$	20	50	1000
$100 < x \leq 150$	55	125	6875
$150 < x \leq 200$	66	175	11550
$200 < x \leq 250$	35	225	7875
$250 < x \leq 400$	24	325	7800
Sum	200		35100

Estimate of the mean is

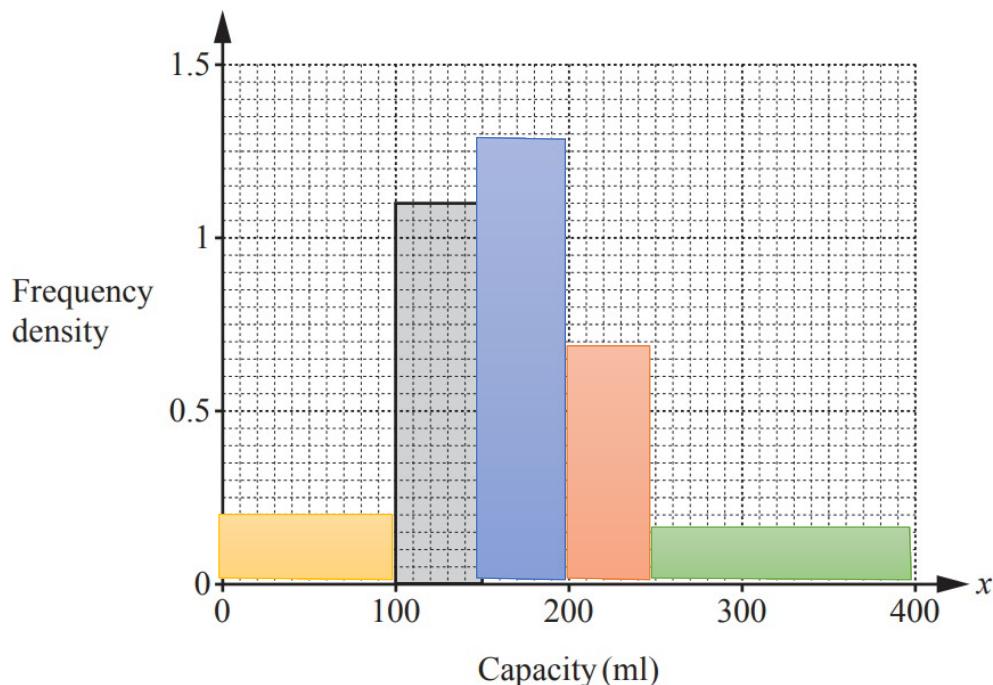
$$\mu = \frac{\sum f \times m}{\sum f}$$

$$= 35100 \div 200$$

$$= 175.5$$

(ii) Complete the histogram.

[4]



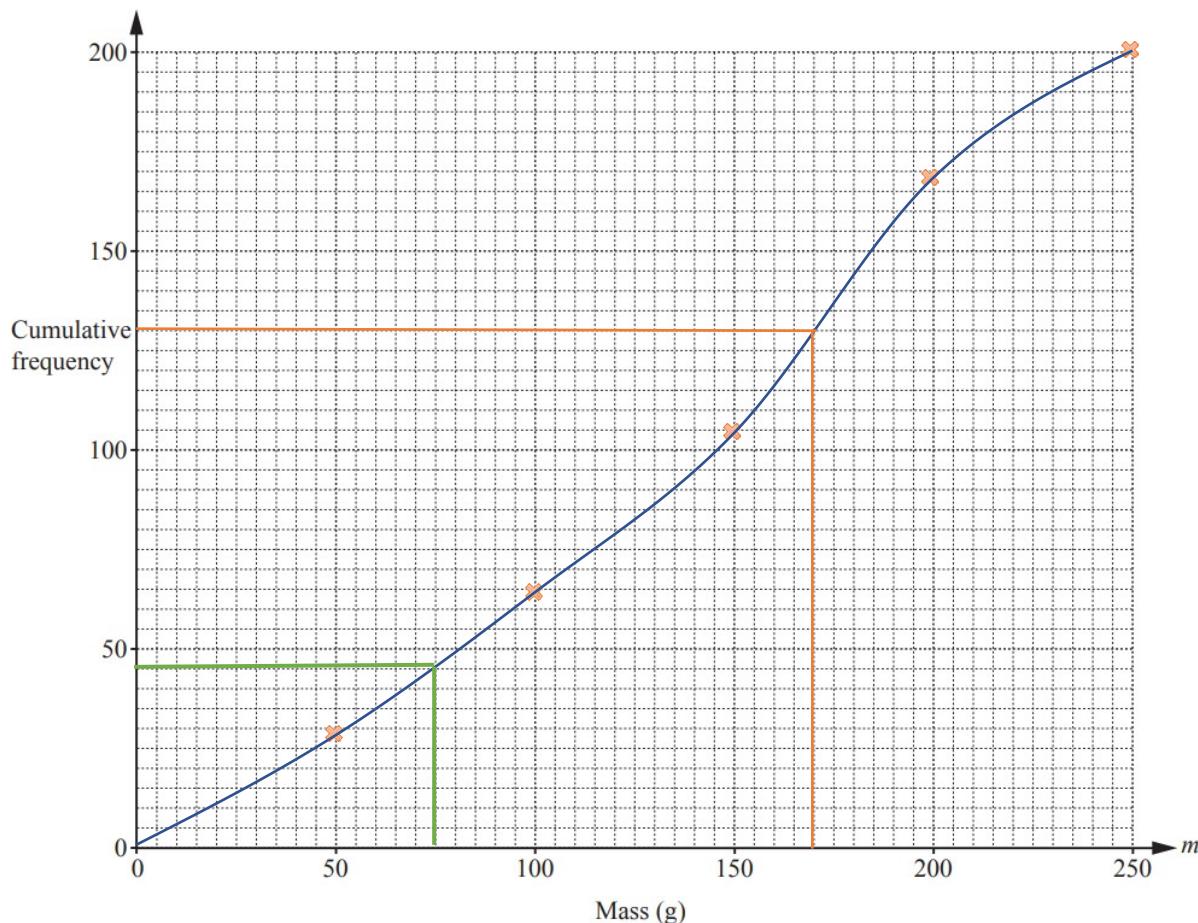
(b) The 200 students also estimate the mass, m grams, of a small rock.

The results are shown in the cumulative frequency table.

Mass (m grams)	$m \leq 50$	$m \leq 100$	$m \leq 150$	$m \leq 200$	$m \leq 250$
Cumulative frequency	28	64	104	168	200

(i) On the grid, draw a cumulative frequency diagram.

[3]



(ii) Find

(a) the 65th percentile,

[1]

The orange line

$$65\% \text{ of } 200 = 130$$

$$\textbf{65th percentile} = 170$$

(b) the number of students who estimated more than 75 g.

[2]

The green line is students who estimates 75g or less

$$200 - 45$$

$$= 155$$

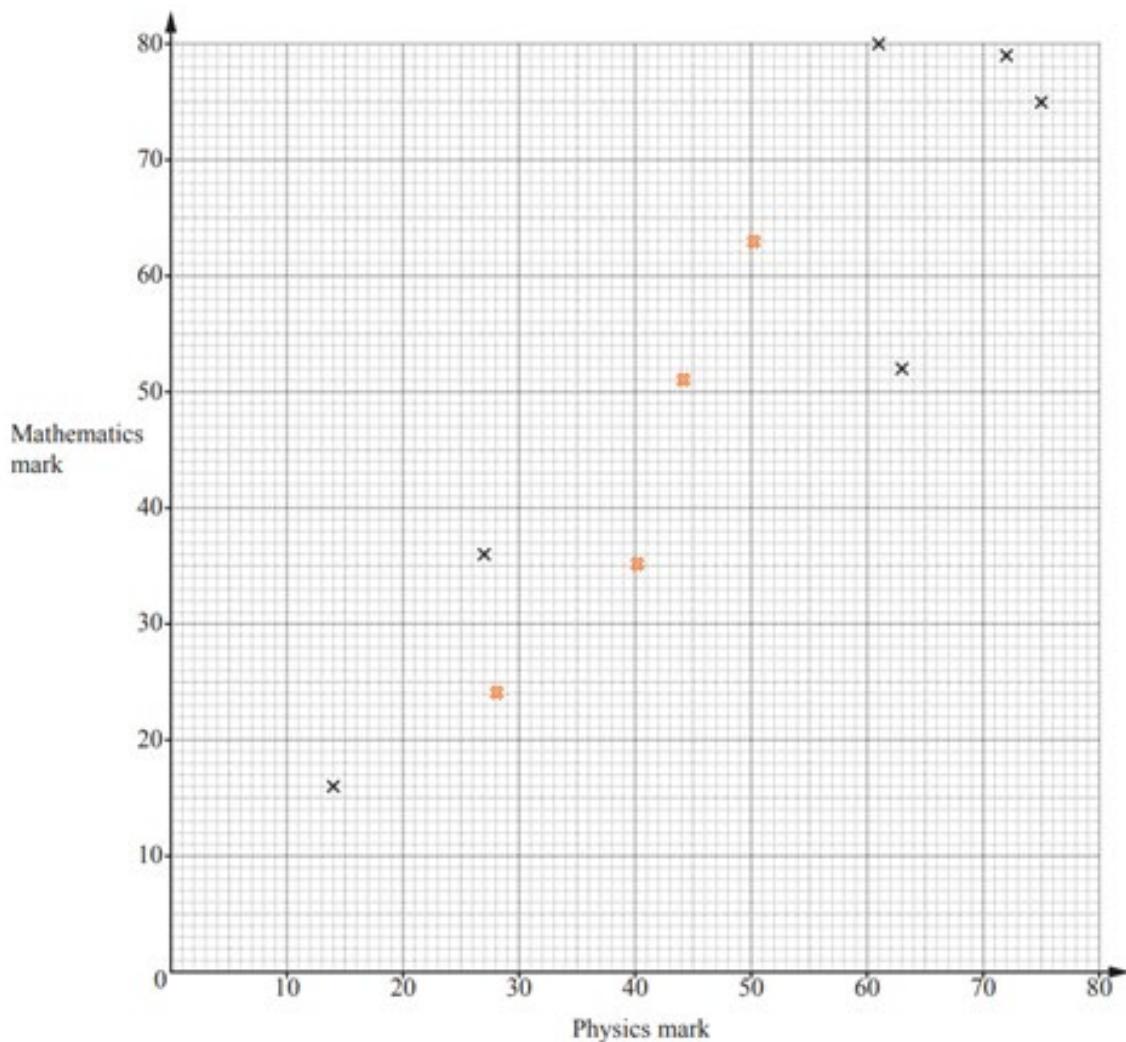
Question 5

(a) The table shows the marks gained by 10 students in their physics test and their mathematics test.

Physics mark	63	61	14	27	72	75	44	40	28	50
Mathematics mark	52	80	16	36	79	75	51	35	24	63

- (i) Complete the scatter diagram below.
The first six points have been plotted for you.

[2]



- (ii) What type of correlation is shown in the scatter diagram?

Positive correlation.

[1]

(b) The marks of 30 students in a spelling test are shown in the table below.

Mark	0	1	2	3	4	5
Frequency	2	4	5	5	6	8

Find the mean, median, mode and range of these marks.

[7]

The mean is

$$\frac{0 \times 2 + 1 \times 4 + 2 \times 5 + 3 \times 5 + 4 \times 6 + 5 \times 8}{30}$$

$$\text{mean} = 3.1$$

The mode is the number that occurs the most

$$\text{mode} = 5$$

The median is the average of the middle marks (15th and 16th)

$$\frac{3 + 3}{2}$$

$$\text{median} = 3$$

The range is the largest minus the smallest mark

$$5 - 0$$

$$\text{range} = 5$$

- (c) The table shows the marks gained by some students in their English test.

Mark	52	75	91
Number of students	x	45	11

The mean mark for these students is 70.3 .

Find the value of x .

[3]

We have that

$$\frac{52x + 75 \times 45 + 91 \times 11}{x + 45 + 11} = 70.3$$

$$\rightarrow \frac{52x + 4376}{x + 56} = 70.3$$

Multiply through by $(x + 56)$

$$\rightarrow 52x + 4376 = 70.3x + 3936.8$$

Now move x to one side and the numbers to the other

$$\rightarrow (70.3 - 52)x = 4376 - 3936.8$$

$$\rightarrow 18.3x = 439.2$$

$$\rightarrow x = 24$$

Question 6

The table shows information about the time taken by 400 people to complete a race.

Time taken (m minutes)	$45 < m \leq 50$	$50 < m \leq 60$	$60 < m \leq 70$	$70 < m \leq 90$	$90 < m \leq 100$	$100 < m \leq 120$
Frequency	23	64	122	136	26	29

- (a) Calculate an estimate of the mean timetaken. [4]

The mean can be estimated by taking the middle of each class, and multiplying by the number within each class

$$\text{mean} = \frac{(23 \times 47.5) + (64 \times 55) + (122 \times 65) + (136 \times 80) + (26 \times 95) + (29 \times 110)}{400}$$

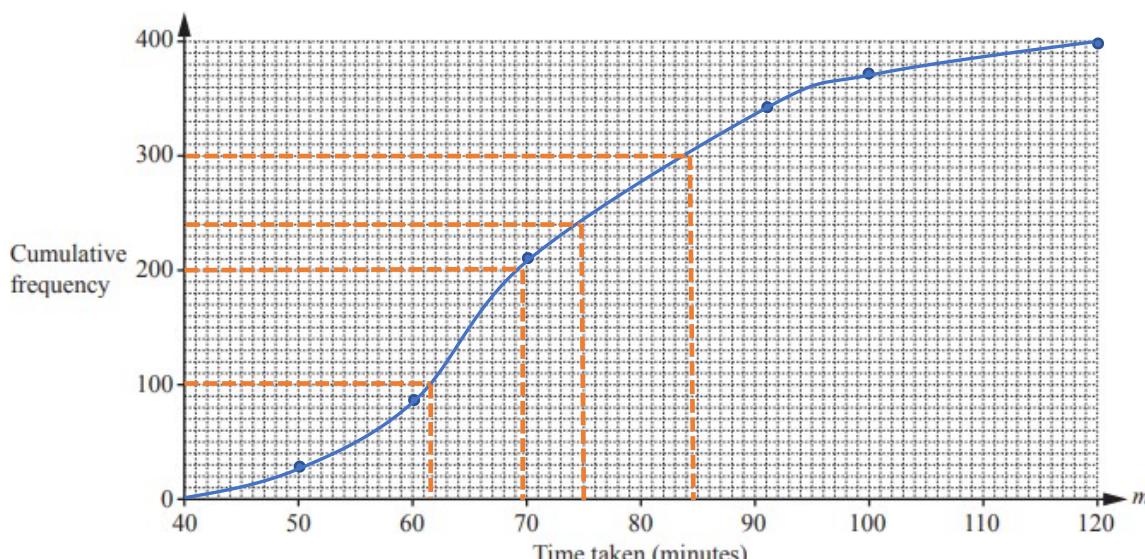
$$\text{mean} = 72.70625$$

$$\text{mean} = 72.7 \text{ mins (3.s.f)}$$

- (b) (i) Complete the cumulative frequency table. [2]

Time taken (m minutes)	$m \leq 50$	$m \leq 60$	$m \leq 70$	$m \leq 90$	$m \leq 100$	$m \leq 120$
Cumulative frequency	23	87	209	345	371	400

- (ii) On the grid, draw a cumulative frequency diagram to show this information. [3]



(iii) Use your diagram to estimate

(a) the median,

[1]

The median occurs when the middle (200th) person crosses the finish line. Reading across we can see this happens at about 69.5 minutes.

(b) the inter-quartile range,

[2]

The interquartile range is from the 100th to the 300th person, which we can see is 61.5minutes-84.5minutes, a range of 23 minutes.

(c) the 60th percentile.

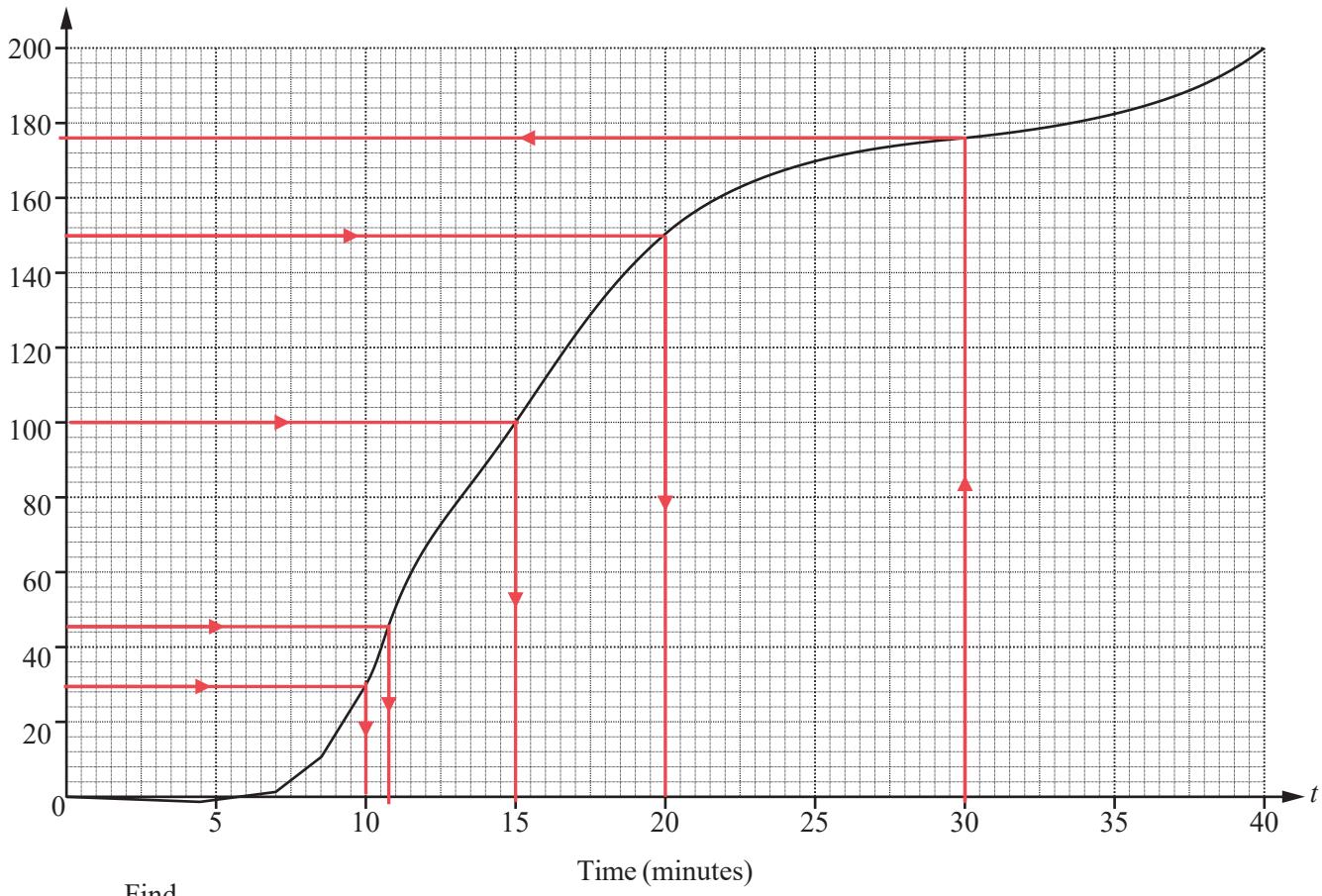
[2]

The 60th percentile is when the 240th person crosses the line, which happens at 75 minutes.

Question 7

- (a) 200 students record the time, t minutes, for their journey from home to school.
The cumulative frequency diagram shows the results.

Cumulative
frequency



Find

- (i) the median, [1]

Read across and down from 50% = 100: **Median = 15**

- (ii) the lower quartile, [1]

Read across and down from 25% = 50: **Lower Quartile = 10.75**

- (iii) the inter-quartile range, [1]

For Upper Quartile, read across and down from 75% = 150: **Upper Quartile = 20**

$$\text{Inter-Quartile Range} = \text{Upper Quartile} - \text{Lower Quartile}$$

$$\text{Inter-Quartile Range} = 20 - 10.75 = 9.25$$

- (iv) the 15th percentile, [1]

Read across and down from 15% = 30: **15th Percentile = 10**

- (v) the number of students whose journey time was more than 30 minutes.

[2]

Read up and across from 30 minutes to get:

$$\text{Number with a journey time LESS THAN } 30 \text{ minutes} = 176$$

$$\text{Number with a journey time MORE THAN } 30 \text{ minutes} = 200 - 176$$

$$\text{Number with a journey time more than } 30 \text{ minutes} = 24$$

- (b) The 200 students record the time, t minutes, for their journey from school to home.

The frequency table shows the results.

Time (t minutes)	$0 < t \leq 10$	$10 < t \leq 15$	$15 < t \leq 20$	$20 < t \leq 30$	$30 < t \leq 60$
Frequency	48	48	60	26	18

$$FD = \frac{\text{Frequency}}{\text{Class Width}} \quad \frac{48}{10} = 4.8 \quad \frac{48}{5} = 9.6 \quad \frac{60}{5} = 12 \quad \frac{26}{10} = 2.6 \quad \frac{18}{30} = 0.6$$

- (i) Calculate an estimate of the mean.

[4]

Use the Mid-Interval Values as the x values (t values) in the formula for the mean:

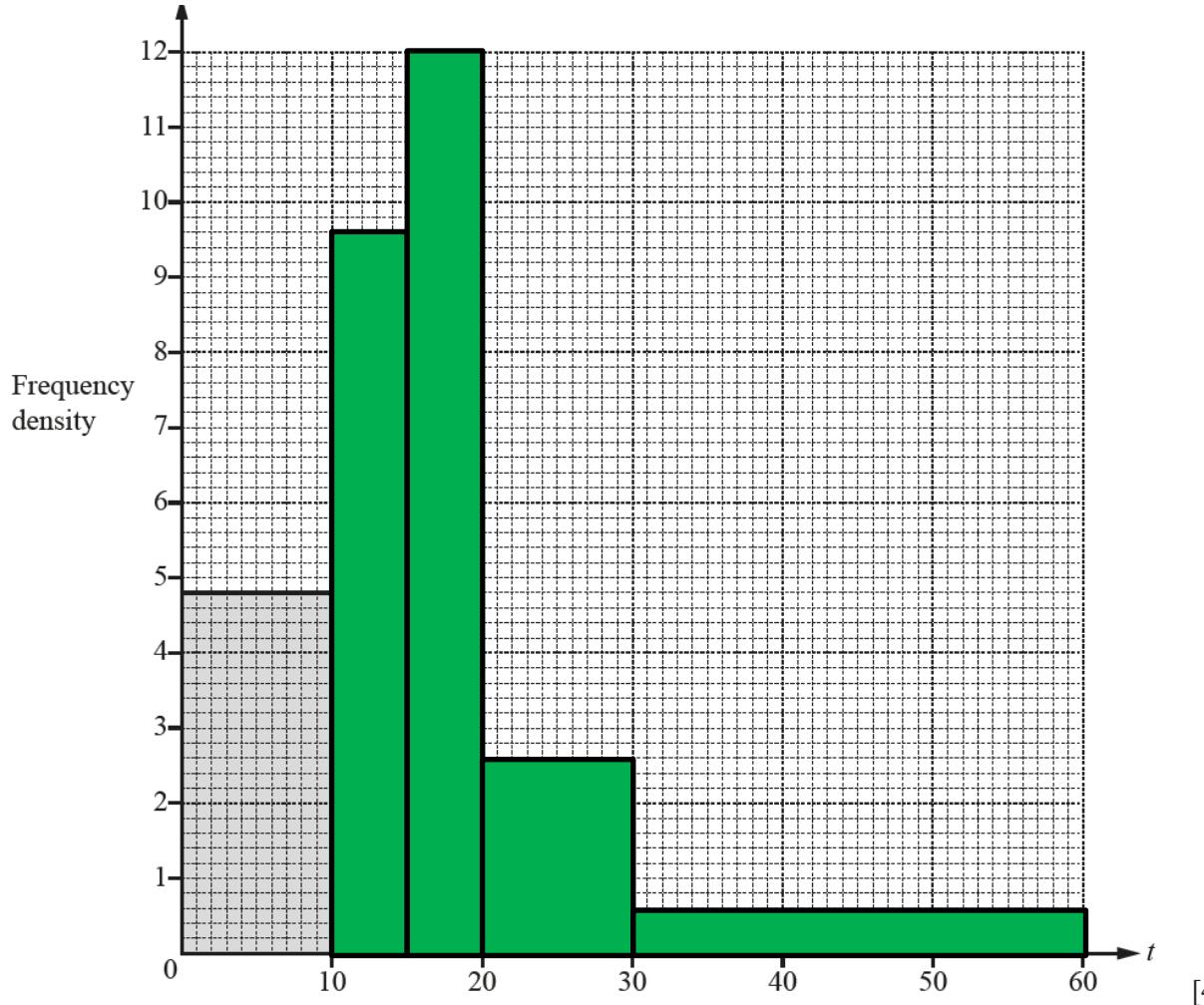
$$\text{Mean} = \frac{\sum fx}{n}$$

$$\text{Mean} = \frac{48 \times 5 + 48 \times 12.5 + 60 \times 17.5 + 26 \times 25 + 18 \times 45}{200}$$

$$\text{Mean} = \frac{3350}{200} = 16.75 \text{ minutes}$$

- (ii) On the grid, complete the histogram to show the information in the frequency table.

Add a Frequency Density (*FD*) row to the table above and plot appropriate bars on the grid:



Statistics

Difficulty: Hard

Model Answers 2

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Hard
Booklet	Model Answers 2

Time allowed: 105 minutes

Score: /91

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

200 people run 10 km.

The table shows some information about the times, t minutes, taken to run the 10 km.

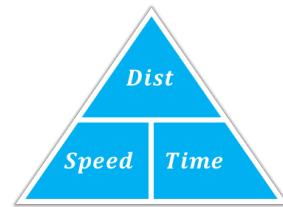
Time (t minutes)	$30 < t \leq 40$	$40 < t \leq 45$	$45 < t \leq 50$	$50 < t \leq 55$	$55 < t \leq 60$	$60 < t \leq 80$
Frequency	8	22	95	55	14	6

$$FD = \frac{\text{Frequency}}{\text{Class Width}} \quad \frac{8}{10} = 0.8 \quad \frac{22}{5} = 4.4 \quad \frac{95}{5} = 19 \quad \frac{55}{5} = 11 \quad \frac{14}{5} = 2.8 \quad \frac{6}{10} = 0.6$$

(a) Howard takes 40 minutes to run the 10 km.

Calculate his average speed in kilometres per hour.

$$\text{Speed} = \frac{\text{Dist}}{\text{Time}} = \frac{10}{\frac{2}{3}} = 15 \text{ kmh}$$



[2]

(b) Calculate an estimate of the mean time.

[4]

Use the Mid-Interval Values as the x values (t values) in the formula for the mean:

$$\text{Mean} = \frac{\sum fx}{n}$$

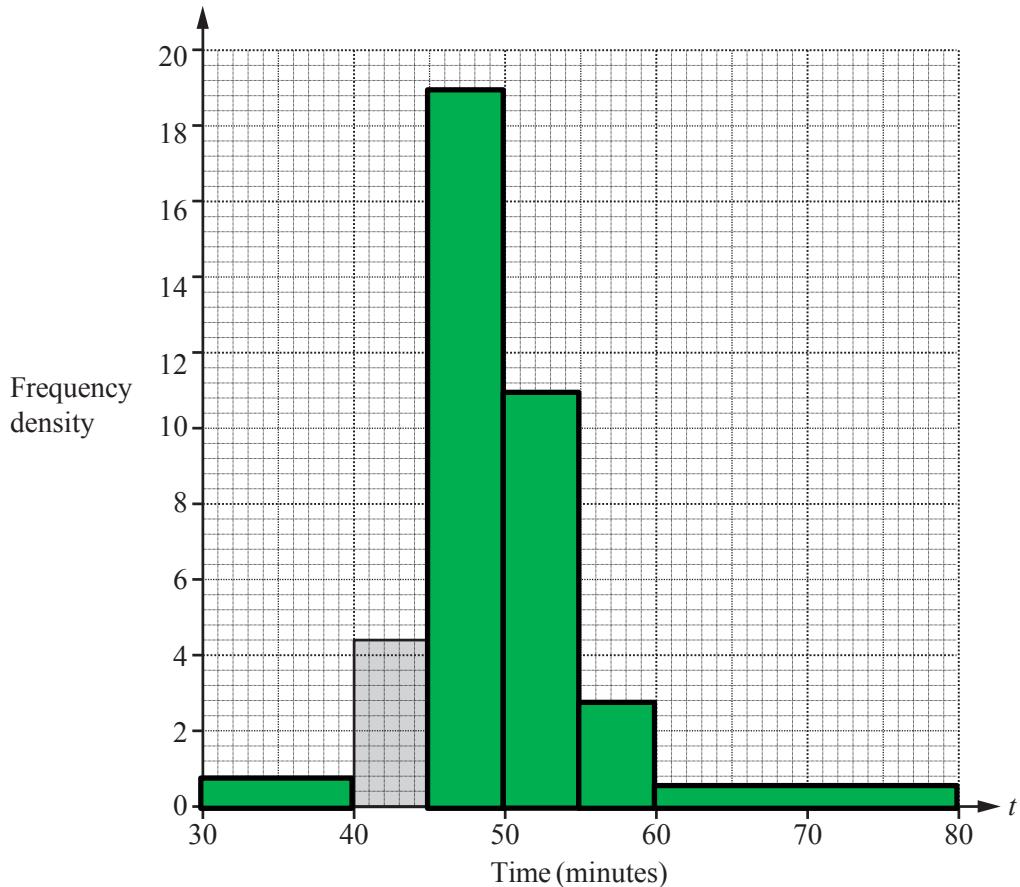
$$\text{Mean} = \frac{8 \times 35 + 22 \times 42.5 + 95 \times 47.5 + 55 \times 52.5 + 14 \times 57.5 + 6 \times 70}{200}$$

$$\text{Mean} = \frac{9840}{200} = 49.2 \text{ minutes}$$

- (c) Complete the histogram to show the information in the table.

[4]

Add a Frequency Density (FD) row to the table above and plot appropriate bars on the grid:

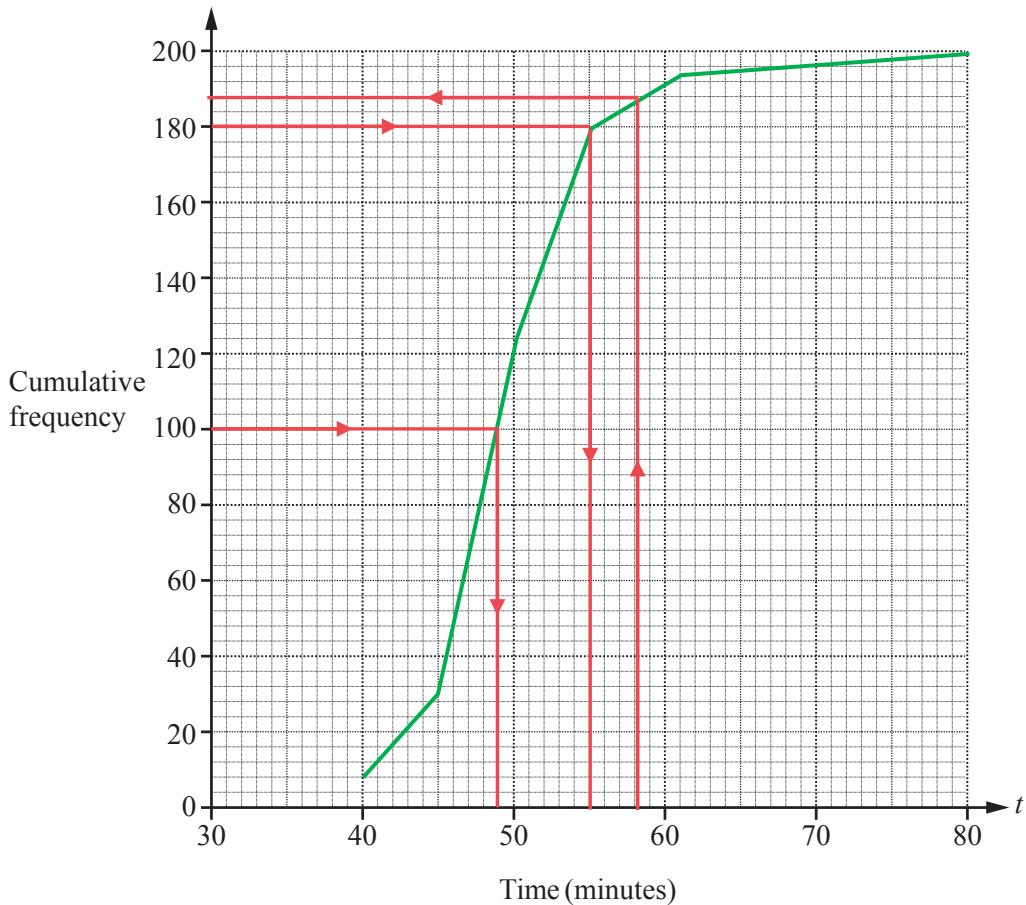


- (d) (i) Use the frequency table opposite to complete the cumulative frequency table.

Time (t minutes)	$t \leq 40$	$t \leq 45$	$t \leq 50$	$t \leq 55$	$t \leq 60$	$t \leq 80$
Cumulative frequency	8	30	$30 + 95 = 125$	$125 + 55 = 180$	194	200

[1]

- (ii) Draw a cumulative frequency diagram to show the information in the table above.



[3]

- (iii) Use your diagram to find

- (a) the median,

[1]

Read across and down from 50% = 100: Median = 49 minutes

- (b) the 90th percentile,

[1]

Read across and down from 90% = 180: 90th Percentile = 55 minutes

- (c) the number of people who took more than 58 minutes to run the 10 km.

[2]

Read up and across from 58 minutes to get:

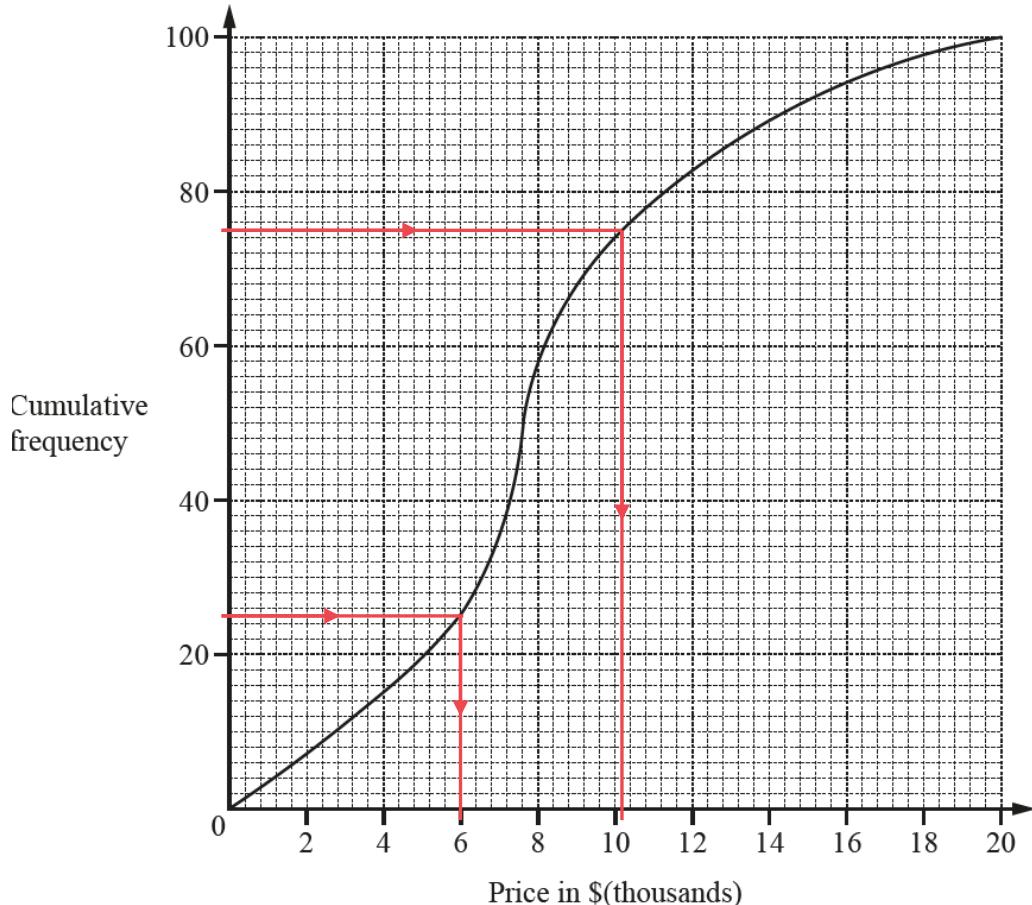
Number with a journey time LESS THAN 58 minutes = 190

Number with a journey time MORE THAN 58 minutes = 200 – 190

Number with a journey time more than 58 minutes = 10

Question 2

(a) (i)



The cumulative frequency diagram shows information about the prices of 100 cars on Website A. Use the information to complete this table.

Read across and down from 25% and 75% to get LQ and UQ. Then use $IQR = UQ - LQ$.

Lower quartile	Median	Upper quartile	Inter-quartile range
\$6000	\$7600	\$10200	\$4200

[2]

(ii) This table shows information about the prices of cars on Website B.

Lower quartile	Median	Upper quartile	Inter-quartile range
\$7600	\$10800	\$13600	\$6000

Here are two statements comparing the distributions of the prices of cars on Website A and Website B.

For each statement write True or False.

Give a reason for each answer, stating clearly which statistic you use to make your decision.

(a) The prices of cars on Website A are lower than the prices of cars on Website B.

True because the Median price is lower on Website A [1]

- (b) A greater percentage of cars have a price more than \$13600 on Website A compared to Website B.

False because **the Upper Quartile on Website A is less than \$13600** [1]

- (b) The table shows the prices of cars on Website B.

Price (\$P)	Number of cars
$0 < P \leq 6000$	9
$6000 < P \leq 8000$	29
$8000 < P \leq 10000$	20
$10000 < P \leq 12000$	14
$12000 < P \leq 14000$	21
$14000 < P \leq 22000$	27

Calculate an estimate of the mean price of the 120 cars.

[4]

Use the Mid-Interval Values as the x values (P values) in the formula for the mean:

$$\text{Mean} = \frac{\sum fx}{n}$$

Mean

$$= \frac{9 \times 3000 + 29 \times 7000 + 20 \times 9000 + 14 \times 11000 + 21 \times 13000 + 27 \times 18000}{9 + 29 + 20 + 14 + 21 + 27}$$

$$\text{Mean} = \frac{1323000}{120} = \$11025$$

- (c) The price of a car is \$8760.

Bryan pays a deposit of 25% of this price and then 24 equal monthly payments.
After 24 months, he will have paid a total of \$9948.

Calculate the cost of one monthly payment.

[3]

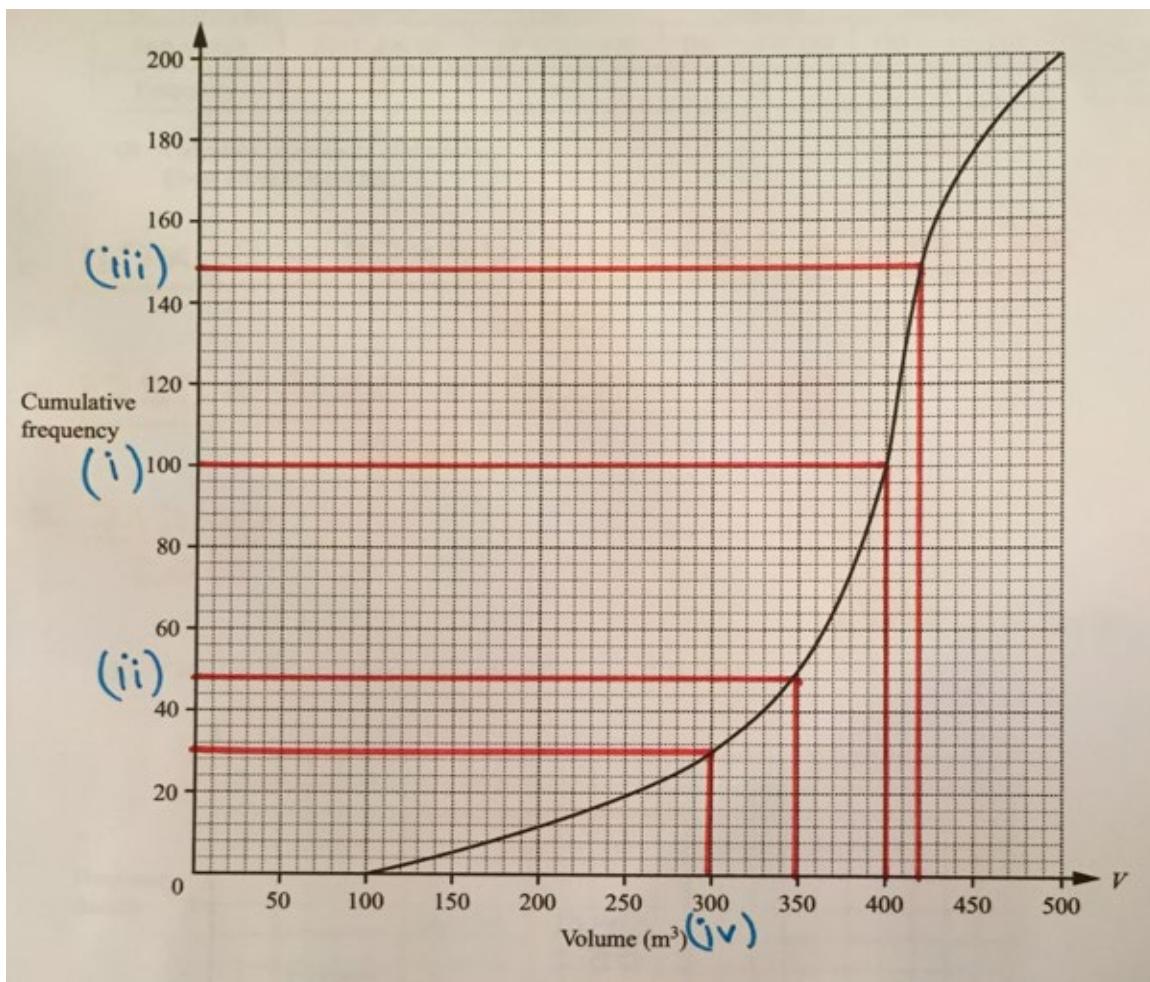
$$\text{Deposit} = 25\% \text{ of } \$8760 = 0.25 \times 8760 = \$2190$$

$$\text{Balance to pay} = 9948 - 2190 = \$7758$$

$$\text{One monthly payment} = \frac{7758}{24} = \$323.35$$

Question 3

- (a) 200 students estimate the volume, V m³, of a classroom.
The cumulative frequency diagram shows their results.



Find

- (i) the median,

[1]

Find the position of the median:

$$200 \div 2 = 100$$

Trace across from 100 on y-axis and read result off x-axis:

400

(ii) the lower quartile,

[1]

Find position of the lower quartile:

$$200 \div 4 = 50$$

Trace across from 50 on the y-axis and read result off x-axis:

350

(iii) the inter-quartile range,

[1]

Find position of the upper quartile:

$$(200 \div 4) \times 3 = 150$$

Trace across from 150 on y-axis and read result off x-axis:

420

Interquartile range = upper quartile – lower quartile

$$420 - 350 = 70$$

(iv) the number of students who estimate that the volume is greater than 300m³.

[2]

Trace across from 300 on x-axis and read result off y-axis:

30

$$200 - 30 = 170$$

- (b) The 200 students also estimate the total area, $A \text{ m}^2$, of the windows in the classroom.
 The results are shown in the table.

Area ($A \text{ m}^2$)	$20 < A \leq 60$	$60 < A \leq 100$	$100 < A \leq 150$	$150 < A \leq 250$
Frequency	32	64	80	24

- (i) Calculate an estimate of the mean.
 Show all your working.

[4]

Mid Points

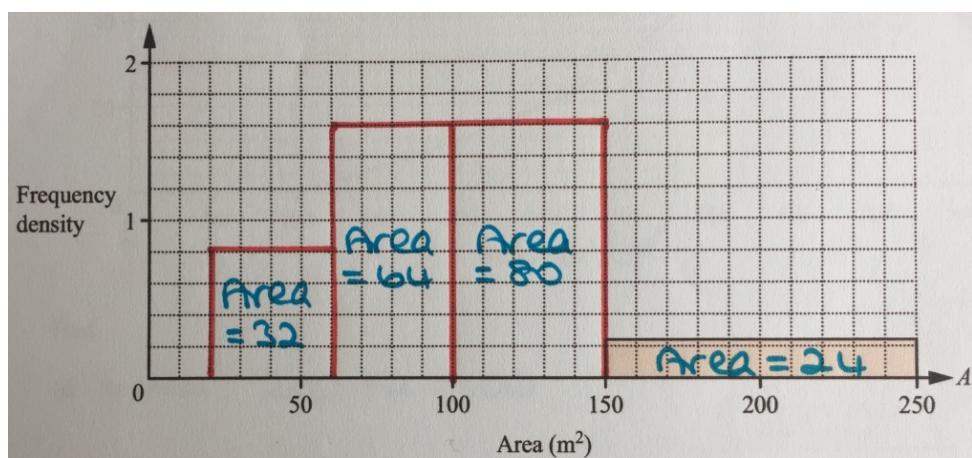
$$\text{Mean} = \frac{40 \times 32 + 80 \times 64 + 125 \times 80 + 200 \times 24}{200}$$

$$= \frac{21200}{200}$$

$$= 106$$

- (ii) Complete the histogram to show the information in the table.

[4]



- (iii) Two of the 200 students are chosen at random.

Find the probability that they both estimate that the area is greater than 100m^2 .

[2]

Student 1 Student 2

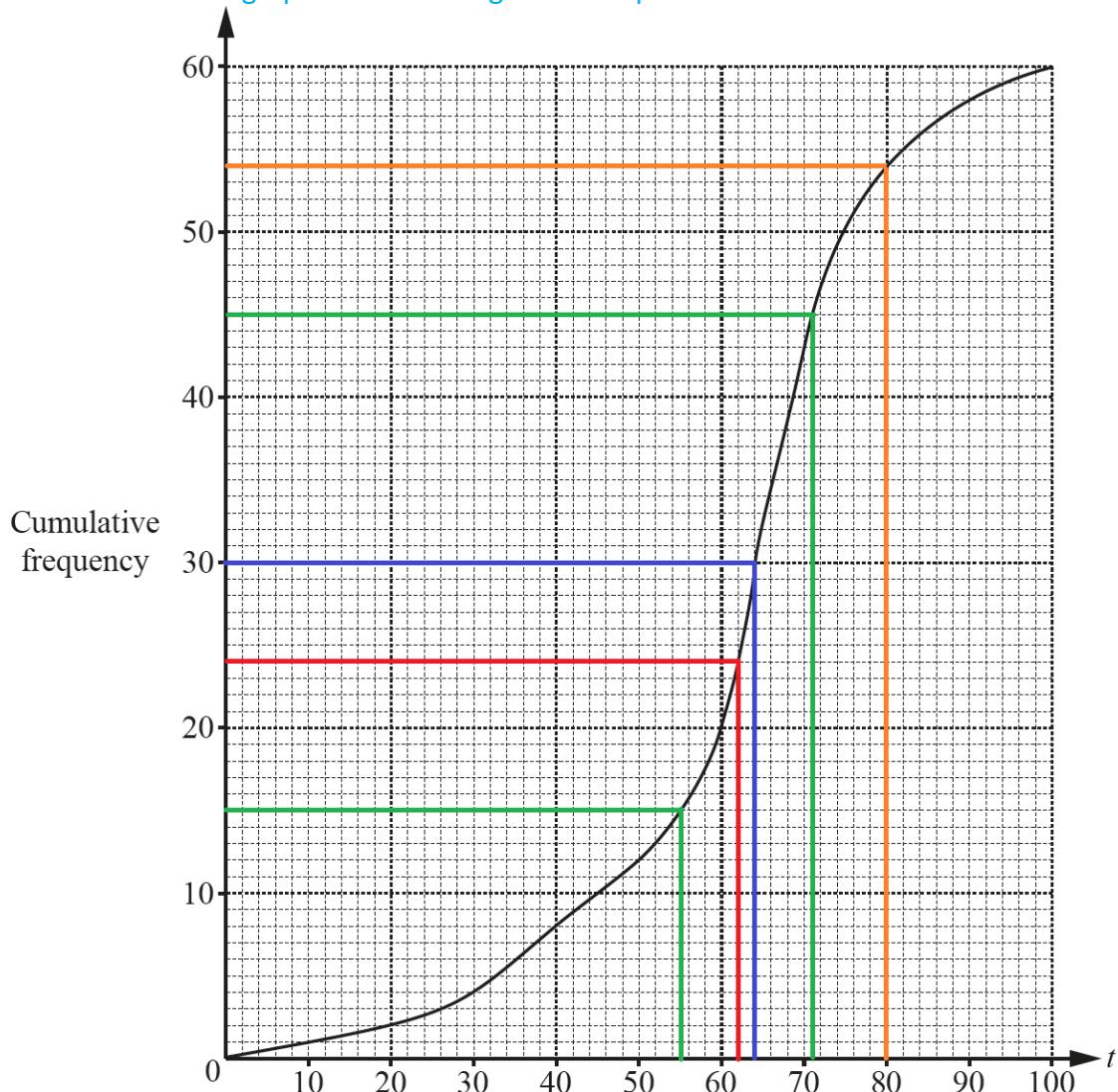
$$\text{Probability} = \frac{104}{200} \times \frac{103}{199}$$

$$= \frac{10712}{39800}$$

Question 4

The cumulative frequency diagram shows information about the time taken, t minutes, by 60 students to complete a test.

We will refer to the graph below throughout this question.



(a) Find

(i) the median,

[1]

The median is the 50th percentile. As there are 60 students in total, it is the time taken of 30th student (blue lines). From the graph, this is

64 minutes.

- (ii) the inter-quartile range,

[2]

The inter-quartile range is the difference between the upper and the lower quartile. The lines representing these values are green.

The lower quartile is the value of student who is 25% of 60, which is 15.

This is 55 min.

The upper quartile is the value of student who is 75% of 60, which is 45.

This is 71 min.

The inter-quartile range is therefore

$$\mathbf{71\text{min} - 55\text{min} = 16 \text{ minutes}.}$$

- (iii) the 40th percentile,

[2]

The 40th percentile is the value of student who is 40% of 60, which is 24 (follow the red line).

$$\mathbf{40\text{th percentile} = 62 \text{ minutes}.}$$

- (iv) the number of students who took more than 80 minutes to complete the test.

[2]

To find the number of student who took 80 minutes or less to complete the test, we follow the orange line.

$$(< 80 \text{ min}) = 54 \text{ students}$$

Now we know that there are 54 students who took 80 minutes or less. Subtract this from the total number of students to get the number of students who took more than 80 minutes.

$$(> 80 \text{ min}) = \text{total} - (< 80 \text{ min})$$

$$(> 80 \text{ min}) = 60 - 54$$

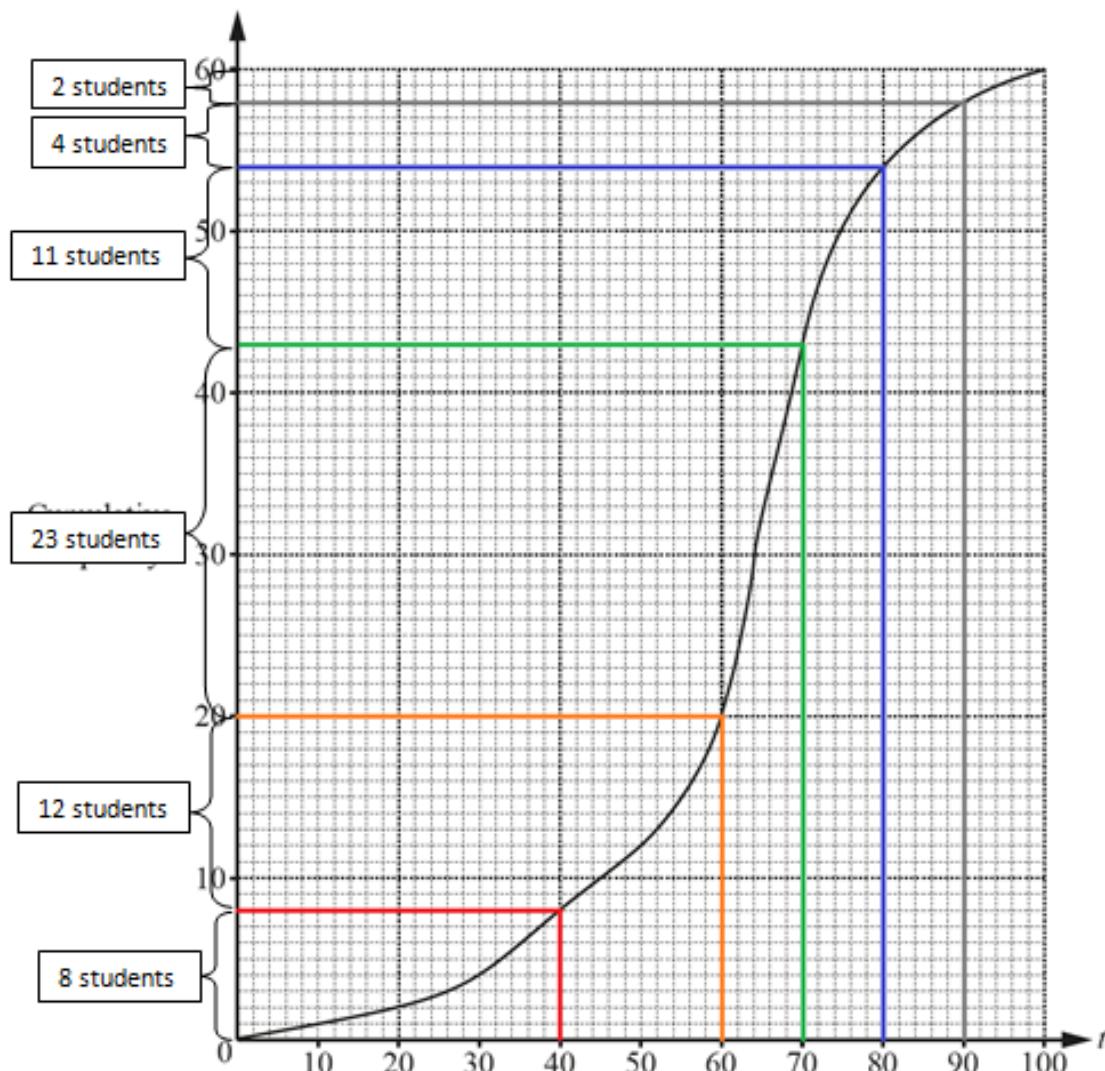
$$(> 80 \text{ min}) = 6 \text{ students}$$

- (b) Use the cumulative frequency diagram to complete the frequency table below.

Time taken (t minutes)	$0 < t \leq 40$	$40 < t \leq 60$	$60 < t \leq 70$	$70 < t \leq 80$	$80 < t \leq 90$	$90 < t \leq 100$
Frequency	8				4	

Draw lines according to the boundaries set in the frequency table.

Calculate the number of students between each boundary.



Fill the frequency table:

Time taken (t minutes)	$0 < t \leq 40$	$40 < t \leq 60$	$60 < t \leq 70$	$70 < t \leq 80$	$80 < t \leq 90$	$90 < t \leq 100$
Frequency	8	12	23	11	4	2

- (c) On the grid below, complete the histogram to show the information in the table in **part (b)**. [4]

The frequency density is calculated using the previous table as the ratio of the frequency and the range of the corresponding interval (difference between the upper and the lower bounds).

$$\text{frequency density} = \frac{\text{frequency}}{\text{range of values}}$$

For the second interval, the frequency is 12 and the range of values is 20min (allowing values from 40min and 60min so range is the difference of these two values). The frequency density:

$$\text{frequency density } (40 < t \leq 60) = \frac{12}{20}$$

$$\text{frequency density } (40 < t \leq 60) = 0.6 \quad (\text{red rectangle})$$

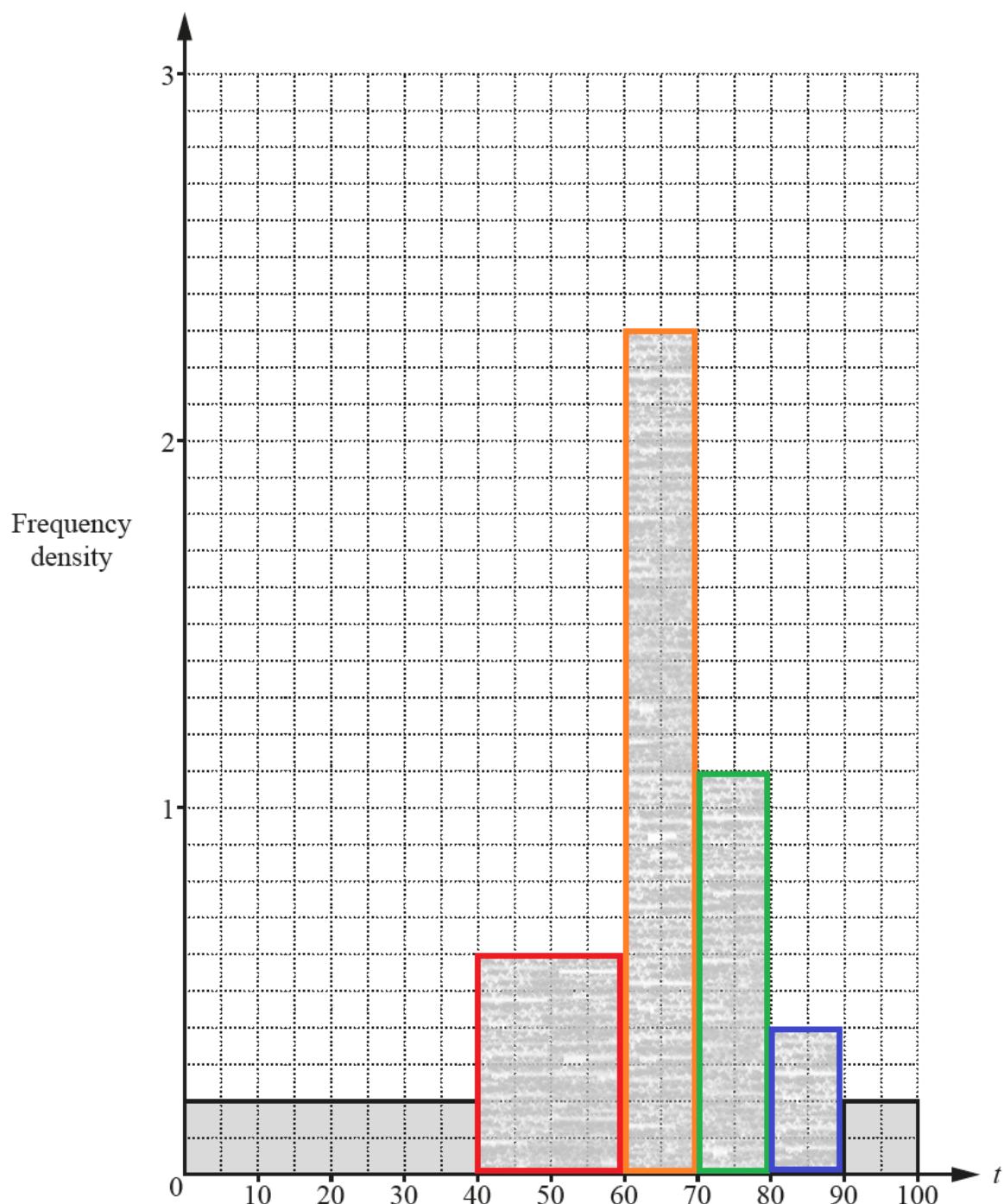
Similarly for the other intervals:

$$\text{frequency density } (60 < t \leq 70) = \frac{23}{10} = 2.3 \quad (\text{orange rectangle})$$

$$\text{frequency density } (70 < t \leq 80) = \frac{11}{10} = 1.1 \quad (\text{green rectangle})$$

$$\text{frequency density } (80 < t \leq 90) = \frac{4}{10} = 0.4 \quad (\text{blue rectangle})$$

We plot the frequency density into the histogram, making sure that the correct value is assigned to the corresponding mass. Height of the histogram changes when the interval changes!



Question 5

The table shows the time, t minutes, that 400 people take to complete a test.

Time taken (t mins)	$0 < t \leq 10$	$10 < t \leq 24$	$24 < t \leq 30$	$30 < t \leq 40$	$40 < t \leq 60$	$60 < t \leq 70$
Frequency	10	90	135	85	70	10

- (a) (i) Write down the modal time interval.

[1]

A modal interval is an interval with the highest frequency.

In our case, it is the interval $24 < t \leq 30$

- (ii) Calculate an estimate of the mean time taken to complete the test.

[4]

The mean value is estimated by assuming that all people within a given group have equal time taken to complete a test to the mid-value of that group (i.e. 5min, 17min, 27min, 35min, 50min, 65min).

We sum the products of the mid-value and the frequency of given group and then divide by the total number of people to get the mean estimate. The mean is therefore:

$$\text{mean} = \frac{5 \times 10 + 17 \times 90 + 27 \times 135 + 35 \times 85 + 50 \times 70 + 65 \times 10}{400}$$

$$\text{mean} = 30.9 \text{ min}$$

- (b) (i) Complete the table of cumulative frequencies. [2]

Time taken (t mins)	$t \leq 10$	$t \leq 24$	$t \leq 30$	$t \leq 40$	$t \leq 60$	$t \leq 70$
Cumulative frequency	10	100				400

For example, to get the cumulative frequency for $t \leq 30$, we sum the corresponding frequencies (up to $24 < t \leq 30$) from the previous table.

$$t \leq 30 = 10 + 90 + 135 = 235$$

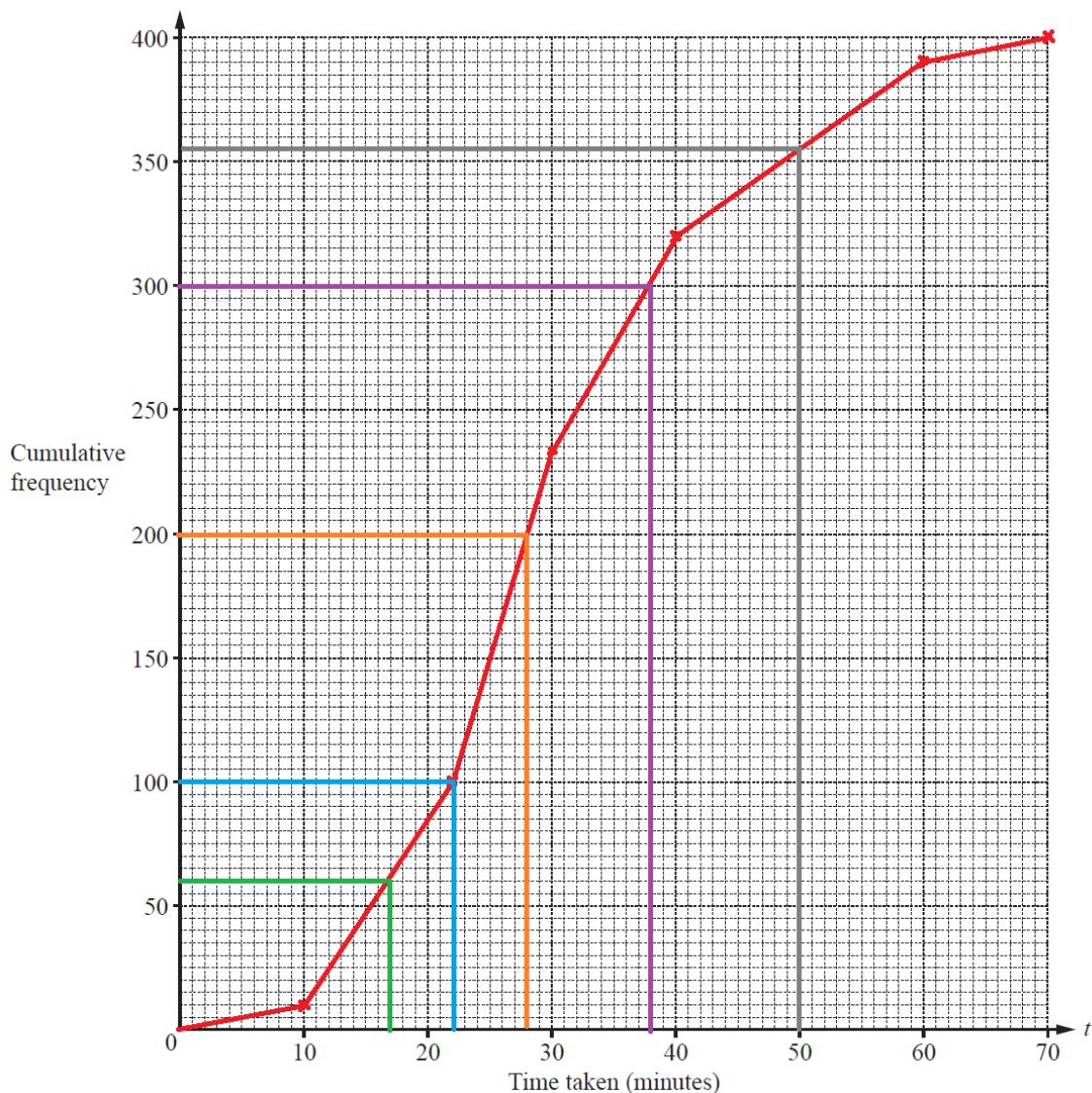
The cumulative frequency for $t \leq 30$ is 235.

We use the same method to complete the cumulative frequency table.

$$t \leq 30 = 235 \quad t \leq 40 = 320 \quad t \leq 60 = 390$$

- (ii) On the grid opposite, draw a cumulative frequency diagram to show this information. [3]

Plot the points from the table on the grid and connect them with a curve (red).



(c) Use your graph to estimate

- (i) the median time, [1]

The median is the 50th percentile. As there are 400 people, it is the time taken of 200th person (orange lines). From the graph, this is **28 minutes**.

- (ii) the inter-quartile range, [2]

The inter-quartile range is the difference between the upper and the lower quartile.

The lower quartile is the value of student who is 25% of 400, which is 100. This is 22 min.

(blue line)

The upper quartile is the value of student who is 75% of 400, which is 300. This is 38 min.

(purple)

The inter-quartile range is therefore **16 minutes**.

- (iii) the 15th percentile, [2]

We calculate 15% of 400 (total number of people). This is 60.

Therefore we are looking for the value of 60th reading (green lines).

From the graph, the 60th percentile is **17 minutes**.

- (iv) the number of people who took more than 50 minutes. [2]

From the graph, there are 355 people who took 50 minutes or less to complete the test

(grey).

Because the total number of people who took the test was 400, there are **45 people** who took more than 50 minutes.

Question 6

- (a) A group of 50 students estimated the mass, M grams, of sweets in a jar.
The results are shown in the table.

Mass (M grams)	Number of students
$0 < M \leq 200$	5
$200 < M \leq 300$	9
$300 < M \leq 350$	18
$350 < M \leq 400$	12
$400 < M \leq 500$	6

- (i) Calculate an estimate of the mean. [4]

The mean value is estimated by assuming that all students within a given group have mass of sweets equal to the mid-value of that group (i.e. 100g, 250g, 325g, 375g, 450g).

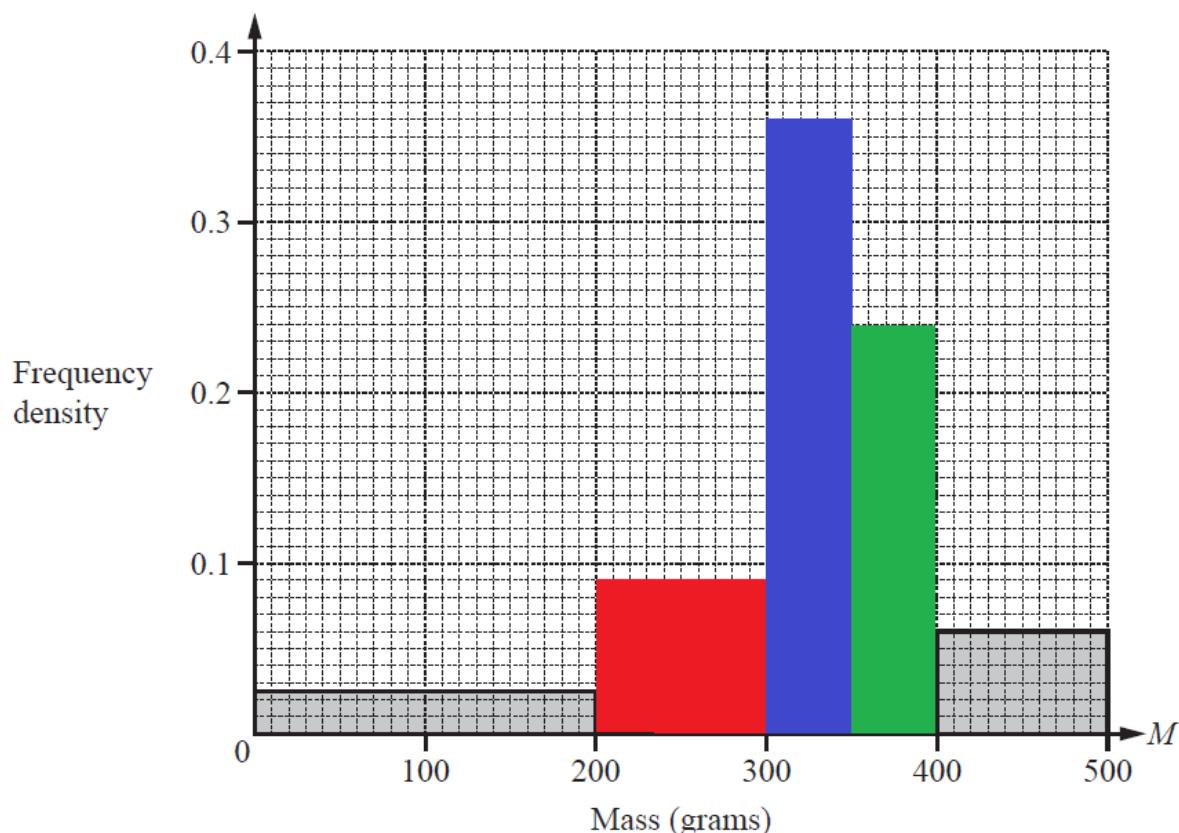
We sum the products of the mid-value and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

$$\text{mean} = \frac{100g \times 5 + 250g \times 9 + 325g \times 18 + 375g \times 12 + 450g \times 6}{50 \text{ students}}$$

$$\text{mean} = 316g$$

- (ii) Complete this histogram to show the information in the table.

[3]



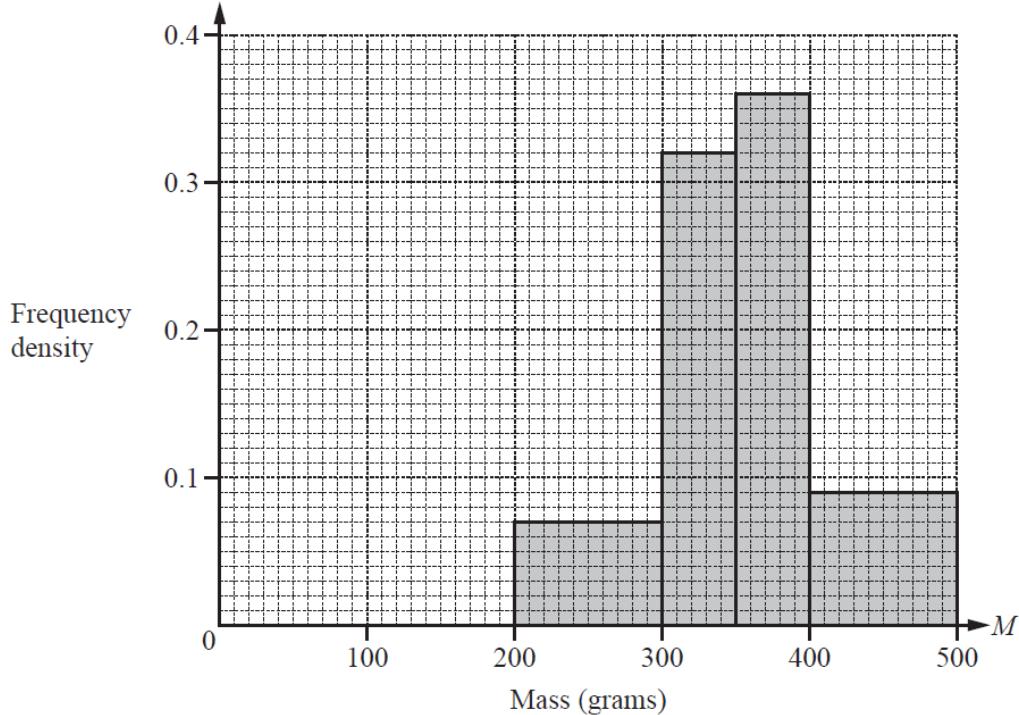
To get the right histogram, each bar has to have an area equal to the frequency of the given group.

- The width of the second group is 100 and the frequency is 9, so the height of the bar is $9/100=0.09$ (red)
- The width of the third group is 50 and the frequency is 18, so the height of the bar is $18/50=0.36$. (blue)
- The width of the fourth group is 50 and the frequency is 12, so the height of the bar is $12/50=0.24$. (green)

- (b) A group of 50 adults also estimated the mass, M grams, of the sweets in the jar.
The histogram below shows information about their estimates.

Use the histograms to make two comparisons between the distributions of the estimates of the students and the adults.

[2]



- Students have a greater range of estimates (as some of them also estimated the mass to be in the range $0 < M \leq 200\text{g}$)
- On average, adults estimated a greater mass of the sweets in the jar

Question 7

The table shows the times, t minutes, taken by 200 students to complete an IGCSE paper.

Time (t minutes)	$40 < t \leq 60$	$60 < t \leq 70$	$70 < t \leq 75$	$75 < t \leq 90$
Frequency	10	50	80	60

(a) By using mid-interval values, calculate an estimate of the mean time. [3]

The mean value is estimated by assuming that the time taken to complete an IGCSE paper for all students within a given group was equal to the mid-value of that group (i.e. 50min, 65min, 72.5min, 82.5min).

We sum the products of the mid-value and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

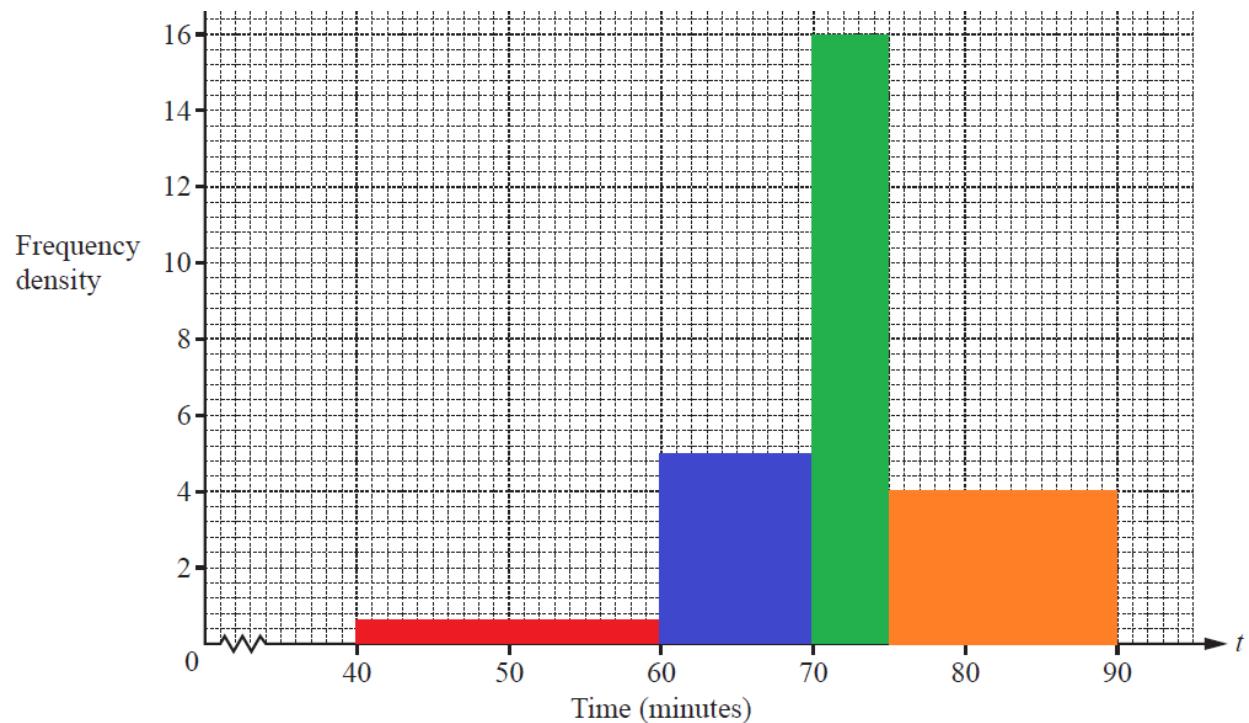
$$\text{mean} = \frac{50\text{min} \times 10 + 65\text{min} \times 50 + 72.5\text{min} \times 80 + 82.5\text{min} \times 60}{200 \text{ students}}$$

mean = 72.5min

(b) On the grid, draw a histogram to show the information in the table. [4]

To get the right histogram, each bar needs to have an area equal to the frequency of the given group.

- The width of the first group is 20 and the frequency is 10, so the height of the bar is $10/20=0.5$ (red)
- The width of the second group is 10 and the frequency is 50, so the height of the bar is $50/10=5.0$. (blue)
- The width of the third group is 5 and the frequency is 80, so the height of the bar is $80/5=16$. (green)
- The width of the forth group is 15 and the frequency is 60, so the height of the bar is $60/15=4.0$. (orange)



Statistics

Difficulty: Hard

Model Answers 3

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Hard
Booklet	Model Answers 3

Time allowed: 90 minutes

Score: /78

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

The table shows the height, h cm, of 40 children in a class.

Height (h cm)	$120 < h \leq 130$	$130 < h \leq 140$	$140 < h \leq 144$	$144 < h \leq 150$	$150 < h \leq 170$
Frequency	3	14	4	6	13

- (a) Write down the class interval containing the median.

[1]

The median is reached, when we take into the account more (or exactly) 50% of the subjects.

There are 40 children, so 50% is 20 children.

We need to sum the frequency in first three groups, to get above 20, therefore the median is in the third group:

$$\mathbf{140 < h \leq 144}$$

- (b) Calculate an estimate of the mean height.

[4]

The mean value is estimated by assuming that all students in a given group have the mid-value of the height for that group (i.e. 125cm, 135 cm, 142 cm, 147 cm, 160cm).

We sum the products of the mid-height and the frequency of given group and then divide by the total number of students to get the mean estimate. The mean is therefore:

$$\text{mean} = \frac{125\text{cm} \times 3 + 135\text{cm} \times 14 + 142\text{cm} \times 4 + 147\text{cm} \times 6 + 160\text{cm} \times 13}{40 \text{ students}}$$

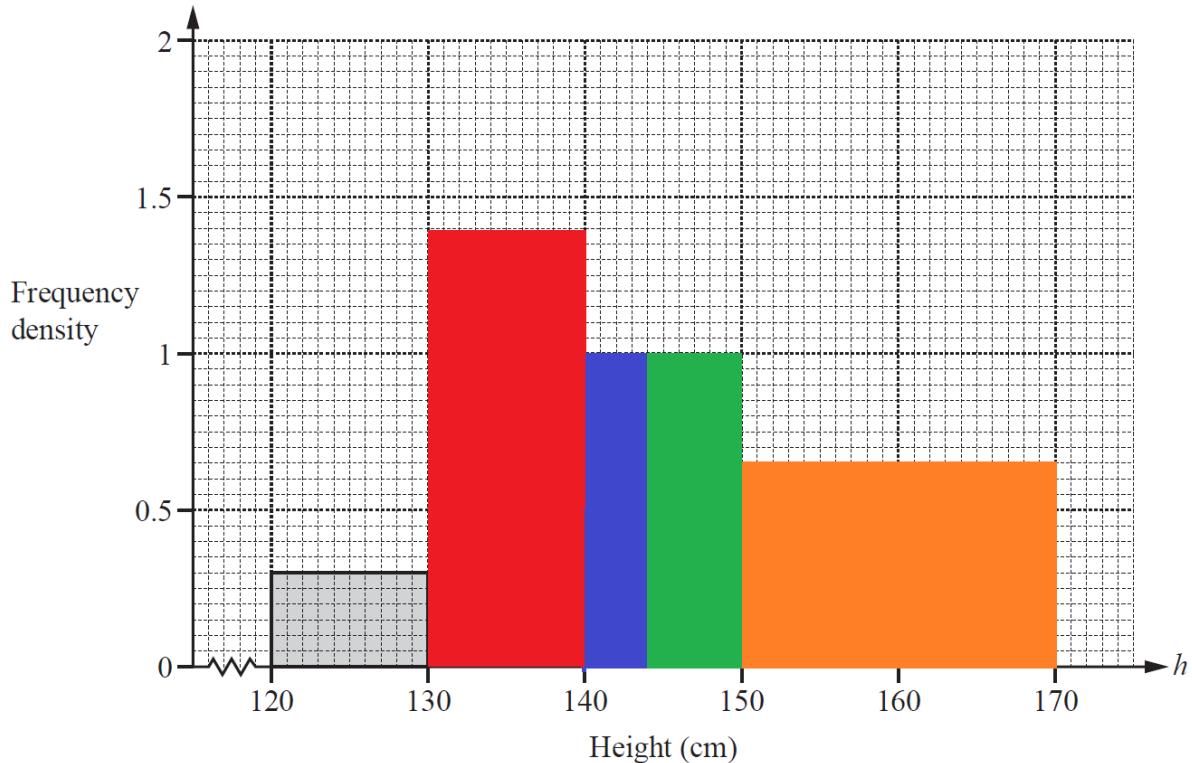
$$\text{mean} = \mathbf{144.875 \text{ cm}}$$

(c) Complete the histogram.

[4]

To get the right histogram, each bar has to have an area equal to the frequency of the given group.

- The width of the second group is 10 and the frequency is 14, so the height of the bar is $14/10=1.4$. (red)
- The width of the third group is 4 and the frequency is 4, so the height of the bar is $4/4=1.0$. (blue)
- The width of the fourth group is 6 and the frequency is 6, so the height of the bar is $6/6=1.0$. (green)
- The width of the fifth group is 20 and the frequency is 13, so the height of the bar is $13/20=0.65$. (orange)



Question 2

A company tested 200 light bulbs to find the lifetime, T hours, of each bulb. The results are shown in the table.

Lifetime (T hours)	Number of bulbs
$0 < T \leq 1000$	10
$1000 < T \leq 1500$	30
$1500 < T \leq 2000$	55
$2000 < T \leq 2500$	72
$2500 < T \leq 3500$	33

(a) Calculate an estimate of the mean lifetime for the 200 light bulbs.

[4]

Lifetime (T hours)	Number of bulbs (f)	Midpoint (M)	$f \times M$
$0 < T \leq 1000$	10	500	5000
$1000 < T \leq 1500$	30	1250	37500
$1500 < T \leq 2000$	55	1750	96250
$2000 < T \leq 2500$	72	2250	162000
$2500 < T \leq 3500$	33	3000	99000
Sum	200		399750

Estimate of the mean (μ) is

$$\mu = \frac{\sum f \times m}{\sum f}$$

$$= \frac{399750}{200}$$

$$= 1998.75$$

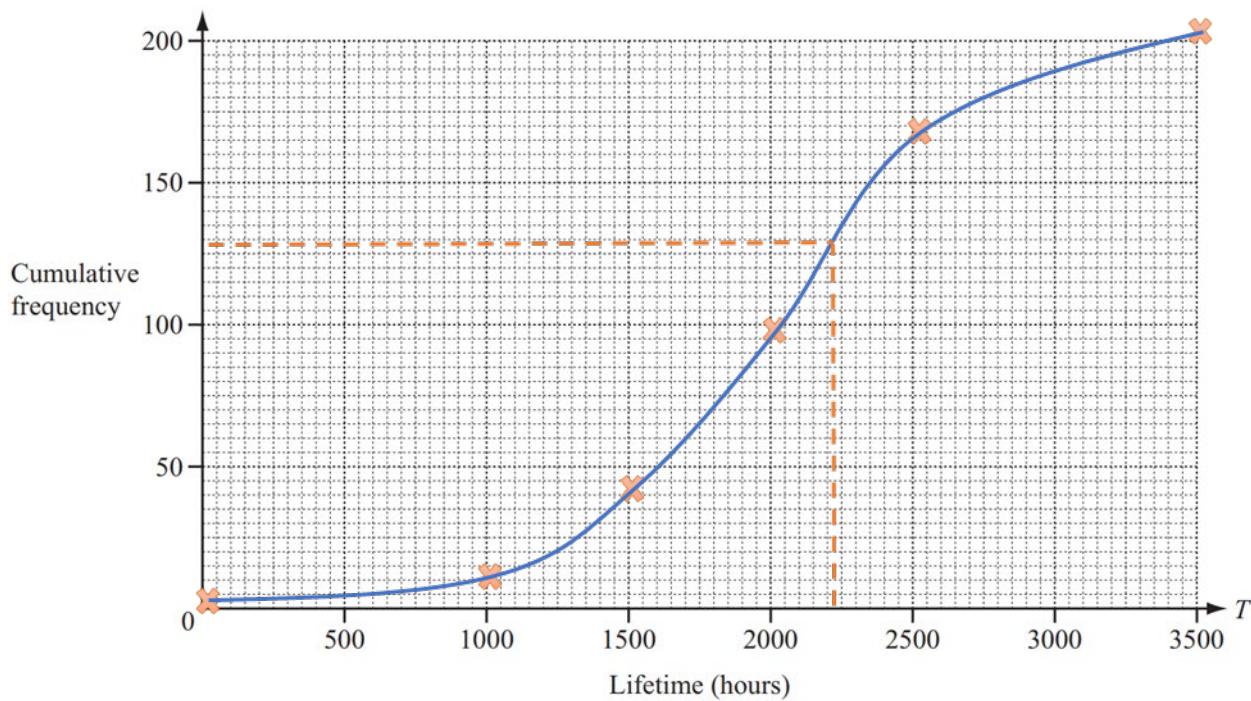
(b) (i) Complete the cumulative frequency table.

[2]

Lifetime (T hours)	$T \leq 1000$	$T \leq 1500$	$T \leq 2000$	$T \leq 2500$	$T \leq 3500$
Number of bulbs	10	40	95	167	200

(ii) On the grid, draw a cumulative frequency diagram to show this information.

[3]



(iii) The company says that the average lifetime of a bulb is 2200 hours.

Estimate the number of bulbs that lasted longer than 2200 hours.

[2]

Dashed orange line on graph above shows approximately

125 bulbs last 2200 hours or less, therefore:

$$200 - 125$$

$$= 75$$

(c) Robert buys one energy saving bulb and one halogen bulb.

The probability that the energy saving bulb lasts longer than 3500 hours is $\frac{9}{10}$.

The probability that the halogen bulb lasts longer than 3500 hours is $\frac{3}{5}$.

Work out the probability that exactly one of the bulbs will last longer than 3500 hours.

[4]

We need the probability that one *does* last longer

multiplied by the probability that the other *does not*:

$$\frac{9}{10} \times \frac{2}{5} + \frac{1}{10} \times \frac{3}{5}$$

$$= \frac{18}{50} + \frac{3}{50}$$

$$= \frac{21}{50}$$

Question 3

The time, t seconds, taken for each of 50 chefs to cook an omelette is recorded.

Time (t seconds)	$20 < t \leq 25$	$25 < t \leq 30$	$30 < t \leq 35$	$35 < t \leq 40$	$40 < t \leq 45$	$45 < t \leq 50$
Frequency	2	6	7	19	9	7

- (a) Write down the modal time interval.

[1]

35 < $t \leq 40$

- (b) Calculate an estimate of the mean time.

Show all your working.

[4]

Time (t seconds)	Frequency (f)	Midpoint (M)	$f \times M$
$20 < t \leq 25$	2	22.5	45
$25 < t \leq 30$	6	27.5	165
$30 < t \leq 35$	7	32.5	227.5
$35 < t \leq 40$	19	37.5	712.5
$40 < t \leq 45$	9	42.5	382.5
$45 < t \leq 50$	7	47.5	332.5
Sum	50		1865

Estimate of the mean (μ) is:

$$\mu = \frac{\sum f \times M}{\sum f}$$

$$= \frac{1865}{50}$$

$$= 37.3$$

(c) A new frequency table is made from the results shown in the table opposite.

Time (t seconds)	$20 < t \leq 35$	$35 < t \leq 40$	$40 < t \leq 50$
Frequency			

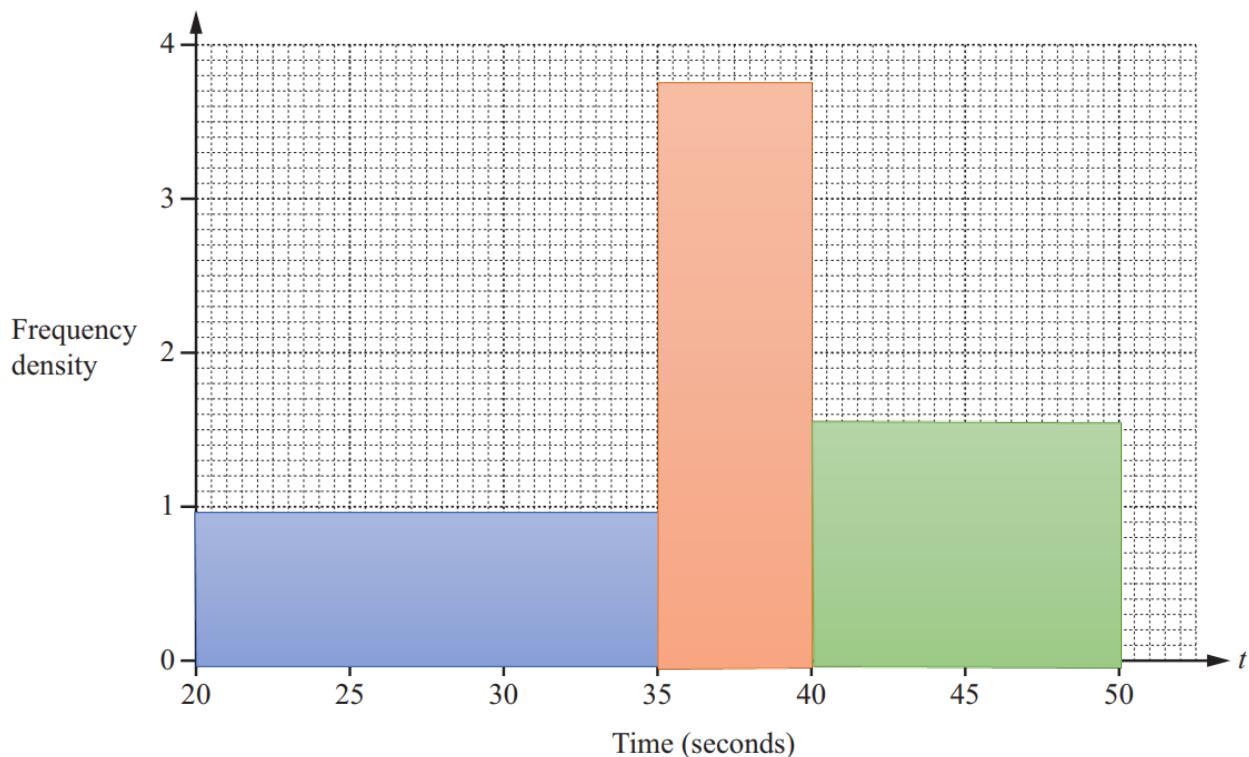
[1]

(i) Complete the table.

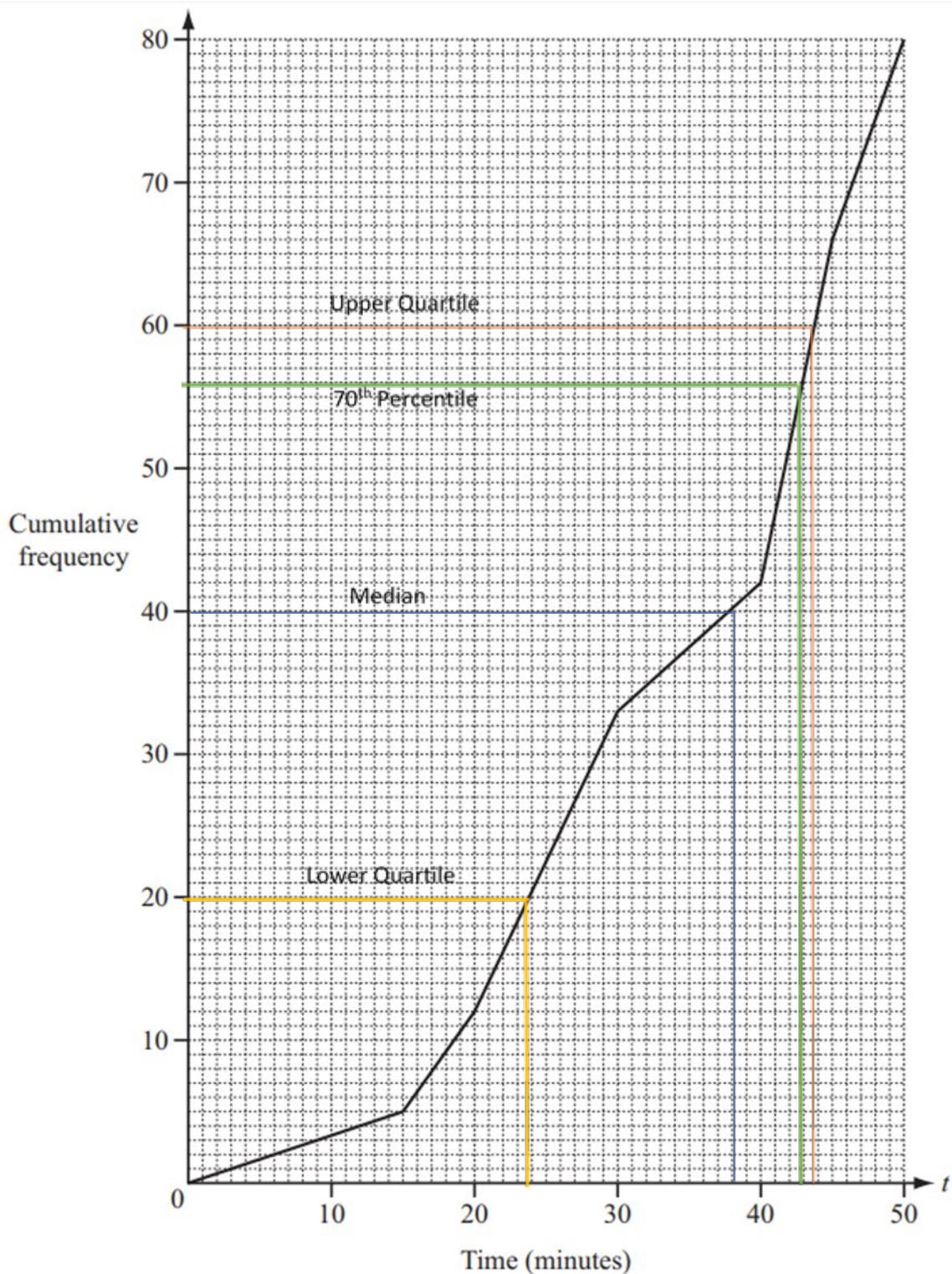
Time (t seconds)	$20 < t \leq 35$	$35 < t \leq 40$	$40 < t \leq 50$
Frequency	15	19	16

(ii) On the grid, draw a histogram to show the information in this new table.

[3]



Question 4



The times (t minutes) taken by 80 people to complete a charity swim were recorded.
The results are shown in the cumulative frequency diagram above.

[1]

(a) Find

(i) the median,

Median (blue line) is **38.5**

(ii) the inter-quartilerange,

[2]

Range is upper quartile (orange line) minus the lower quartile (yellow line)

$$44 - 24$$

$$= 20$$

(iii) the 70th percentile.

[2]

$$0.7 \times 80 = 56$$

70th percentile (green line) is 43

(b) The times taken by the 80 people are shown in this grouped frequencytable.

Time (t minutes)	$0 < t \leq 20$	$20 < t \leq 30$	$30 < t \leq 45$	$45 < t \leq 50$
Frequency	12	21	33	14

(i) Calculate an estimate of the mean time.

[4]

Time (t minutes)	Frequency (f)	Midpoint (m)	$f \times m$
$0 < t \leq 20$	12	10	120
$20 < t \leq 30$	21	25	525
$30 < t \leq 45$	33	37.5	1237.5
$45 < t \leq 50$	14	47.5	665
Sum	80		2547.5

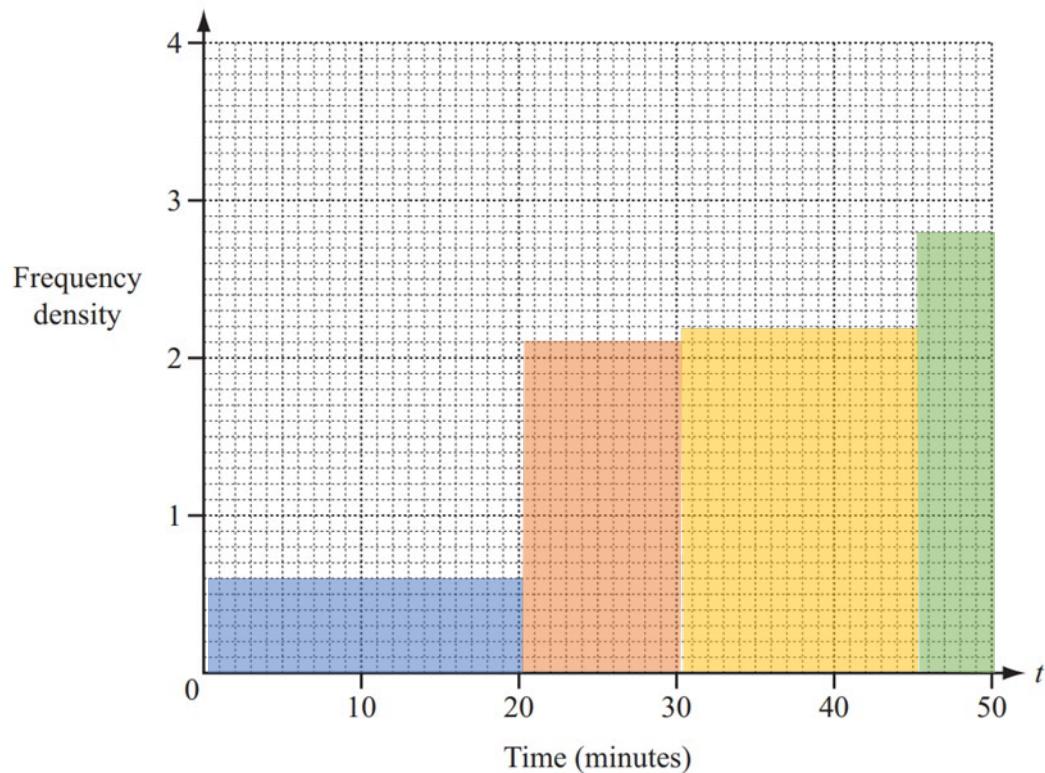
Estimate of the mean is

$$\mu = \frac{\sum f \times m}{\sum f}$$

$$= \frac{2547.5}{80}$$

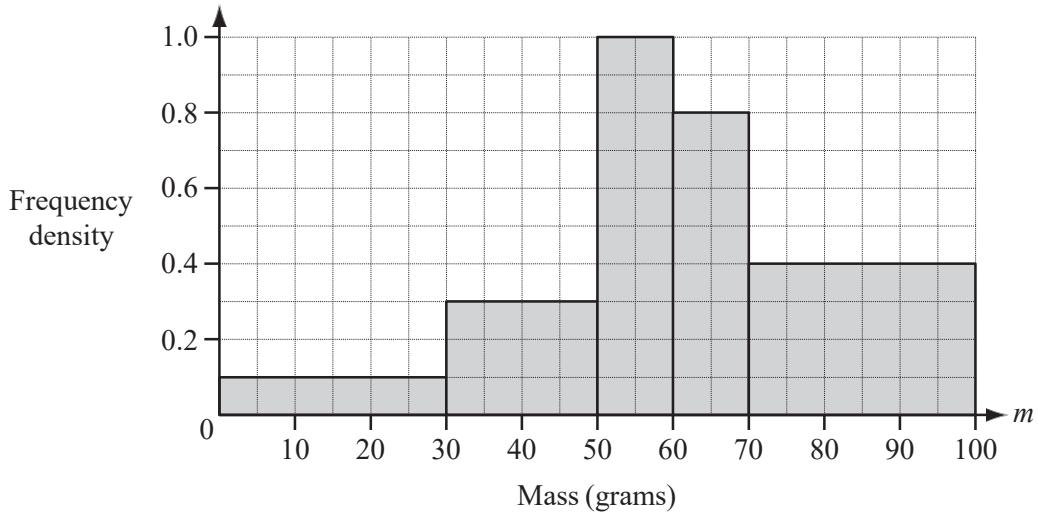
$$= 31.8$$

- (ii) Draw a histogram to represent the grouped frequency table. [4]



Question 5

(a)



The histogram shows some information about the masses (m grams) of 39 apples.

- (i) Show that there are 12 apples in the interval $70 < m \leq 100$. [1]

$$30 \times 0.4$$

$$= 12$$

- (ii) Calculate an estimate of the mean mass of the 39 apples. [5]

The data can be represented as

Mass (grams)	Frequency, f	Midpoint, m	$f \times m$
$0 < m \leq 30$	3	15	45
$30 < m \leq 50$	6	40	240
$50 < m \leq 60$	10	55	550
$60 < m \leq 70$	8	65	520
$70 < m \leq 100$	12	85	1020
Sum	39		2375

The estimate of the mean is

$$\mu = \frac{\sum f \times m}{\sum f}$$

$$= \frac{2375}{39}$$

$$= 60.9$$

(b) The mean mass of 20 oranges is 70 g.

One orange is eaten.

The mean mass of the remaining oranges is 70.5g.

Find the mass of the orange that was eaten.

[3]

$$20 \times 70 - 19 \times 70.5$$

$$= 60.5$$

Question 6

- (a) 80 students were asked how much time they spent on the internet in one day.
This table shows the results.

Time (t hours)	$0 < t \leq 1$	$1 < t \leq 2$	$2 < t \leq 3$	$3 < t \leq 5$	$5 < t \leq 7$	$7 < t \leq 10$
Number of students	15	11	10	19	13	12

- (i) Calculate an estimate of the mean time spent on the internet by the 80 students. [4]

Take the midpoint of all the times:

Time	0.5	1.5	2.5	4	6	8.5
Number	15	11	10	19	13	12

Now multiply the number of students by the time and then divide through by 80:

Mean Time

$$= \frac{(15 \times 0.5) + (11 \times 1.5) + (10 \times 2.5) + (19 \times 4) + (13 \times 6) + (12 \times 8.5)}{80}$$

$$= 3.81\text{hrs}$$

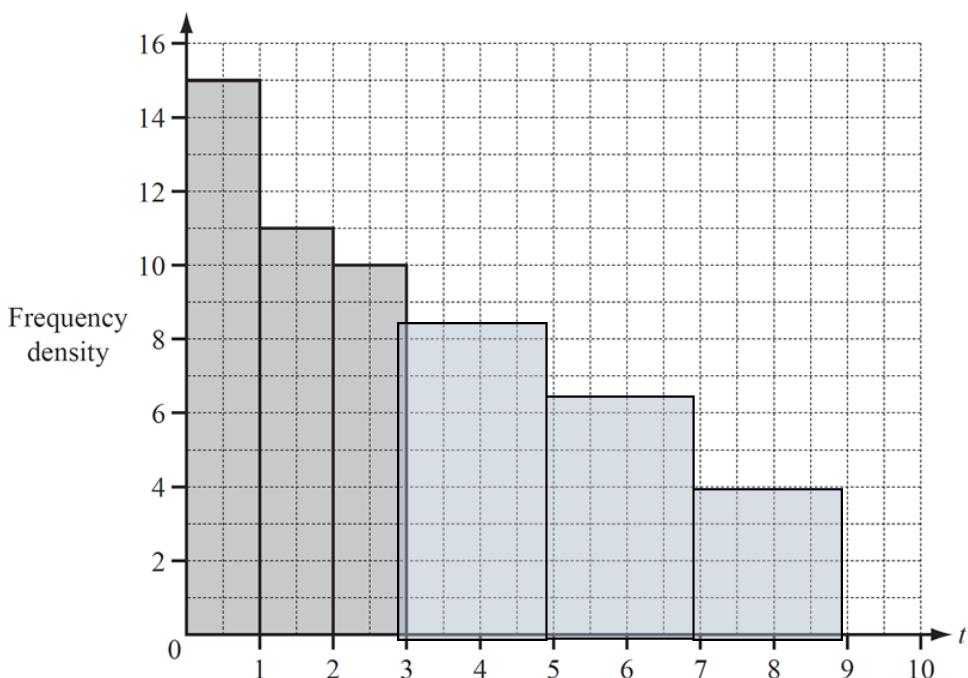
- (ii) On the grid, complete the histogram to show this information. [4]

To adjust the frequency for different class widths, we identify a minimum class size, which is the smallest class size found on the table. The minimum class size here is 1.

To find adjusted frequency, we take:

$$\text{Adjusted frequency} = \frac{\text{Minimum class size}}{\text{Class size}} \times \text{frequency}$$

Time	$0 < t \leq 1$	$1 < t \leq 2$	$2 < t \leq 3$	$3 < t \leq 5$	$5 < t \leq 7$	$7 < t \leq 10$
Frequency	15	11	10	19	13	12
Class size	1	1	1	2	2	3
Adjusted Freq.	15	11	10	8.5	6.5	4

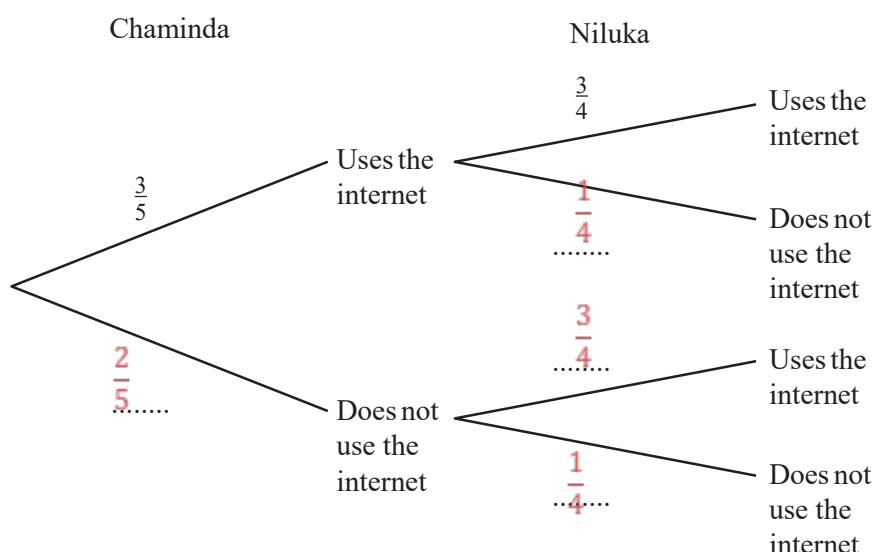


(b) The probability that Chaminda uses the internet on any day is $\frac{3}{5}$.

The probability that Niluka uses the internet on any day is $\frac{3}{4}$.

(i) Complete the tree diagram.

[2]



The probability Niluka uses the internet is $\frac{3}{4}$.

This is equivalent to:

$$\left(\frac{3}{5} \times \frac{3}{4}\right) + \left(\frac{2}{5} \times P(N, \text{uses internet})\right) = \frac{3}{4}$$

$$P(N, \text{uses internet}) = \frac{3}{4}$$

- (ii) Calculate the probability, that on any day, at least one of the two students uses the internet. [3]

The probability is the sum of the probability of Chaminda using the internet, and the probability that if Chaminda does not use the internet, then Niluka uses the internet.

$$P(\text{At least 1 uses internet}) = \frac{3}{5} + \left(\frac{2}{5} \times \frac{3}{4}\right)$$

$$= \frac{9}{10}$$

- (iii) Calculate the probability that Chaminda uses the internet on three consecutive days. [2]

$$P(C, 3 \text{ consecutive days}) = \frac{3}{5} \times \frac{3}{5} \times \frac{3}{5}$$

$$= \frac{27}{125}$$

Question 7

Height (h cm)	$150 < h \leq 160$	$160 < h \leq 165$	$165 < h \leq 180$	$180 < h \leq 190$
Frequency	5	9	18	10

The table shows information about the heights of a group of 42 students.

- (a) Using mid-interval values, calculate an estimate of the mean height of the students.
Show your working.

[3]

You can find the mean using the following equation: (Σ = The sum of)

$$\text{Mean height} = \frac{\sum(h * \text{frequency})}{\text{frequency}}$$

By first finding the midpoints of each of the intervals, we can sub in to the equation like so:

$$= \frac{155*5 + 162.5*9 + 172.5*18 + 185*10}{42}$$

$$= \frac{7192.5}{42}$$

$$= 171.25\text{cm}$$

- (b) Write down the interval which contains the lower quartile.

[1]

To find the lower quartile, we first find what a quarter of the total frequency:

$$\rightarrow \frac{42}{4}$$

$$= 10.5$$

Hence we must find where the 10.5th tallest students lies.

$$160 < h \leq 165$$

(c) Complete the histogram to show the information in the table.

One column has already been drawn for you.

[4]

Before we can complete the histogram we must first find out the frequency density for each interval using the following equation:

$$\text{Frequency density} = \frac{\text{frequency}}{\text{width of interval}}$$

$$F(160 < h \leq 165) = :$$

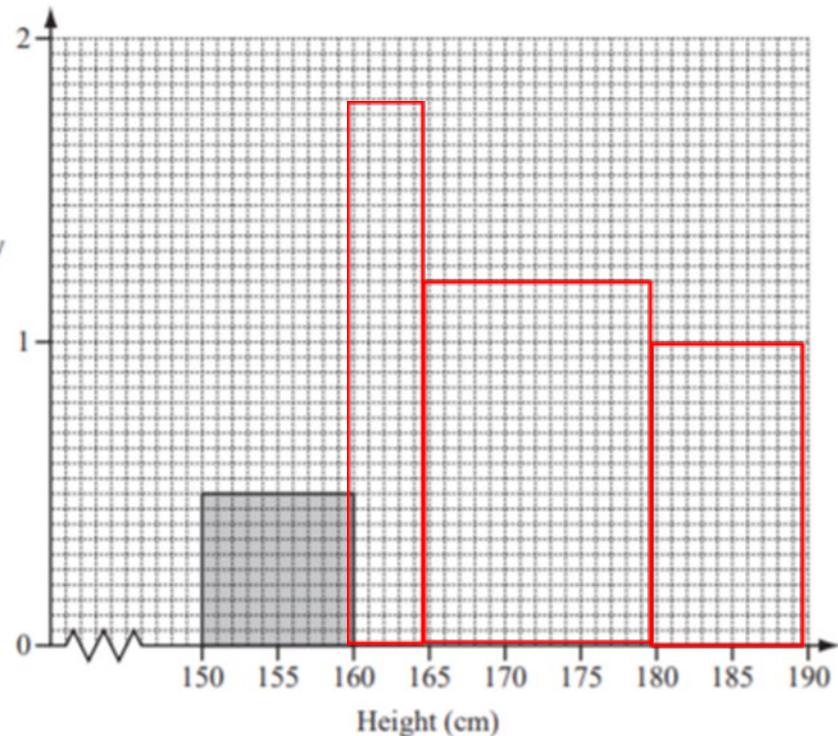
$$= 1.8$$

$$F(165 < h \leq 180) = :$$

$$= 1.2$$

$$F(180 < h \leq 190) = :$$

$$= 1$$



Now we know the frequency density for each interval, the bars can be drawn onto the histogram.

Statistics

Difficulty: Hard

Model Answers 4

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Hard
Booklet	Model Answers 4

Time allowed: 107 minutes

Score: /93

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

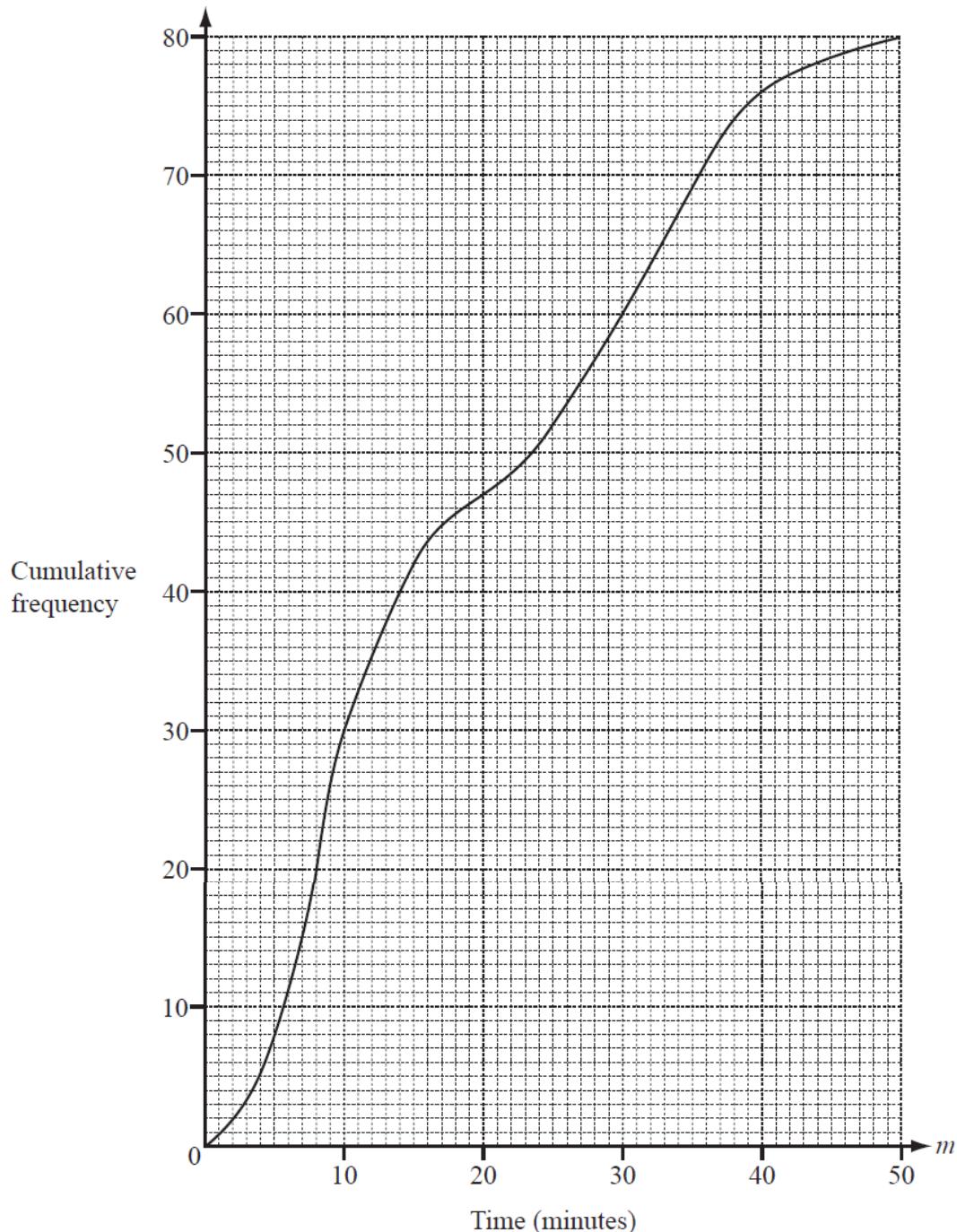
A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

Sam asked 80 people how many minutes their journey to work took on one day.
The cumulative frequency diagram shows the times taken (m minutes).



(a) Find

(i) the median,

[1]

There are 80 people in total therefore the median is the time at half that value.

Draw a line at $y = 40$ and read off the x value to find the median:

$$\mathbf{q_2 = 14}$$

(ii) the lower quartile,

[1]

The lower quartile is found at the $\frac{n+1}{4}$ th value

$$\frac{80 + 1}{4} = 20.25$$

Draw a line at $y = 20.25$ and read off the x value to find the lower quartile:

$$q_1 = 8 \text{ minutes}$$

(iii) the inter-quartile range.

[1]

To find the IQR we first need to find the upper quartile then subtract the lower quartile value from it.

The upper quartile is found at the $\frac{3}{4}(n + 1)$ th value

$$\frac{3}{4}(80 + 1) = 60.75$$

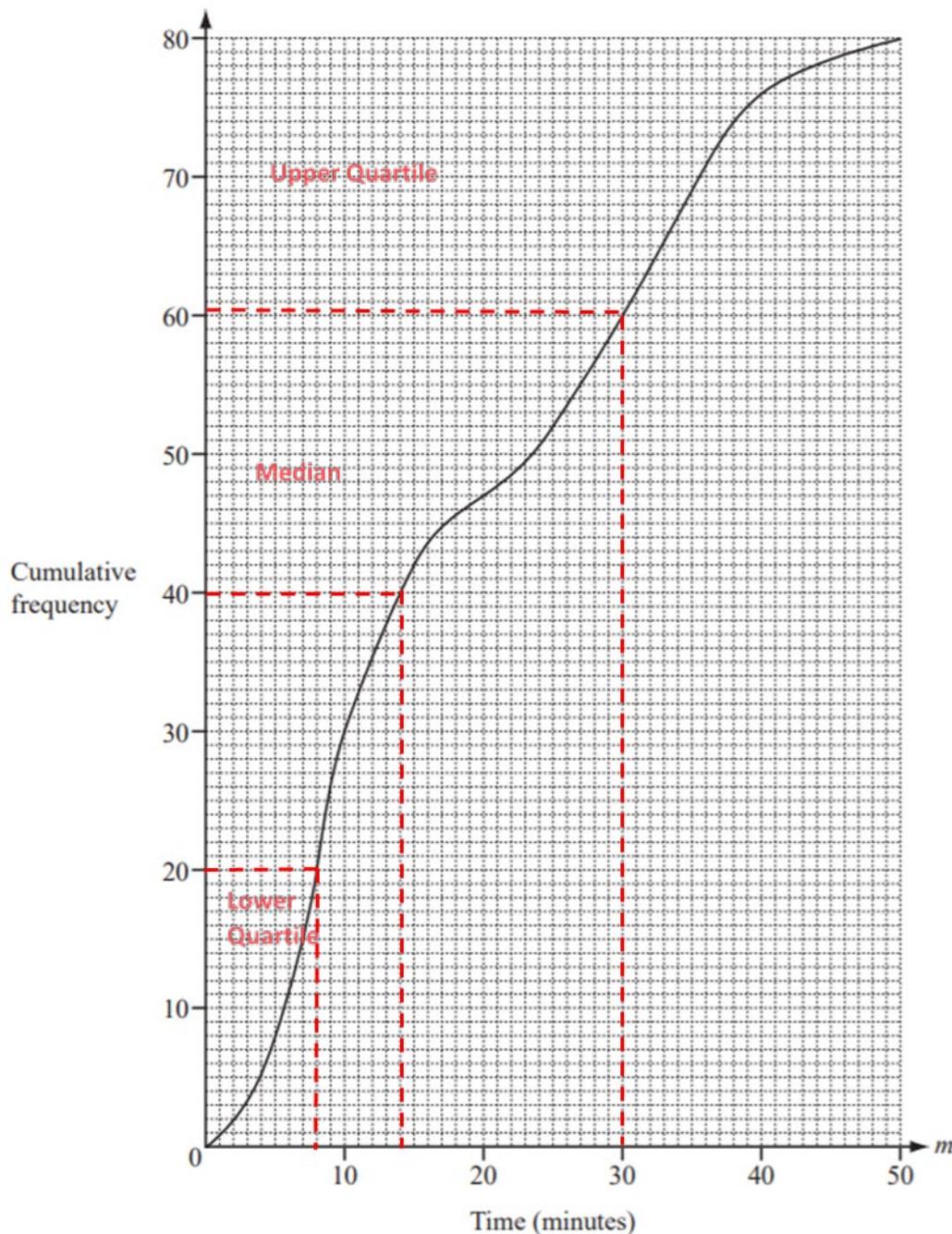
Draw a line at $y = 60.75$ and read off the x value to find the upper quartile:

$$q_3 = 30$$

$$IQR = q_3 - q_1$$

$$= 30 - 8 = 22$$

$$= 22$$



(b) One of the 80 people is chosen at random.

Find the probability that their journey to work took more than 35 minutes.
Give your answer as a fraction.

[2]

Draw a vertical line from 35 minutes to the curve and find the cumulative frequency:

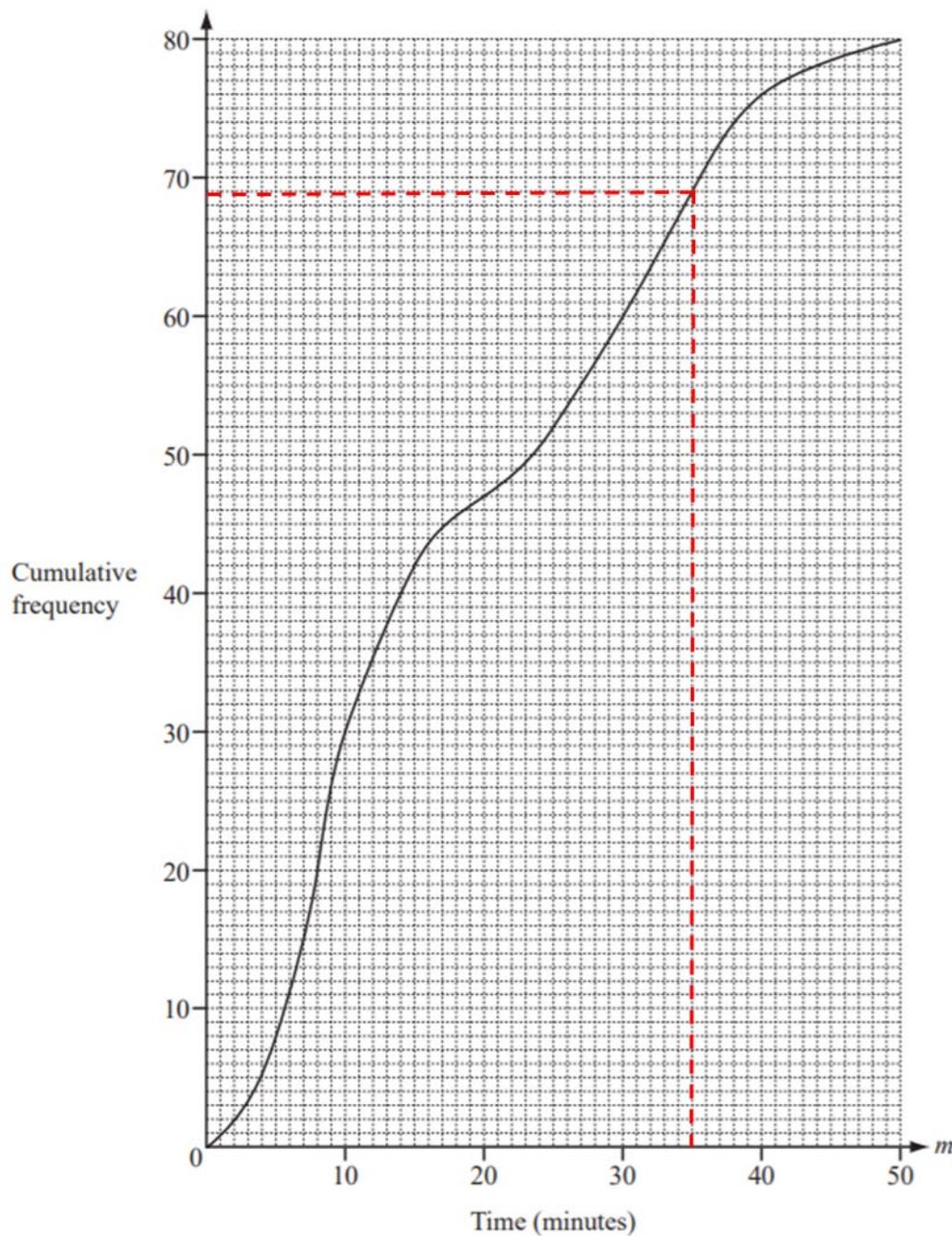
$$\text{cumulative frequency} = 69$$

Find the amount of people above 35 minutes:

$$80 - 69 = 11$$

Find the probability:

$$\frac{11}{80}$$

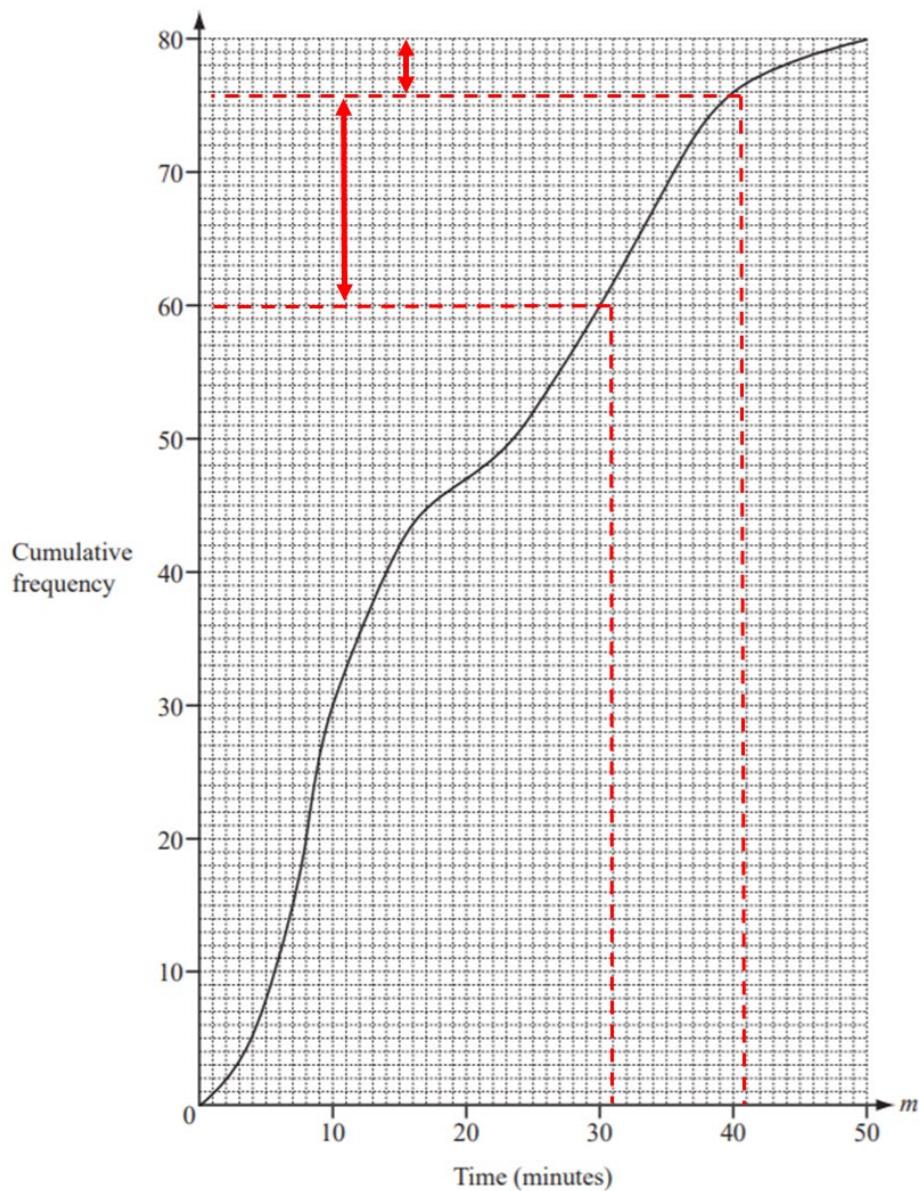


(c) Use the cumulative frequency diagram to complete this frequency table. [2]

Time (m minutes)	$0 < m \leq 10$	$10 < m \leq 15$	$15 < m \leq 30$	$30 < m \leq 40$	$40 < m \leq 50$
Frequency	30	12	18	16	4

Draw lines from 30 and 40 to find the difference in cumulative frequency at 30 and 40

and 40 and 50.



- (d) Using mid-interval values, calculate an estimate of the mean journey time for the 80 people. [3]

The mean is calculated by:

Time (m minutes)	$0 < m \leq 10$	$10 < m \leq 15$	$15 < m \leq 30$	$30 < m \leq 40$	$40 < m \leq 50$
Frequency	30	12	18	16	4
Frequency dens.	3	2.4	1.2	1.6	0.4

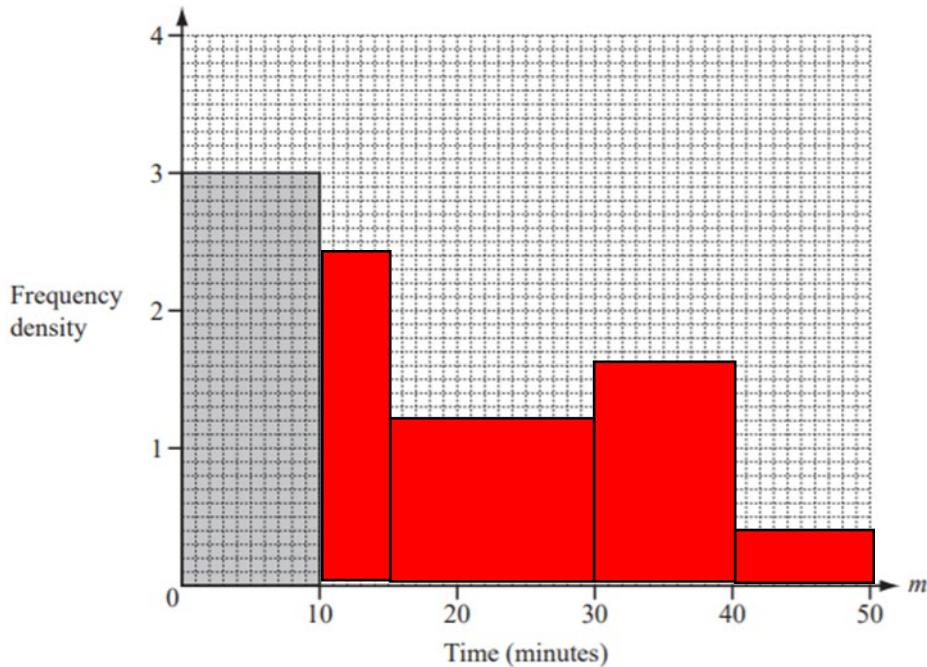
$$\frac{\sum \text{midpoint} \times \text{frequency}}{\text{frequency}}$$

$$\frac{5 \times 30 + 12.5 \times 12 + 22.5 \times 18 + 35 \times 16 + 45 \times 4}{80} = 18.1 \text{ minutes}$$

- (e) Use the table in **part (c)** to complete the histogram to show the times taken by the 80 people.
One column has already been completed for you.

[5]

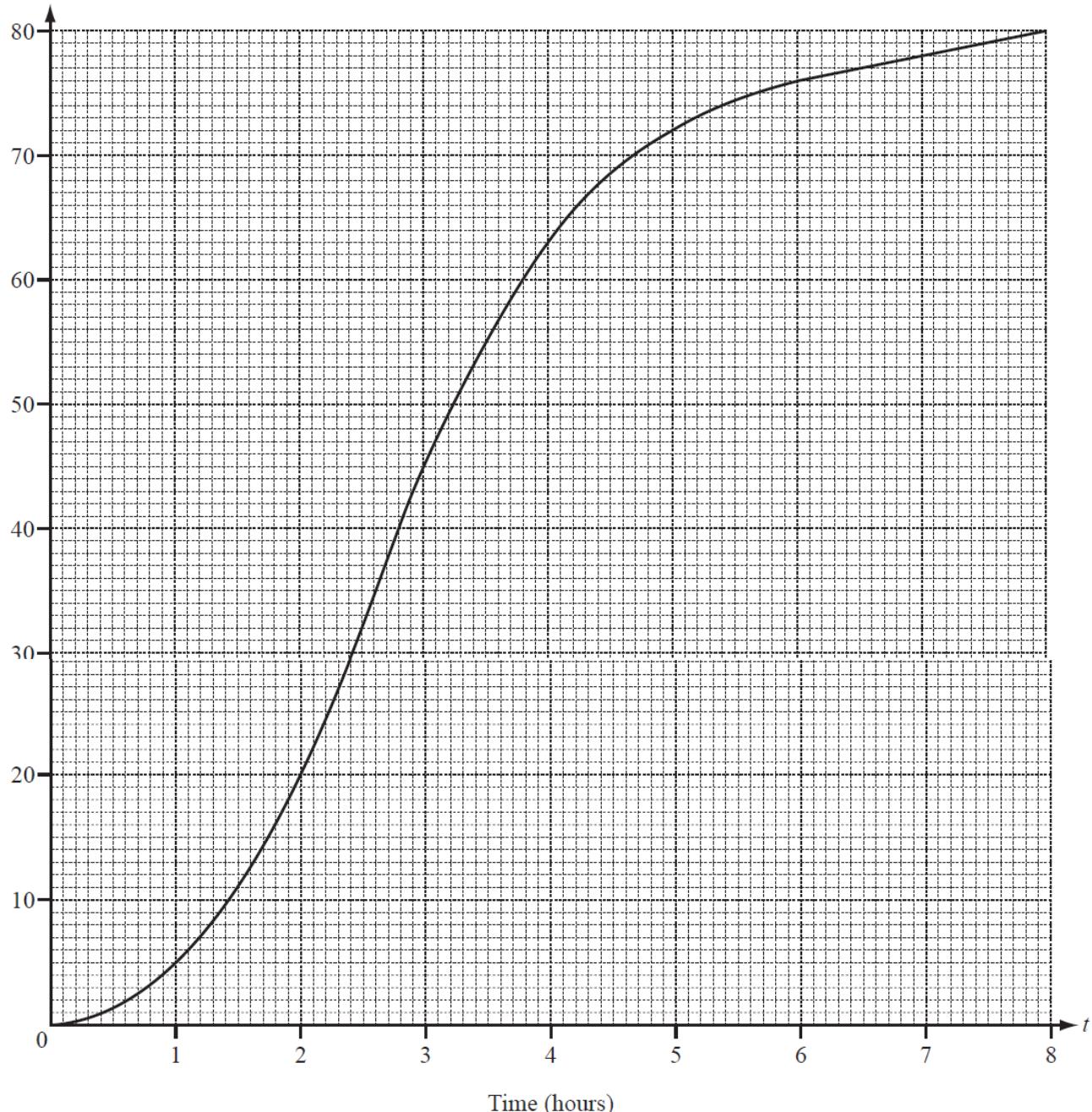
Frequency density is the frequency divided by the range:



Question 2

Felix asked 80 motorists how many hours their journey took that day.
He used the results to draw a cumulative frequency diagram.

Cumulative
frequency



(a) Find

- (i) the median, [1]

2.8

- (ii) the upper quartile, [1]
3.8

(iii) the inter-quartile range.

[1]

Lower quartile is 2, hence the inter-quartile range is

$$3.8 - 2$$

$$= 1.8$$

(b) Find the number of motorists whose journey took more than 5 hours but no more than 7 hours.

[1]

Number of motorists whose journey took 7 hours or less is 78.

Number of motorists whose journey took 5 hours or less is 72.

Hence the number of motorists whose journey took more than 5 but less than 7 is

$$(80 - 72) - (80 - 78)$$

$$= 6$$

(c) The frequency table shows some of the information about the 80 journeys.

[2]

Time in hours (t)	Frequency (f)	Midpoint (M)	$f \times M$
$0 < t \leq 2$	20	1	20
$2 < t \leq 3$	25	2.5	62.5
$3 < t \leq 4$	18	3.5	63
$4 < t \leq 5$	9	4.5	40.5
$5 < t \leq 6$	4	5.5	22
$6 < t \leq 8$	4	7	28
Sum	80		236

(i) Use the cumulative frequency diagram to complete the table above.

(ii) Calculate an estimate of the mean number of hours the 80 journeys took.

[4]

Using the extended table above we have the estimation of the mean as

$$\mu = \frac{\sum f \times M}{\sum f}$$

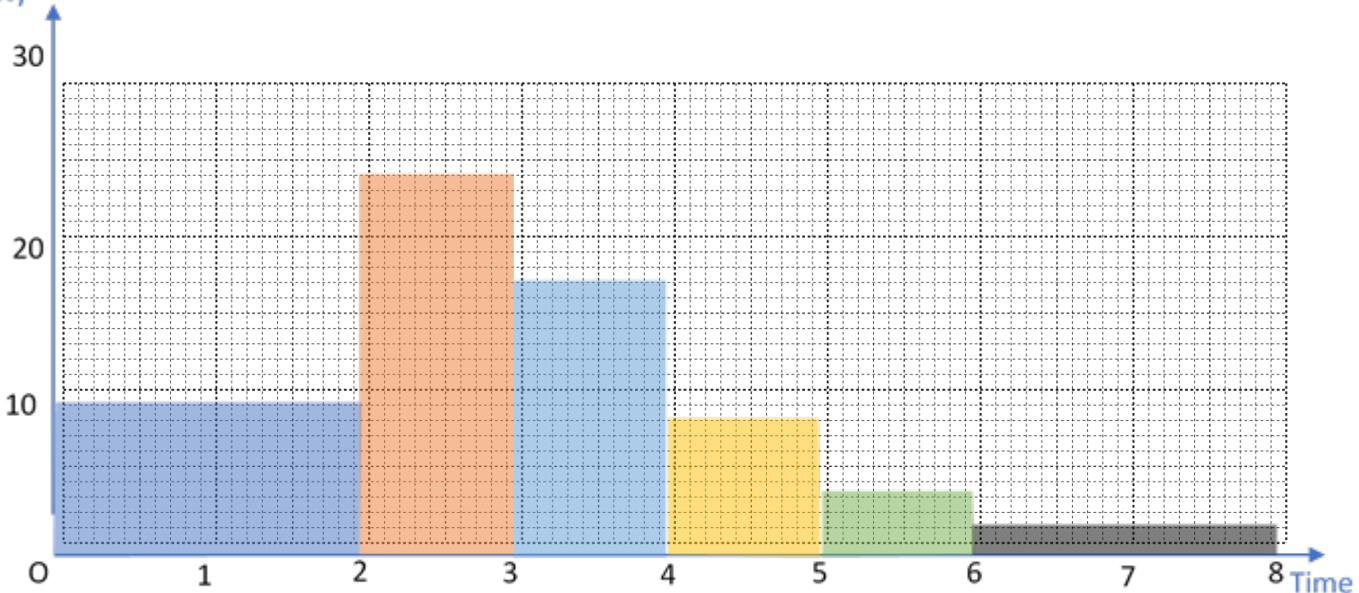
$$= \frac{236}{80}$$

$$= 2.95$$

(d) On the grid, draw a histogram to represent the information in your table in **part (c)**.

[5]

Frequency
density

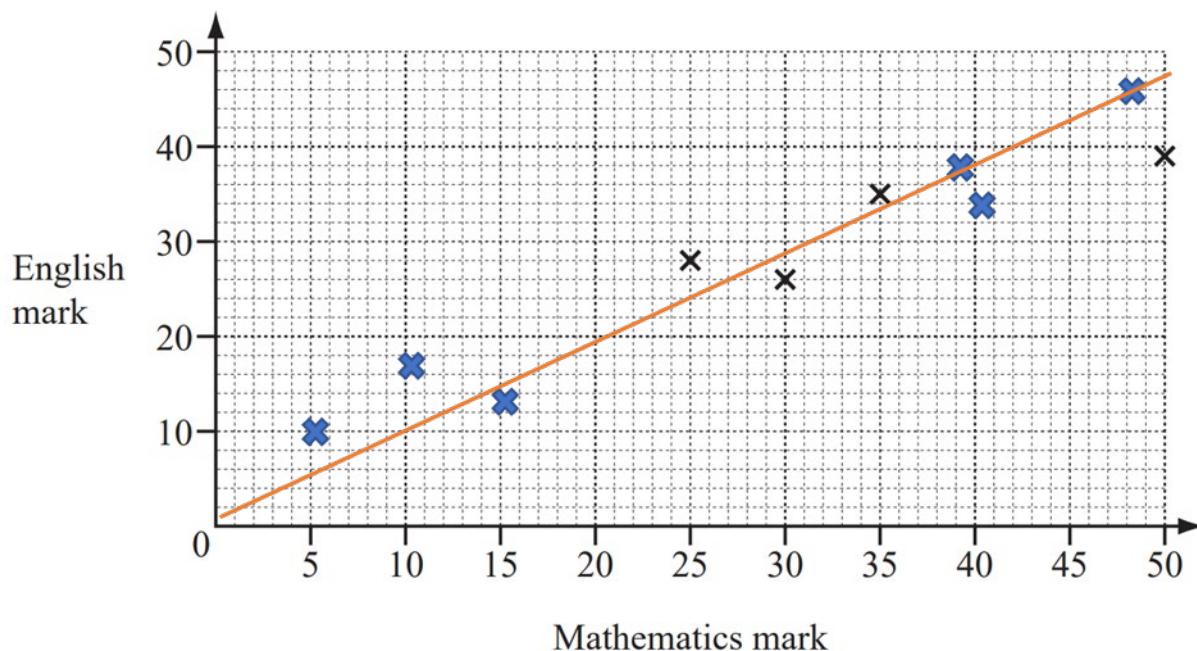


Question 3

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

The table shows the test marks in Mathematics and English for 10 students.

- (a) (i) On the grid, complete the scatter diagram to show the Mathematics and English marks for the 10 students. The first four points have been plotted for you. [2]



- (ii) What type of correlation does your scatter diagram show? [1]

Positive correlation.

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

Orange line drawn above.

- (iv) Ann missed the English test but scored 22 marks in the Mathematics test. Use your line of best fit to estimate a possible English mark for Ann. [1]

Read off the orange line at $M = 22$

(b) Show that the mean English mark for the 10 students is 28.

[2]

$$(26 + 39 + 35 + 28 + 9 + 37 + 45 + 33 + 16 + 12) \div 10$$

$$= 28$$

(c) Two new students do the English test. They both score the **same** mark.

The mean English mark for the 12 students is 31.

[3]

Calculate the English mark for the new students.

Let the new students score be x .

We have

$$28 \times 10 + 2x = 31 \times 12$$

$$\rightarrow 2x = 372 - 280$$

$$\rightarrow x = \frac{92}{2}$$

$$= 46$$

Question 4

The table shows information about the heights of 120 girls in a swimming club.

Height (h metres)	Frequency
$1.3 < h \leq 1.4$	4
$1.4 < h \leq 1.5$	13
$1.5 < h \leq 1.6$	33
$1.6 < h \leq 1.7$	45
$1.7 < h \leq 1.8$	19
$1.8 < h \leq 1.9$	6

- (a) (i) Write down the modal class. [1]

The modal class represents the class which contains the middle value.

The middle value is 60 girls. By adding up the frequencies for the intervals in order, we work out that this value is reached for the interval

$$1.6 < h \leq 1.7.$$

$$(4 + 13 + 33 < 60, 4 + 13 + 33 + 45 > 60)$$

- (ii) Calculate an estimate of the mean height. Show all of your working. [4]

To work out an estimate of the mean we initially need to assume that each frequency represents the middle value of its corresponding interval.

Therefore, we need to work out the middle values:

Height (h meters)	Frequency	Middle value	Frequency x Middle value
$1.3 < t \leq 1.4$	4	1.35	5.4
$1.4 < t \leq 1.5$	13	1.45	18.85
$1.5 < t \leq 1.6$	33	1.55	51.15
$1.6 < t \leq 1.7$	45	1.65	74.25
$1.7 < t \leq 1.8$	19	1.75	33.25
$1.8 < t \leq 1.9$	6	1.85	11.1
Total	120	-	194

Then, for each interval, we multiply the frequency by the middle value to work out the amount of time represented by that interval.

We add up all the values to obtain the total height, 194 m.

To work out the estimate we need to divide the total amount of time by the total number of people.

$$\text{Mean} = \frac{194}{120}$$

$$\text{Mean} = 1.6166$$

- (b) Girls from this swimming club are chosen at random to swim in a race.
Calculate the probability that

- (i) the height of the first girl chosen is more than 1.8 metres, [1]

The table shows that there are only 6 girls from the group of 120 which are taller than 1.8 m.

Therefore, the probability of choosing one of them will be:

$$P = \frac{6}{120}$$

(ii) the heights of **both** the first and second girl chosen are 1.8 metres or less. [3]

The table shows that there are 114 girls from the group of 120 which are shorter than 1.8 m.

Therefore, the probability of the first girl chosen to be shorter than 1.8 m is:

$$P = \frac{114}{120}$$

After choosing the first girl, the total number of girls is reduced by 1, as well as the number of girls shorter than 1.8 m.

$$P = \frac{113}{119}$$

These 2 events need to happen simultaneously, therefore, we multiply the probabilities for the 2 separate events.

$$P = \frac{114}{120} \times \frac{113}{119}$$

$$P = \frac{12882}{14280}$$

$$P = \frac{2147}{2380}$$

$$= 0.9021$$

- (c) (i) Complete the cumulative frequency table for the heights.

[1]

Height (h metres)	Cumulative frequency
$h \leq 1.3$	0
$h \leq 1.4$	4
$h \leq 1.5$	17
$h \leq 1.6$	50
$h \leq 1.7$	
$h \leq 1.8$	114
$h \leq 1.9$	

To work out the cumulative frequency for each inequality we need to add up the frequency for multiple intervals which in the end will represent the inequality.

For example:

$h \leq 1.4$ is represented by the interval $1.3 < h \leq 1.4$, with the frequency 0

$$+ 4 = 4$$

$h \leq 1.6$ is represented by the interval $1.6 < h \leq 1.7$ and the inequality h

≤ 1.6 with the frequency $50 + 45$

$$= 95.$$

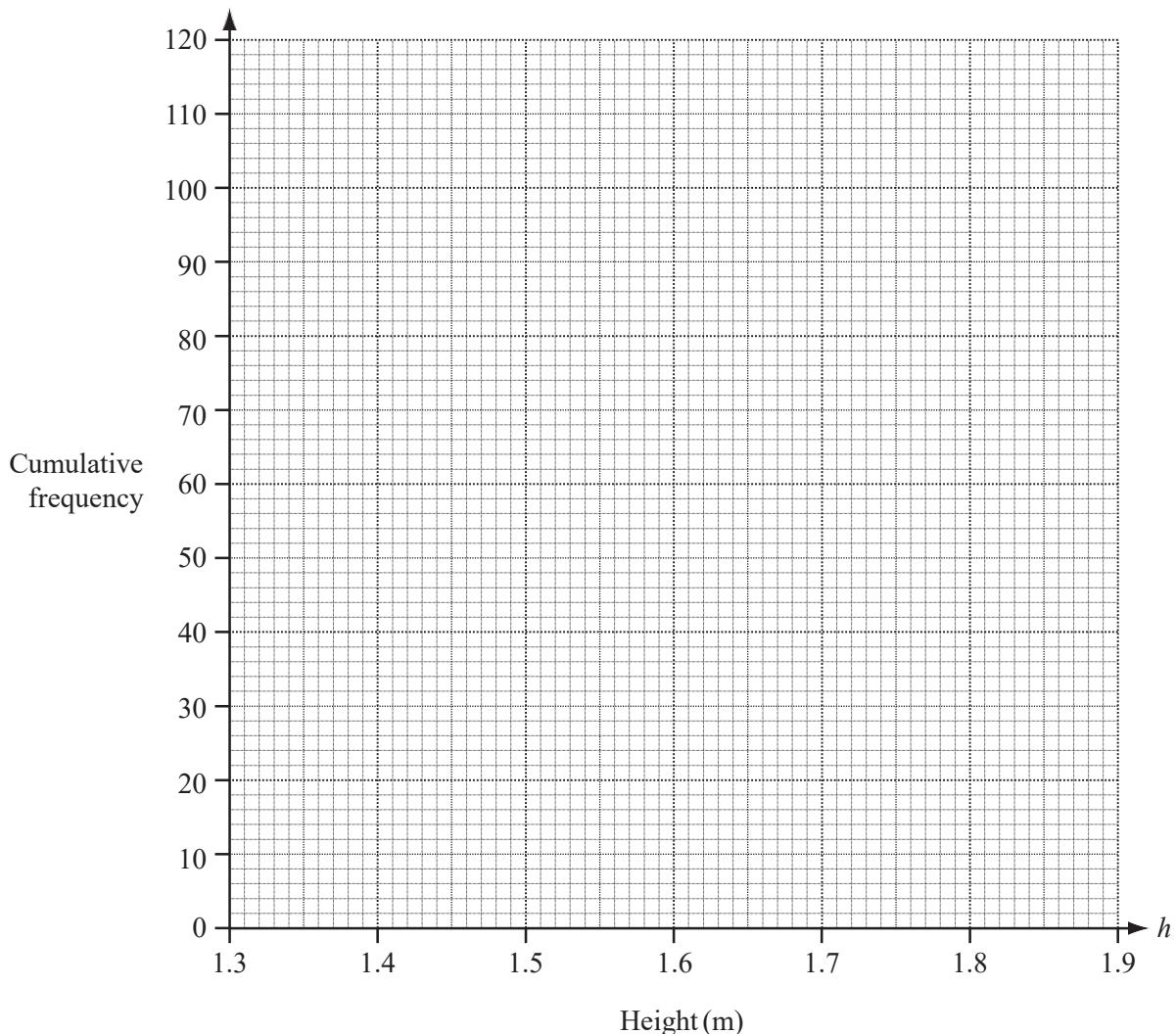
$h \leq 1.9$ is represented by the inequality $h \leq 1.8$ and the interval $1.8 < t \leq$

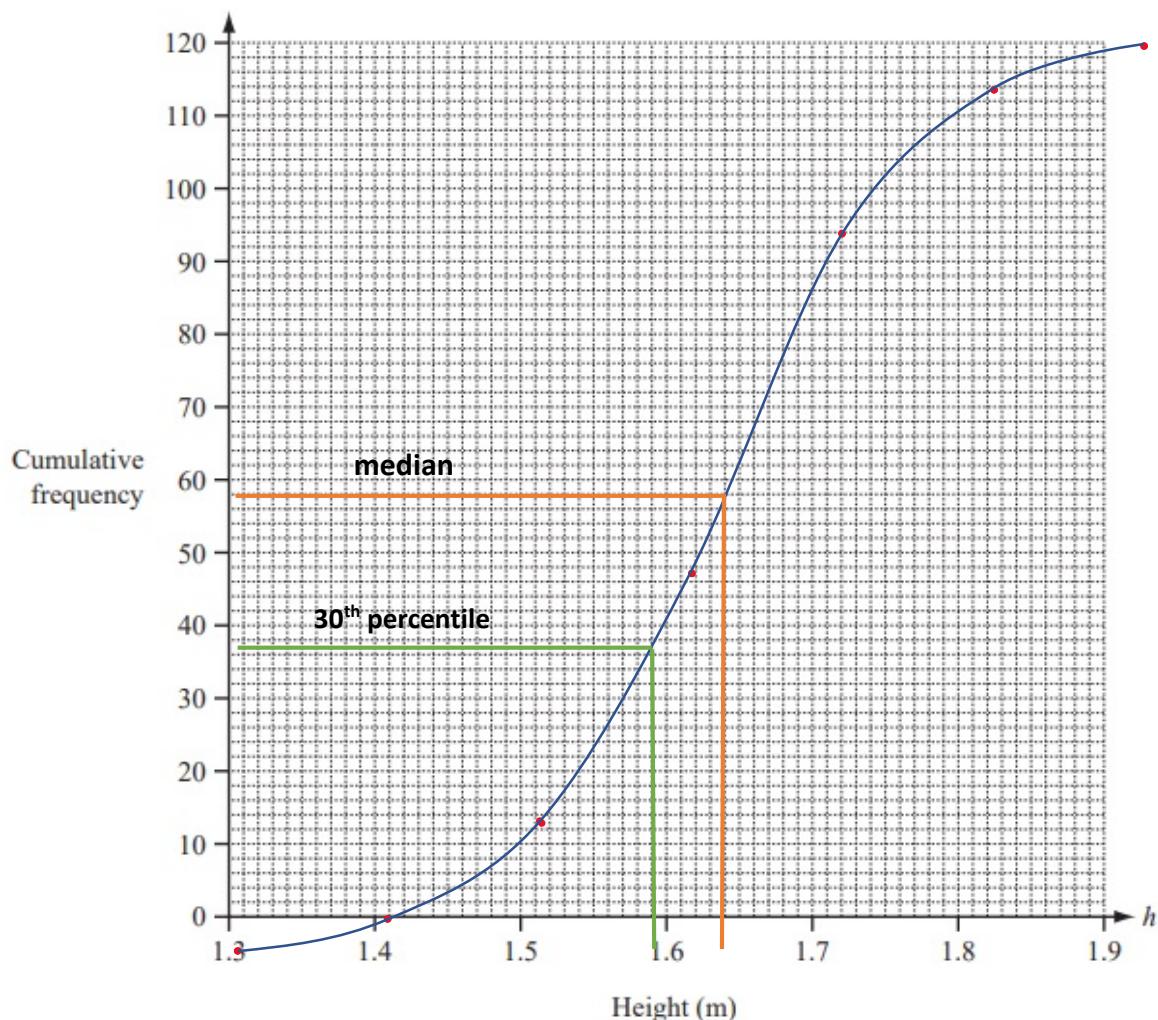
1.9 , with the frequency $114 + 6$

$$= 120.$$

(ii) Draw the cumulative frequency graph on the grid.

[3]





(d) Use your graph to find

(i) the median height,

[1]

The median will be the height corresponding to half of the number of students, 60.

This median is 1.62 m.

(ii) the 30th percentile.

[1]

Similarly, the 30th percentile will be the height corresponding to 30% of the number of students, 40.

This 30th percentile 1.57 minutes.

Question 5

The table below shows the marks scored by a group of students in a test.

Mark	11	12	13	14	15	16	17	18
Frequency	10	8	16	11	7	8	6	9

- (a) Find the mean, median and mode.

[6]

We find the mean by summing the product of marks and their respective frequencies and then divide by the sum of frequencies (i.e. the total number of students).

$$\text{mean} = \frac{\text{sum of (mark} \times \text{frequency)}}{\text{sum of frequencies}}$$

$$\text{mean} = \frac{11 \times 10 + 12 \times 8 + 13 \times 16 + 14 \times 11 + 15 \times 7 + 16 \times 8 + 17 \times 6 + 18 \times 9}{10 + 8 + 16 + 11 + 7 + 8 + 6 + 9}$$

Calculate the mean:

$$\text{mean} = 14.2$$

The median is the 50th percentile. It is the minimum value of the mark, which would give us more than a half of the total number of students if we sum the students who got this mark or lower.

There are 70 students in total, so we are looking for the mark which gives 37.5

We keep summing the number of students as we move to higher marks. At mark 13, we only have 34 students who got this mark or lower, so that is not yet a half of all the students. At mark 14, however, we have 45 students who got mark 14 or lower, so:

$$\text{median} = 14$$

Mode is simply the mark that was awarded to the highest number of students (highest frequency).

There were 16 students who got mark 13.

mode = 13

(b) The table below shows the time (t minutes) taken by the students to complete the test.

Time (t)	$0 < t \leq 10$	$10 < t \leq 20$	$20 < t \leq 30$	$30 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Frequency	2	19	16	14	15	9

(i) Cara rearranges this information into a new table.

Complete her table.

[2]

Time (t)	$0 < t \leq 20$	$20 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Frequency				9

When Cara rearranges the information, the total frequency in a given interval cannot change; hence we just need to add the number of students that belong to the new intervals.

Time (t)	$0 < t \leq 10$	$10 < t \leq 20$	$20 < t \leq 30$	$30 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Frequency	2	19	16	14	15	9

(i) Cara rearranges this information into a new table.

Time (t)	$0 < t \leq 20$	$20 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Frequency	21	30	15	9

(ii) Cara wants to draw a histogram to show the information in **part (b)(i)**.

Complete the table below to show the interval widths and the frequency densities.

[3]

	$0 < t \leq 20$	$20 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Interval width				10
Frequency density				0.9

The interval width is given by the difference between the top and the bottom

boundary of the interval.

In the last interval, the difference between the top boundary (60) and the bottom

boundary (50) is 10, hence that is the interval width. We follow a similar method for

the other values.

	$0 < t \leq 20$	$20 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Interval width	20	20	10	10

To calculate the frequency density, we use the interval width and the frequency

from the previous table. The frequency density is the ratio of the frequency and the

interval width.

$$\text{frequency density} = \frac{\text{frequency}}{\text{interval width}}$$

Using this relationship, we calculate the frequency density:

	$0 < t \leq 20$	$20 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$
Interval width	20	20	10	10
Frequency density	1.05	1.5	1.5	0.9

(c) **Some** of the students were asked how much time they spent revising for the test.

10 students revised for 2.5 hours, 12 students revised for 3 hours and n students revised for 4 hours.

The mean time that **these** students spent revising was 3.1 hours.

Find n .

Show all your working.

[4]

We use the same formula as in part a), but now we do not know the value of n rather than the mean itself. (Frequency denotes the number of students in a given group)

$$\text{mean} = \frac{\text{sum of } (\text{time} \times \text{frequency})}{\text{sum of frequencies}}$$

We substitute the given values keeping n the unknown.

$$3.1 \text{ hours} = \frac{2.5 \text{ hours} \times 10 + 3 \text{ hours} \times 12 + 4 \text{ hours} \times n}{10 + 12 + n}$$

$$3.1 \text{ hours} = \frac{25 + 36 + 4n}{22 + n}$$

Multiply both sides by $(22+n)$.

$$3.1 \text{ hours} \times (22 + n) = 25 + 36 + 4n$$

$$68.2 \text{ hours} + 3.1n \text{ hours} = 25 + 36 + 4n$$

We no longer need to use the hours after each term. Subtract 61 from both sides of the equation.

$$7.2 + 3.1n = 4n$$

Subtract $3.1n$ from both sides of the equation.

$$7.2 = 0.9 \times n$$

Divide both sides of the equation by 0.9 to get the value of n :

$$\mathbf{n = 8}$$

(a) For a set of six integers, the mode is 8, the median is 9 and the mean is 10.

The smallest integer is greater than 6 and the largest integer is 16.

Find the two possible sets of six integers.

[5]

8, 8, 8, 10, 10, 16

(b) One day Ahmed sells 160 oranges.

He records the mass of each orange.

The results are shown in the table.

Mass (m grams)	$50 < m \leq 80$	$80 < m \leq 90$	$90 < m \leq 100$	$100 < m \leq 120$	$120 < m \leq 150$
Frequency	30	35	40	40	15

(i) Calculate an estimate of the mean mass of the 160 oranges.

[4]

Mass (m grams)	Frequency	Midpoint	<i>frequency</i> \times midpoint
$50 < m \leq 80$	30	65	1950
$80 < m \leq 90$	35	85	2975
$90 < m \leq 100$	40	95	3800
$100 < m \leq 120$	40	110	4400
$120 < m \leq 150$	15	135	2025
Sum		160	15150

Question 6

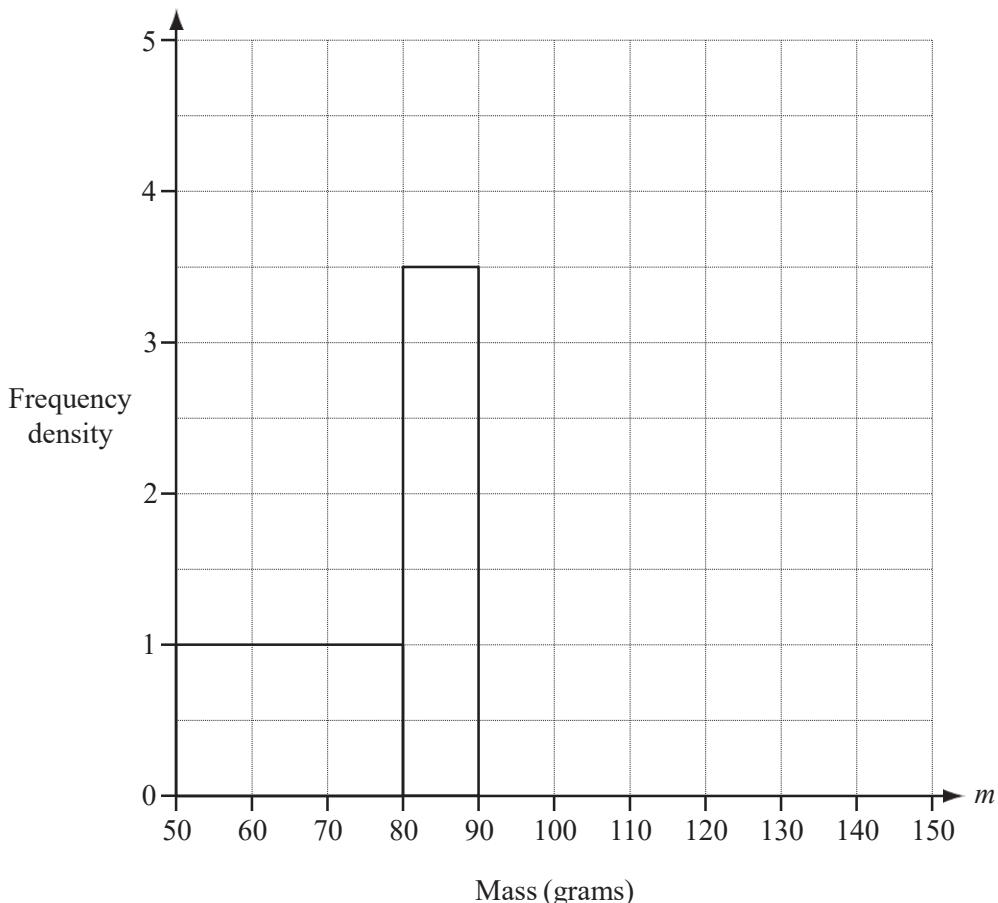
Estimate of the mean is

$$\mu = \frac{\sum f \times m}{\sum f}$$

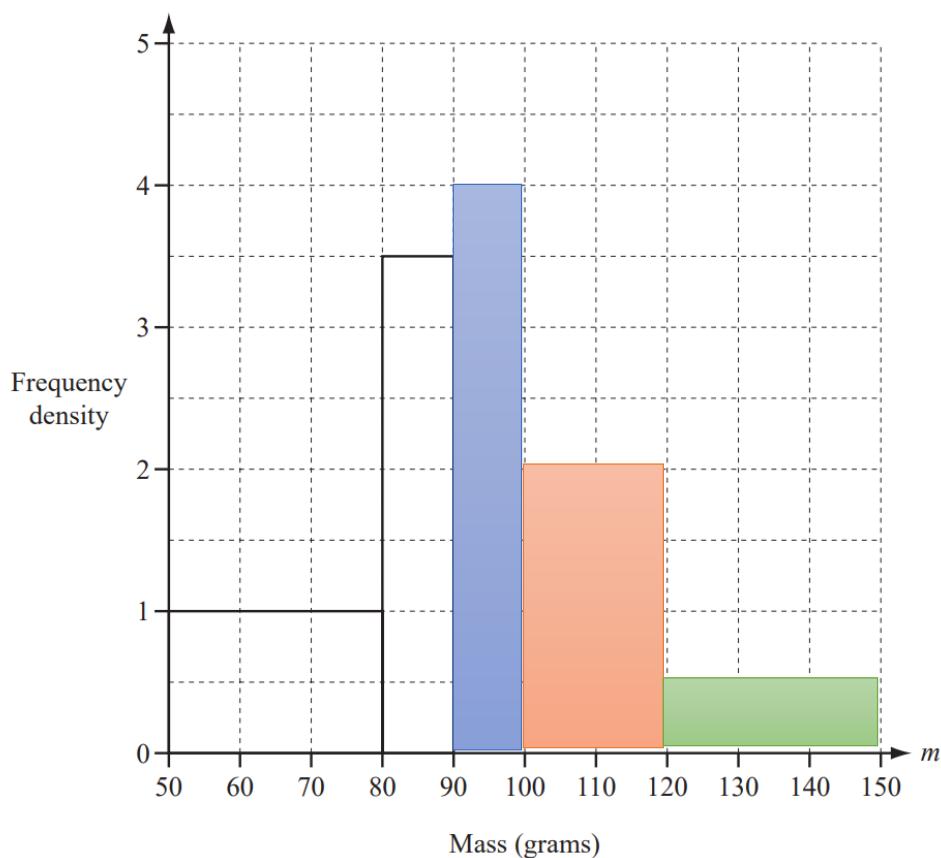
$$= \frac{15150}{160}$$

$$= 94.7$$

(ii) On the grid, complete the histogram to show the information in the table.



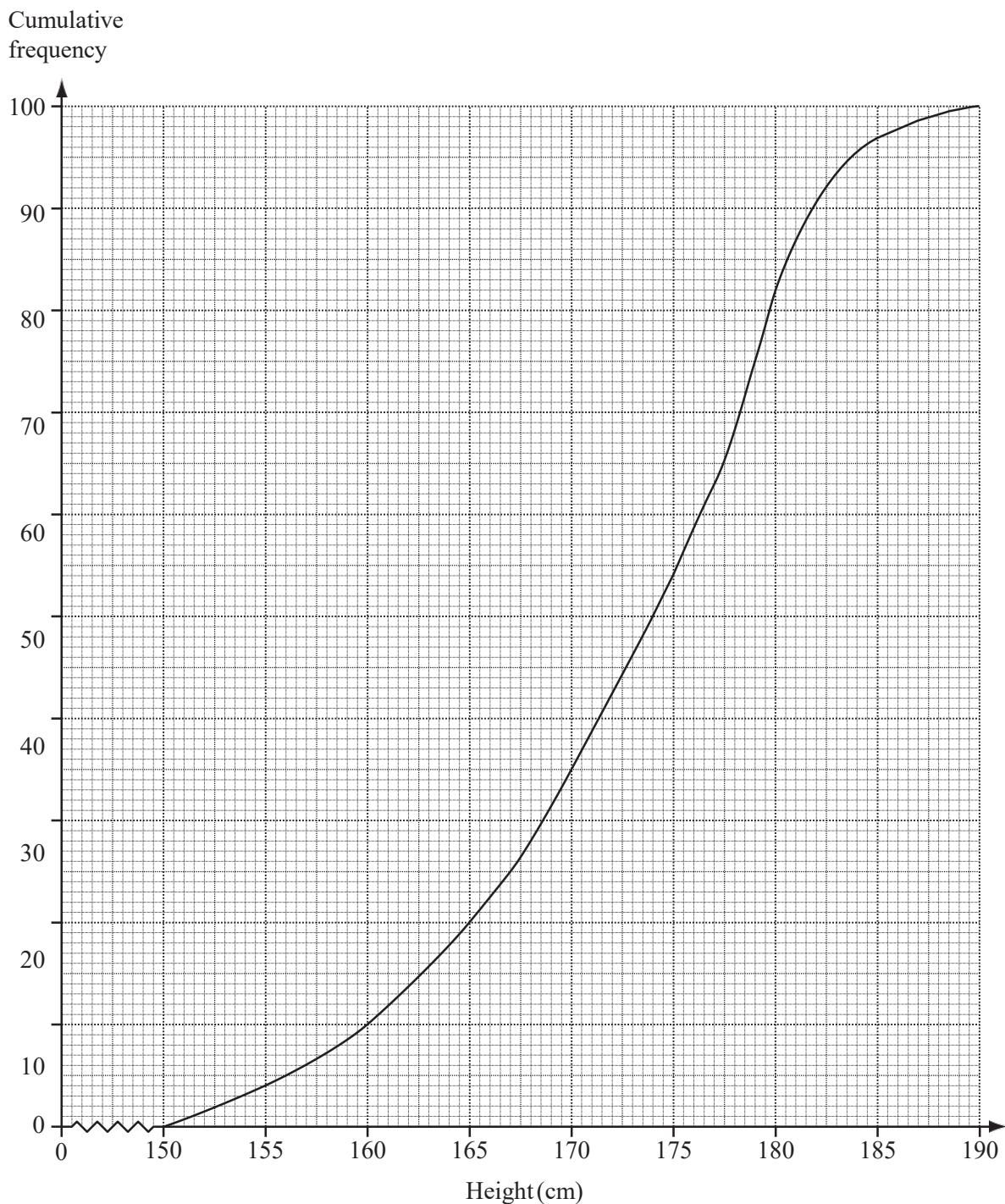
[4]



Question 7

The heights of 100 students are measured.

The results have been used to draw this cumulative frequency diagram.



(a) Find

(i) the median height,

[1]

The median represents the height corresponding to the middle cumulative frequency value.

In our case, the middle frequency value is 50.

Median height = 174 cm

(ii) the lower quartile,

[1]

The lower quartile represents the height corresponding to $\frac{1}{4}$ of the cumulative frequency value.

In our case, $\frac{1}{4}$ of the cumulative frequency is 25.

Lower quartile = 167

(iii) the inter-quartile range,

[1]

The interquartile range represents the difference between the higher and the lower quartiles.

The higher quartile represents the height corresponding to $\frac{3}{4}$ of the cumulative frequency value.

In our case, $\frac{3}{4}$ of the cumulative frequency is 75.

higher quartile = 179

The interquartile range = $179 - 167$

= **12**

- (iv) the number of students with a height greater than 177 cm. [2]

The cumulative frequency corresponding to a height of 177 cm is: 63.

The total cumulative frequency is 100.

Therefore, the number of people which are taller than 177 cm is: $100 - 63$

$$= 37$$

- (b) The frequency table shows the information about the 100 students who were measured.

Height (h cm)	$150 < h \leq 160$	$160 < h \leq 170$	$170 < h \leq 180$	$180 < h \leq 190$
Frequency			47	18

- (i) Use the cumulative frequency diagram to complete the table above. [1]

Height (h cm)	$150 < h \leq 160$	$160 < h \leq 170$	$170 < h \leq 180$	$180 < h \leq 190$
Frequency	10	25	47	18

By looking at the graph, we observe that the cumulative frequency corresponding to the interval between 150 and 160 cm is

$$10.$$

By looking at the graph, we observe that the cumulative frequency corresponding to less than 170 cm is 35.

From this value we subtract the frequency corresponding to the previous interval to obtain the frequency only for the interval between 160 and 170.

$$35 - 10$$

$$= 25$$

(ii) Calculate an estimate of the mean height of the 100 students.

[4]

To calculate an estimate of the mean height we need to consider the mid-point of each of the height intervals in the table above. This would be the height estimate representing each interval, which is then multiplied by its corresponding frequency to work out the estimate of the total height. This total height is divided by the total frequency to obtain the mean.

Mid-points for each interval: 155, 165, 175, 185.

The estimate of the total height considered is:

$$155 \times 10 + 165 \times 25 + 175 \times 47 + 185 \times 18 = 17230$$

The estimate of the mean height is: $17230/100$

$$= 172.3 \text{ cm}$$

Statistics

Difficulty: Hard

Model Answers 5

Level	IGCSE
Subject	Maths (0580/0980)
Exam Board	CIE
Topic	Statistics
Paper	Paper 4
Difficulty	Hard
Booklet	Model Answers 5

Time allowed: 90 minutes

Score: /78

Percentage: /100

Grade Boundaries:

CIE IGCSE Maths (0580)

A*	A	B	C	D
>83%	67%	51%	41%	31%

CIE IGCSE Maths (0980)

9	8	7	6	5	4
>95%	87%	80%	69%	58%	46%

Question 1

(a) Each student in a class is given a bag of sweets.

The students note the number of sweets in their bag.

The results are shown in the table, where $0 \leq x < 10$.

Number of sweets	30	31	32
Frequency (number of bags)	10	7	x

(i) State the mode.

[1]

30

(ii) Find the possible values of the median.

[3]

Consider the extreme values of x : 0 and 10.

If $x = 0$ the median will be the 9th value which is 30.

If $x = 10$ the median will be the 14th value, which is 31.

Hence the median can only be

30, 30.5, 31

(iii) The mean number of sweets is 30.65.

Find the value of x .

[3]

$$\mu = \frac{(30 \times 10 + 31 \times 7 + 32x)}{17 + x} = 30.65$$

$$\rightarrow 30.65 \times (17 + x) = 517 + 32x$$

$$\rightarrow 1.35x = 521.05 - 517$$

$$\rightarrow x = \frac{4.05}{1.35}$$

$$= 3$$

(b) The mass, m grams, of each of 200 chocolates is noted and the results are shown in the table.

Mass (m grams)	$10 < m \leq 20$	$20 < m \leq 22$	$22 < m \leq 24$	$24 < m \leq 30$
Frequency	35	115	26	24

(i) Calculate an estimate of the mean mass of a chocolate.

[4]

Mass (m grams)	Frequency (f)	Midpoint (M)	$f \times M$
$10 < m \leq 20$	35	15	525
$20 < m \leq 22$	115	21	2415
$22 < m \leq 24$	26	23	598
$24 < m \leq 30$	24	27	648
Sum	200		4183

We have that the estimate of the mean is

$$\mu = \frac{\sum f \times M}{\sum f}$$

$$= \frac{4183}{200}$$

$$= 20.9$$

(ii) On a histogram, the height of the column for the $20 < m \leq 22$ interval is 11.5 cm.

Calculate the heights of the other three columns.

Do not draw the histogram.

[5]

Height is the frequency density, so

$$f = (22 - 20) \times 11.5 \times s$$

Where s is the scalar on the y -axis

$$= 2s \times 11.5$$

$$= 23s$$

$$\rightarrow 23s = 115$$

$$\rightarrow s = 5$$

The $10 < m \leq 20$ column is

$$\frac{35}{10s}$$

$$= 0.7 \text{ cm}$$

The $22 < m \leq 24$ column is

$$\frac{26}{2s}$$

$$= 2.6 \text{ cm}$$

The $24 < m \leq 30$ column is

$$\frac{24}{6s}$$

$$= 0.8 \text{ cm}$$

Question 2

- (a) The numbers 0, 1, 1, 1, 2, k , m , 6, 9, 9 are in order ($k \neq m$).
Their median is 2.5 and their mean is 3.6.

(i) Write down the mode.

[1]

Mode = most frequent value in the data set

Mode = 1

(ii) Find the value of k .

[1]

Median = middle value which separates the data set in 2 halves.

The number of values is even, 10, therefore, the median will be the mean of the 2 middle numbers.

$$2.5 = \frac{2+k}{2}$$

k = 3

(iii) Find the value of m .

[2]

$$\text{mean} = \frac{\text{the sum of all values}}{\text{the number of values}}$$

$$\text{mean} = \frac{0+1+1+1+2+k+m+6+9+9}{10}$$

$$\frac{29+k+m}{10} = 3.6$$

$$k + m = 7$$

$$m = 4$$

- (iv) Maria chooses a number at random from the list.

The probability of choosing this number is $\frac{1}{5}$. Which number does she choose?

[1]

$$\text{Probability} = \frac{\text{number of favourable cases}}{\text{number of total cases}}$$

The number of favourable cases is the number of times the value appears in the set and the number of total cases is the total number of values in the set, 10.

$$\frac{1}{5} = \frac{\text{the number of times the value appears}}{10}$$

The value appears twice so the value is 9.

- (b) 100 students are given a question to answer.

The time taken (t seconds) by each student is recorded and the results are shown in the table.

t	$0 < t \leq 20$	$20 < t \leq 30$	$30 < t \leq 35$	$35 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$	$60 < t \leq 80$
Frequency	10	10	15	28	22	7	8

- (i) Calculate an estimate of the mean time taken.

[4]

To work out the estimate of the mean we multiply each frequency by the midpoint of its corresponding interval. The estimate amount of time is summed up and divided by the total frequency.

$$\text{Mean} = \frac{10 \times 10 + 10 \times 25 + 15 \times 32.5 + 28 \times 37.5 + 22 \times 45 + 7 \times 55 + 8 \times 70}{100}$$

$$\text{Mean} = 38.2$$

(ii) Two students are picked at random.

What is the probability that they both took more than 50 seconds?

Give your answer as a fraction in its lowest terms.

[3]

The number of students that took more than 50 seconds is: $8 + 7 = 15$.

The probability that the first student picked took more than 50 seconds is: $P = \frac{15}{100}$

After the first student is picked, the number of total students and the number of those who took more than 50 s is reduced by 1.

The probability that the second student picked took more than 50 seconds is: $P = \frac{14}{99}$

The probability that both took more than 50s = $\frac{15}{100} \times \frac{14}{99} = \frac{210}{9900}$

Simplified: $P = \frac{7}{330}$

Answer part (c) on a sheet of graph paper.

(c) The data in part (b) is re-grouped to give the following table.

t	$0 < t \leq 30$	$30 < t \leq 60$	$60 < t \leq 80$
Frequency	p	q	8

(i) Write down the values of p and q .

[2]

$$p = 10 + 10 = 20$$

$$q = 15 + 28 + 22 + 7 = 72$$

- (ii) Draw an accurate histogram to show these results.

Use a scale of 1 cm to represent 5 seconds on the horizontal time axis.

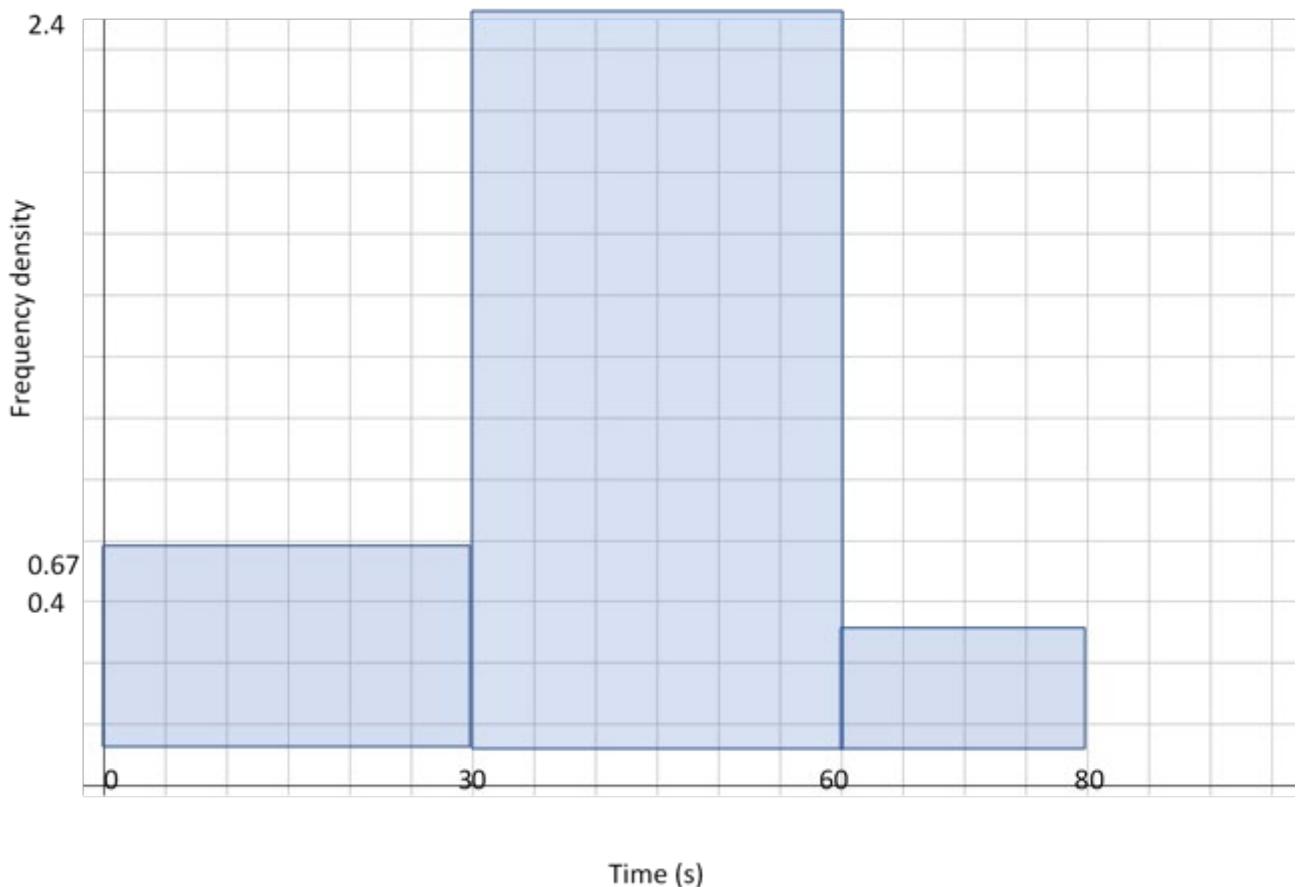
Use a scale of 1 cm to 0.2 units of frequency density (so that 1 cm² on your histogram represents 1 student). [4]

$$\text{frequency density} = \text{frequency}/\text{class width}$$

$$\text{For } 0 < t \leq 30, \text{ the frequency density} = 20/30 = 0.67$$

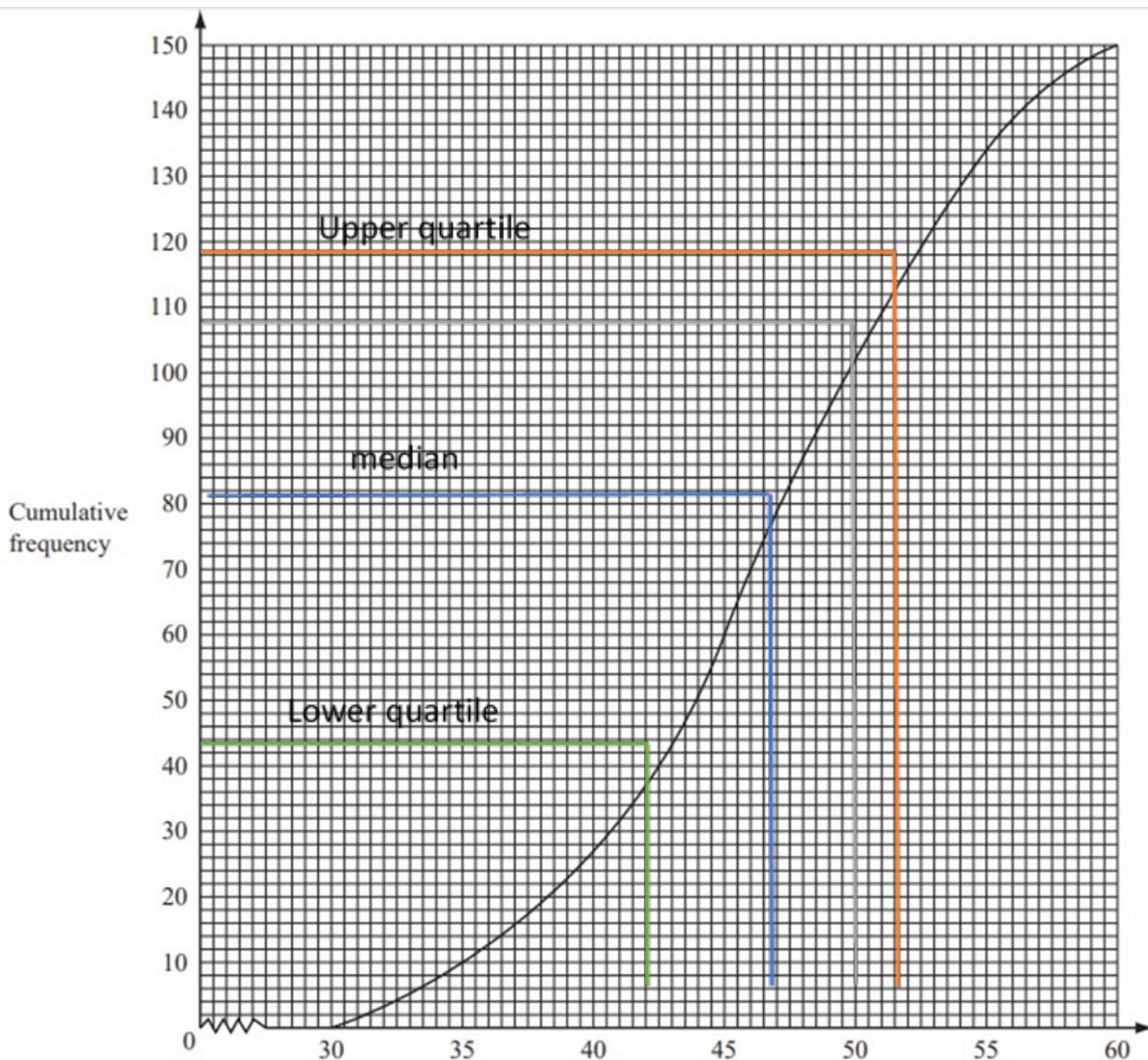
$$\text{For } 30 < t \leq 60, \text{ the frequency density} = 72/30 = 2.4$$

$$\text{For } 60 < t \leq 80, \text{ the frequency density} = 8/20 = 0.4$$



Question 3

The speeds (v kilometres/hour) of 150 cars passing a 50 km/h speed limit sign are recorded. A cumulative frequency curve to show the results is drawn below.



(a) Use the graph to find

(i) the median speed,

[1]

The total cumulative frequency represented on the y axis is 250, the total number of students.

The median represents the middle value in a group.

In this case, the median will be $75 = \frac{150}{2}$

The x coordinate corresponding to $y = 75$ represents the median height

, $x = 46.5$

- (ii) the inter-quartile range of the speeds, [2]

To work out the upper quartile we divide the total cumulative frequency by 4. The upper quartile will represent the x value corresponding to the y value representing 3 quarters of the total frequency. The lower quartile will represent the x value corresponding to the y value representing one quarter of the total frequency.

$$\frac{150}{4} = 37.5$$

The upper quartile will be the x value corresponding to y = 112.5.

$$x = 51.5$$

The lower quartile will be the x value corresponding to y = 37.5.

$$x = 42$$

The inter-quartile range is the difference between the upper quartile and the lower quartile.

The inter-quartile range is:

$$51.5 - 42 = 9.5$$

- (iii) the number of cars travelling with speeds of more than 50 km/h. [2]

On the graph, the value corresponding to a speed of 50 km/h is 102 cars. The number of cars which travel with more than 50 km/h is the total number of cars subtracted by the number of cars which travel slower than 50 km/h.

$$\text{Number of cars} = 150 - 102 = 48$$

(b) A frequency table showing the speeds of the cars is

Speed (v km/h)	$30 < v \leq 35$	$35 < v \leq 40$	$40 < v \leq 45$	$45 < v \leq 50$	$50 < v \leq 55$	$55 < v \leq 60$
Frequency	10	17	33	42	n	16

(i) Find the value of n . [1]

The total frequency, number of cars, is 150.

$$n = 150 - 10 - 17 - 33 - 42 - 16$$

$$\mathbf{n = 32}$$

(ii) Calculate an estimate of the mean speed. [4]

To estimate the mean speed, we know for example that there are 10 cars with the speed between 30 km/h and 35 km/h. Since we do not know an exact speed, we will estimate that each car in this group has exactly 32.5 km/h, the midpoint of the group.

Similarly, we work out the mid-point for every interval.

Speed (km/h)	Frequency	Mid-point	Mid-point x Frequency
$30 < v \leq 35$	10	32.5	325
$35 < h \leq 40$	17	37.5	637.5
$40 < h \leq 45$	33	42.5	1402.5
$45 < h \leq 50$	42	47.5	1995
$50 < h \leq 55$	32	52.5	1680
$55 < h \leq 60$	16	57.5	920
Total	150	-	6960

We then calculate the mid-point \times frequency for each interval to work out the total speed of all the cars, estimating the mid-point for each height interval.

$$\text{Mean} = \frac{6960}{150}$$

$$\text{Mean} = 46.4 \text{ km/h}$$

(c) Answer this part of this question on a sheet of graph paper.

Another frequency table for the same speeds is

Speed (v km/h)	$30 < v \leq 40$	$40 < v \leq 55$	$55 < v \leq 60$
Frequency	27	107	16

Draw an accurate histogram to show this information.

Use 2 cm to represent 5 units on the speed axis and 1 cm to represent 1 unit on the frequency density axis (so that 1 cm² represents 2.5 cars). [5]

To draw a histogram, we need to know the frequency density for each interval.

Frequency density = frequency/group width

For $30 < v \leq 40$:

Freq density = $27 / 10$

Freq density = 2.7

For $40 < v \leq 55$:

Freq density = $107 / 15$

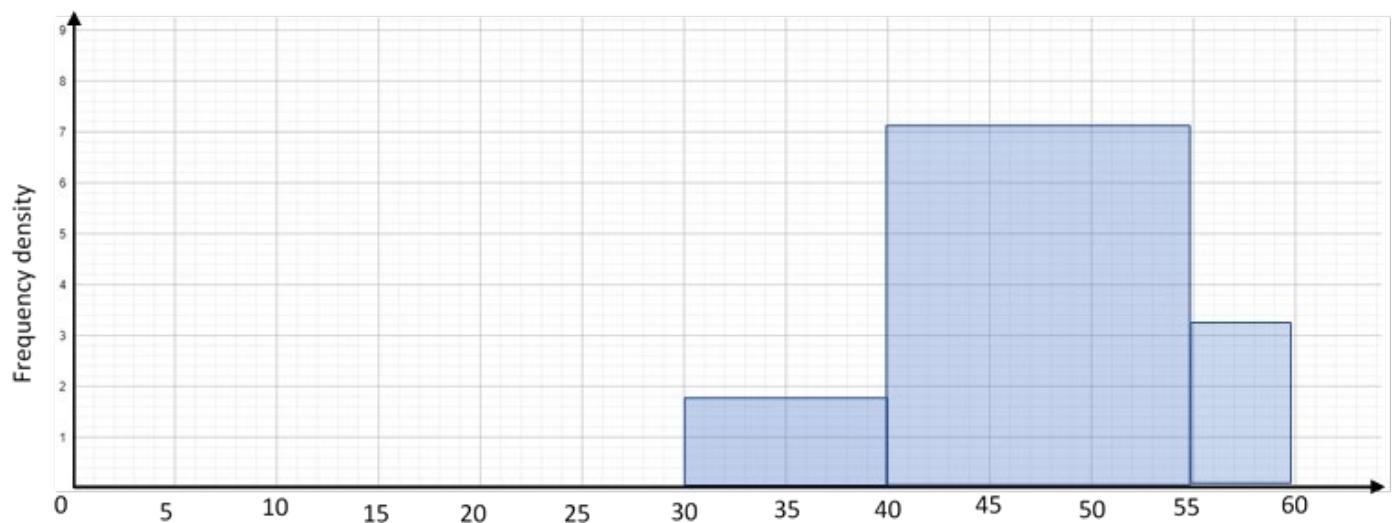
Freq density = 7.1

For $55 < v \leq 60$:

$$\text{Freq density} = 16/5$$

$$\text{Freq density} = 3.2$$

Using these values, we obtain the histogram below:



Question 4

Answer the whole of this question on a sheet of graph paper.

120 passengers on an aircraft had their baggage weighed. The results are shown in the table.

Mass of baggage (M kg)	$0 < M \leq 10$	$10 < M \leq 15$	$15 < M \leq 20$	$20 < M \leq 25$	$25 < M \leq 40$
Number of passengers	12	32	28	24	24

- (a) (i) Write down the modal class. [1]

The modal class represents the interval which corresponds to the highest frequency of passengers.

Modal class = $10 < M \leq 15$ with a frequency of 32.

- (ii) Calculate an estimate of the mean mass of baggage for the 120 passengers. Show all your working. [4]

To work out the estimate of the mean mass we assume that the frequency for each interval corresponds to the mid-point of the interval. We multiply the frequency by this corresponding precise weight and then sum up to obtain the total estimated weight, which we divide by the total frequency.

Mass of baggage (M kg)	$0 < M \leq 10$	$10 < M \leq 15$	$15 < M \leq 20$	$20 < M \leq 25$	$25 < M \leq 40$	Total
Nr. of passengers	12	32	28	24	24	120
Mid-point	5	12.5	17.5	22.5	32.5	-
Mid-point x frequency	60	400	490	540	780	2270

$$\text{Mean} = \frac{2270}{120}$$

Mean = 18.9 kg

(iii) Sophia draws a pie chart to show the data.

What angle should she have in the $0 < M \leq 10$ sector?

[1]

The total pie chart would represent the total of 120 people by 360° .

Therefore, the interval $0 < M \leq 10$ corresponding to 12 people can be

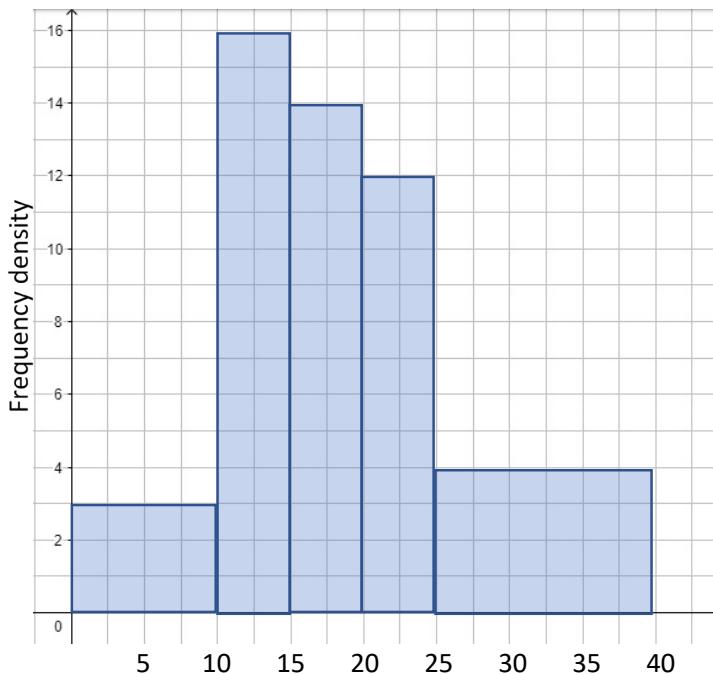
represented by an angle of

36° .

(b) Using a scale of 2 cm to represent 5 kg, draw a horizontal axis for $0 < M \leq 40$.

Using an area scale of 1 cm^2 to represent 1 passenger, draw a histogram for this data.

[7]



1 cm² represents 1 passenger.

For example, for the first interval $0 < M \leq 10$, the x axis shows 4 cm² because $2 \text{ cm}^2 = 5 \text{ kg}$.

$$12 \text{ passengers}/4 \text{ cm}^2 = 3$$

The height of the histogram will be 3 units.

Similarly, we draw the histogram for all the intervals.

k = 1

Question 5

In a survey, 200 shoppers were asked how much they had just spent in a supermarket. The results are shown in the table.

Amount(\$x)	$0 < x \leq 20$	$20 < x \leq 40$	$40 < x \leq 60$	$60 < x \leq 80$	$80 < x \leq 100$	$100 < x \leq 140$
Number of shoppers	10	32	48	54	36	20

- (a) (i) Write down the modal class. [1]

The modal class is the interval with the highest frequency.

In this case: **modal class = $60 < x \leq 80$**

- (ii) Calculate an estimate of the mean amount, giving your answer correct to 2 decimal places. [4]

Amount (\$x)	$0 < x \leq 20$	$20 < x \leq 40$	$40 < x \leq 60$	$60 < x \leq 80$	$80 < x \leq 100$	$100 < x \leq 140$	Total
Nr. of shoppers	10	32	48	54	36	20	200
Mid-point	10	30	50	70	90	120	-
Mid-point x frequency	100	960	2400	3780	3240	2400	12880

To work out the estimate of the mean we assume that the frequency corresponding to each interval is precisely for its mid-point. By multiplying the mid-point by the frequency and adding up all the values we obtain the total amount of money spent, which we can divide by the frequency to obtain the mean.

$$\text{Mean} = \frac{12880}{200}$$

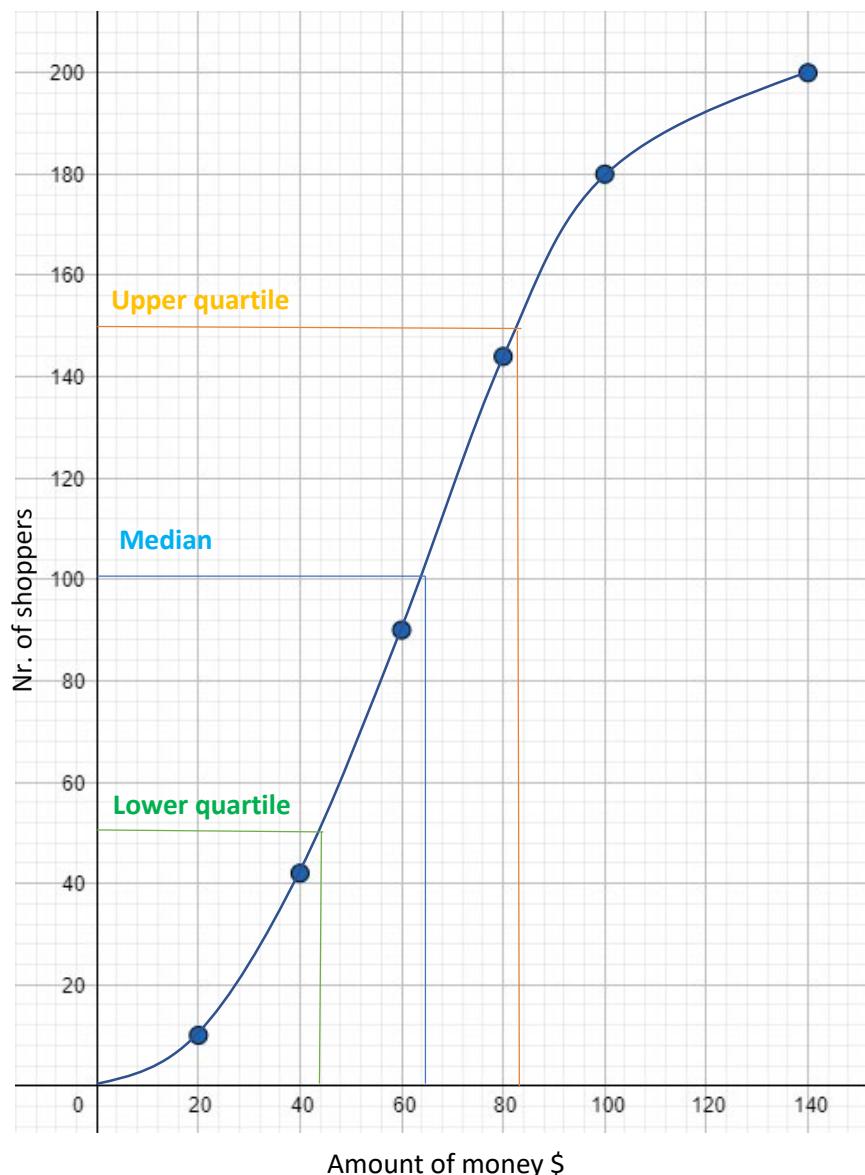
$$\text{Mean} = 64.4$$

- (b) (i) Make a cumulative frequency table for these 200 shoppers. [2]

We obtain a cumulative frequency by adding up the frequency corresponding to the interval with the one from the previous interval.

Amount (\$x)	$x \leq 20$	$x \leq 40$	$x \leq 60$	$x \leq 80$	$x \leq 100$	$x \leq 140$
Cumulative frequency	10	$10 + 32 = 42$	$42 + 48 = 90$	$90 + 54 = 144$	$144 + 36 = 180$	$180 + 20 = 200$

- (ii) Using a scale of 2 cm to represent \$20 on the horizontal axis and 2 cm to represent 20 shoppers on the vertical axis, draw a cumulative frequency diagram for this data. [4]



- (c) Use your cumulative frequency diagram to find
(i) the median amount,

[1]

The median represents the amount corresponding to the middle value of the number of shoppers, 100.

Median amount = 64

- (ii) the upper quartile,

[1]

The upper quartile represents the amount corresponding to 3/4 of the number of shoppers, 150.

Upper quartile = 84

- (iii) the interquartile range,

[1]

The interquartile range is the difference between the upper and the lower quartile.

Lower quartile = 44

IQR = 84 – 44 = 40

- (iv) how many shoppers spent at least \$75.

[2]

The y value corresponding to x = \$75 is y = 126 approximately.

The number of people who spent more than \$75 is: 200 – 126
= \$74