Cambridge IGCSE · Mathematics 0580 · Notes

Cumulative frequency and histograms

Read online at **pastopic.com/cambridge/mathematics-0580/notes/cumulative-frequency-and-histograms**

Last checked 29 September 2026

Read first: **Averages and measures of spread**

What you need to know

Extended only: this whole topic is on the Extended papers (Papers 2 and 4) only. By the end of it you should be able to:

1. **Make and read cumulative frequency tables**: turn a grouped frequency table into running totals, and go back again.
2. **Draw and read cumulative frequency diagrams**: plot each running total at the upper end of its class, mark the points clearly (small crosses) and join them with a smooth curve.
3. **Estimate from a cumulative frequency diagram** the median, the lower and upper quartiles, the interquartile range and any percentile, and say what these tell you about the data.
4. **Draw and read histograms** for grouped data with equal or unequal class widths, with the vertical axis labelled **frequency density**.
5. **Calculate with frequency density**: frequency density $=$ frequency $\div$ class width, so frequency $=$ frequency density $\times$ class width.

Working out an estimate of the mean from a grouped table is in the note on averages and measures of spread, but it turns up in almost every question on this topic, so "How to do it" goes through it again. Box-and-whisker plots are not in the 2025–2027 syllabus.

Understand it

Grouped data and class boundaries

Continuous data, such as times, masses and heights, is put into classes like $20 < t \leqslant 30$. This class holds every value above 20 and up to and including 30. Its **width** is $30 - 20 = 10$, its **lower boundary** is 20 and its **upper boundary** is 30. Classes don't have to be the same width: where values are crowded together, narrow classes show more detail, and where they are spread out, wide classes save space.

Why a histogram uses frequency density

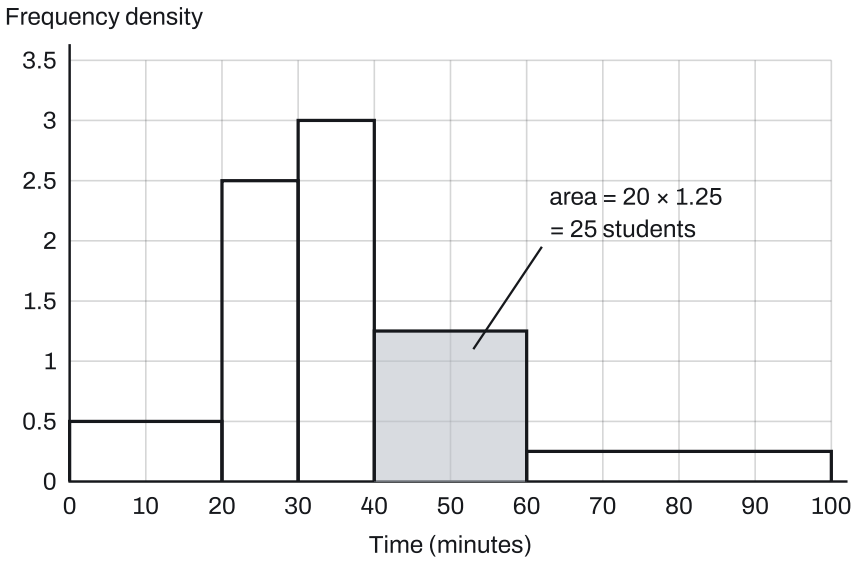
In a bar chart the height of each bar is its frequency. That goes wrong when the classes have different widths. If a class twice as wide as the others holds twice as many values, a tall bar makes it look crowded, when really the values are just as spread out as their neighbours.

A histogram fixes this by making the **area** of each bar stand for its frequency. The height is then the frequency per unit of width:

$$\text{frequency density} = \frac{\text{frequency}}{\text{class width}}.$$

The table shows the time, t minutes, that 100 students spent on homework one evening.

Time (t minutes)	$0 < t \leqslant 20$	$20 < t \leqslant 30$	$30 < t \leqslant 40$	$40 < t \leqslant 60$	$60 < t \leqslant 100$
Frequency	10	25	30	25	10
Class width	20	10	10	20	40
Frequency density	0.5	2.5	3	1.25	0.25



The $20 < t \leqslant 30$ and $40 < t \leqslant 60$ classes both hold 25 students, but the second is twice as wide, so its bar is half as tall. The tallest bar shows where the data is most **densely packed**, not the class with the most students.

Reading a histogram works backwards: **frequency** = **frequency density** × **class width**, the area of the bar. You can also estimate how many values lie in **part** of a class by assuming the values are spread evenly across it. For example, "more than 70 minutes" is the part of the last bar from 70 to 100: $30 \times 0.25 = 7.5$, so about 7 or 8 students.

Cumulative frequency

Cumulative frequency is a running total: how many values are **less than or equal to** each upper boundary.

Time (t minutes)	$t \leqslant 0$	$t \leqslant 20$	$t \leqslant 30$	$t \leqslant 40$	$t \leqslant 60$	$t \leqslant 100$
Cumulative frequency	0	10	35	65	90	100

Each total is the one before plus the next frequency, and the last total is the number of values, 100. Going back, a class's frequency is the difference between two neighbouring totals: $65 - 35 = 30$ students took between 30 and 40 minutes.

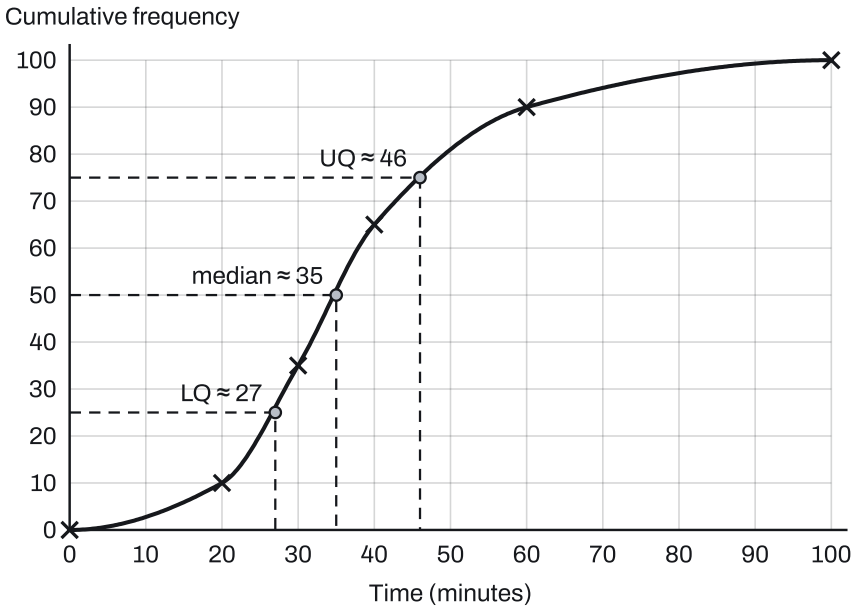
To draw the **cumulative frequency diagram**, plot each total above the **upper boundary** of its class, because only at the end of the class have all its values been counted. Start at the lowest boundary with a total of 0, here $(0, 0)$. Join the crosses with a smooth curve (or with straight lines, which mark schemes also accept). The curve always goes up from left to right, and it is steepest where the data is packed most tightly.

Median, quartiles and percentiles from the curve

The curve turns any count into a value. Go **across** from a cumulative frequency to the curve, then **down** to the horizontal axis.

- **Median**: read across from half the total, $\frac{n}{2}$.
- **Lower quartile (LQ)**: read across from a quarter of the total, $\frac{n}{4}$. A quarter of the values are below it.
- **Upper quartile (UQ)**: read across from three-quarters of the total, $\frac{3n}{4}$.
- **Interquartile range (IQR)** $=$ UQ $-$ LQ, the spread of the middle half of the data.
- **Percentiles**: the 90th percentile is read across from 90% of the total, and so on. The median is the 50th percentile, and the quartiles are the 25th and 75th.

For large grouped data use $\frac{n}{2}$, $\frac{n}{4}$ and $\frac{3n}{4}$, not the $\frac{n+1}{2}$ rule for a short list.



For the 100 students: median ≈ 35 minutes, LQ ≈ 27 minutes, UQ ≈ 46 minutes, so the IQR $\approx 46 - 27 = 19$ minutes. The 90th percentile is read across from 90, which is exactly 60 minutes.

The curve also works the other way round: **up** from a value, then **across** to a count. At $t = 50$ the curve reads about 81, so about 81 students took 50 minutes or less, and about $100 - 81 = 19$ took more than 50 minutes.

These readings are **estimates**, because the curve guesses how values are spread inside each class. Mark schemes accept a small range of answers.

What the numbers tell you. Half the students did less homework than the median, and half did more. The middle half took between about 27 and 46 minutes; a small IQR means the middle values are close together (consistent), and a large IQR means they are spread out. When you compare two groups, compare an average (the median) **and** a spread (the IQR), in the context of the question.

Key facts and formulas

The 0580 Extended papers print no list of formulas for this topic, so you must learn every one.

Formula	Status
Class width $=$ upper boundary $-$ lower boundary	Learn it
Frequency density $=$ frequency $\div$ class width	Learn it
Frequency $=$ frequency density $\times$ class width (area of the bar)	Learn it
Cumulative frequency at an upper boundary $=$ total of all frequencies up to that class	Learn it
Plot cumulative frequencies at the **upper** class boundaries, starting from 0 at the lowest boundary	Learn it
Median: read at $\frac{n}{2}$; LQ at $\frac{n}{4}$; UQ at $\frac{3n}{4}$	Learn it
IQR $=$ UQ $-$ LQ	Learn it
pth percentile: read at $\frac{p}{100} \times n$	Learn it
Number **more than** $x = n-$ reading at x	Learn it
Estimated mean $= \dfrac{\sum fx}{\sum f}$, with x the midpoint of each class	Learn it

How to do it

In the Paper 2 and Paper 4 questions we have tagged for this topic (30 questions, 2021–2025), the types came up this often. 27 of the 30 questions mix two or more types, so the counts add up to more than 30.

Question type	Questions
Estimate of the mean from grouped data	22
Draw or complete a histogram, or find bar heights	20
Read the median, quartiles, IQR or percentiles from a cumulative frequency diagram	14
Complete a cumulative frequency table and draw the diagram	8

Question type	Questions
Read frequencies from a histogram	7
Box-and-whisker plot from a cumulative frequency diagram (papers before 2025)	3

1. Estimate of the mean from grouped data (22 questions)

This is usually the first part, with the histogram or cumulative frequency diagram after it.

1. **Find the midpoint** of each class: $\frac{\text{lower}+\text{upper}}{2}$. For $20 < t \leqslant 40$ it is 30.
2. **Multiply** each midpoint by its frequency and add: $\sum fx$.
3. **Divide** by the total frequency $\sum f$.
4. Give the answer to 3 significant figures, or exactly if it is shorter (such as 53 or 48.375).

For the homework times: $\dfrac{10 \times 10 + 25 \times 25 + 30 \times 35 + 25 \times 50 + 10 \times 80}{100} = \dfrac{3825}{100} = 38.25$ minutes.

Variations you will meet:

- **The data is only in a histogram.** Find every frequency first with frequency density $\times$ class width, then use the midpoints as above.
- **Why it is an estimate:** you don't know the actual values, only which class each is in, so each is replaced by its class's midpoint.

2. Draw or complete a histogram, or find bar heights (20 questions)

1. **Work out each class width** from the boundaries (watch out: they are often unequal).
2. **Frequency density** $=$ **frequency** $\div$ **class width** for each class. Write these values down: if a bar is drawn slightly wrong, the written density can still earn a mark.
3. **Check the scale** on the frequency density axis: count what one small square is worth.
4. **Draw each bar** with a ruler, from its exact lower boundary to its exact upper boundary, up to its frequency density. Bars touch; there are no gaps.

Variations you will meet:

- **"Complete the histogram"** with some bars already drawn: work out only the missing bars, but check one drawn bar to be sure of the scale.
- **A missing frequency.** If you know the total, read the other bars' frequencies (density $\times$ width), subtract from the total, then divide by the width for the missing bar's height.
- **Bar heights in centimetres.** "The bar for the first class is 6 cm tall; find the heights of the others." Work out every frequency density, find the scale factor from the known bar (height in cm $\div$ its frequency density), then multiply the other densities by it. Never use a scale factor from the frequencies: it only works if all widths are equal.
- **Merged classes.** If two classes are combined into one, add their frequencies and divide by the new, wider class width.

3. Read the median, quartiles, IQR or percentiles from a cumulative frequency diagram (14 questions)

1. **Find the total** n: it is the last cumulative frequency, and it is usually given in the question.
2. **Work out the count** to read at: $\frac{n}{2}$ for the median, $\frac{n}{4}$ and $\frac{3n}{4}$ for the quartiles, $\frac{p}{100} \times n$ for the pth percentile.
3. **Read across** from that count to the curve, then **down** to the value. Draw the lines on the graph.
4. For the IQR, subtract: UQ $-$ LQ.

Variations you will meet:

- **"How many are more than x?"** Read **up** from x to the curve and **across** to the count, then subtract it from the total. "x or less" is the reading itself.
- **"How many are between x_1 and x_2?"** Read at both values and subtract the smaller count from the larger.
- **"70% are less than k; find k."** Work out 70% of n first, then read across from that count.
- **A frequency table from the diagram.** Read the total at each boundary and subtract neighbouring totals to get each class's frequency.
- **A probability.** "One is chosen at random; find the probability it is more than x": the count from the graph divided by n.

4. Complete a cumulative frequency table and draw the diagram (8 questions)

1. **Running totals.** Add each frequency to the total before. The last total must equal the number of values in the question: a quick check.
2. **Plot** each total above the **upper boundary** of its class, with a small cross. Include the starting point: the lowest boundary with cumulative frequency 0.
3. **Join** the crosses with a smooth increasing curve (or ruled straight lines). Don't draw bars, and don't plot at the midpoints.

5. Read frequencies from a histogram (7 questions)

1. **Read the height** (frequency density) of the bar from the scale, carefully.
2. **Frequency $=$ frequency density $\times$ class width.**
3. For "show that there are 10 in this class", write the calculation, such as $(1.6 - 1.35) \times 40 = 10$.

Variations you will meet:

- **Part of a class.** For "more than 35" when a class runs from 20 to 50, assume the values are spread evenly: take the fraction $\frac{50-35}{50-20} = \frac{1}{2}$ of that bar's frequency, then add every class above it.
- **A letter in the table.** If the table says the frequency of a class is x and the histogram shows its bar, x is that bar's density $\times$ width. If it is the other way round, use the known frequency with the drawn bar to find the scale, then draw the rest.

6. Box-and-whisker plot from a cumulative frequency diagram (3 questions)

Papers before 2025 only. Box-and-whisker plots are not in the 2025–2027 syllabus, so you won't meet this on your exam. In older questions, read the LQ, median and UQ from the curve as in type 3, draw a box from LQ to UQ with a line at the median, and whiskers out to the smallest and largest values given in the question.

Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- **M1** method mark: for a correct method, even if the numbers slip.
- **A1** accuracy mark: for a correct answer; it needs the M mark before it.
- **B1** mark for a correct result on its own, with no method needed.

Example 1

The table shows the length, L cm, of each of 80 fish.

Length (L cm)	$10 < L \leqslant 20$	$20 < L \leqslant 25$	$25 < L \leqslant 30$	$30 < L \leqslant 40$	$40 < L \leqslant 60$
Frequency	8	18	26	20	8

(a) Calculate an estimate of the mean length. **[4]**

(b) Find the frequency density of each class, ready to draw a histogram. **[2]**

(c) A different histogram of the same data is drawn with the bar for $10 < L \leqslant 20$ at a height of 2 cm. Calculate the height of the bar for $25 < L \leqslant 30$. **[2]**

(a) Midpoints $15, 22.5, 27.5, 35, 50$. **M1**

$\sum fx = 8 \times 15 + 18 \times 22.5 + 26 \times 27.5 + 20 \times 35 + 8 \times 50 = 120 + 405 + 715 + 700 + 400 = 2340$. **M1** for $\sum fx$ with values in the right classes.

Mean $\approx 2340 \div 80 = 29.25$ cm. **M1** for dividing by 80, **A1**.

(b) Class widths $10, 5, 5, 10, 20$. Frequency densities:

$$\frac{8}{10} = 0.8, \quad \frac{18}{5} = 3.6, \quad \frac{26}{5} = 5.2, \quad \frac{20}{10} = 2, \quad \frac{8}{20} = 0.4.$$

B2 for all five (**B1** for three or four).

(c) Scale factor from the known bar: $2 \div 0.8 = 2.5$ cm per unit of frequency density. **M1**

Height $= 5.2 \times 2.5 = 13$ cm. **A1**

Using the frequencies instead ($2 \div 8 = 0.25$, so $26 \times 0.25 = 6.5$ cm) is wrong, because the two classes have different widths.

Example 2

The table shows the mass, m grams, of each of 120 apples.

Mass (m grams)	$100 < m \leqslant 120$	$120 < m \leqslant 140$	$140 < m \leqslant 160$	$160 < m \leqslant 180$	$180 < m \leqslant 200$	$200 < m \leqslant 220$
Frequency	9	21	42	30	12	6

(a) Complete the cumulative frequency table and draw the cumulative frequency diagram. **[5]**

(b) Use your diagram to find an estimate of

(i) the median, **[1]**

(ii) the interquartile range, **[2]**

(iii) the 80th percentile, **[2]**

(iv) the number of apples with a mass greater than 190 g. **[2]**

(a)

Mass (m grams)	$m \leqslant 100$	$m \leqslant 120$	$m \leqslant 140$	$m \leqslant 160$	$m \leqslant 180$	$m \leqslant 200$	$m \leqslant 220$
Cumulative frequency	0	9	30	72	102	114	120

B2 for the table (the last total is 120, as it should be).

Plot $(100, 0)$, $(120, 9)$, $(140, 30)$, $(160, 72)$, $(180, 102)$, $(200, 114)$, $(220, 120)$ and join them with a smooth curve. **B1** for the points at the right masses, **B1** for the right heights, **B1** for an increasing curve through them.

(b)(i) $\frac{120}{2} = 60$. Across from 60, down to the axis: median ≈ 154 g. **B1**

(ii) LQ: across from $\frac{120}{4} = 30$, which is a plotted point, so LQ $= 140$ g. UQ: across from $\frac{3 \times 120}{4} = 90$, giving about 170 g. **B1** for one quartile.

IQR $\approx 170 - 140 = 30$ g. **B1**

(iii) 80% of $120 = 96$. **B1** for 96. Across from 96: about 175 g. **B1**

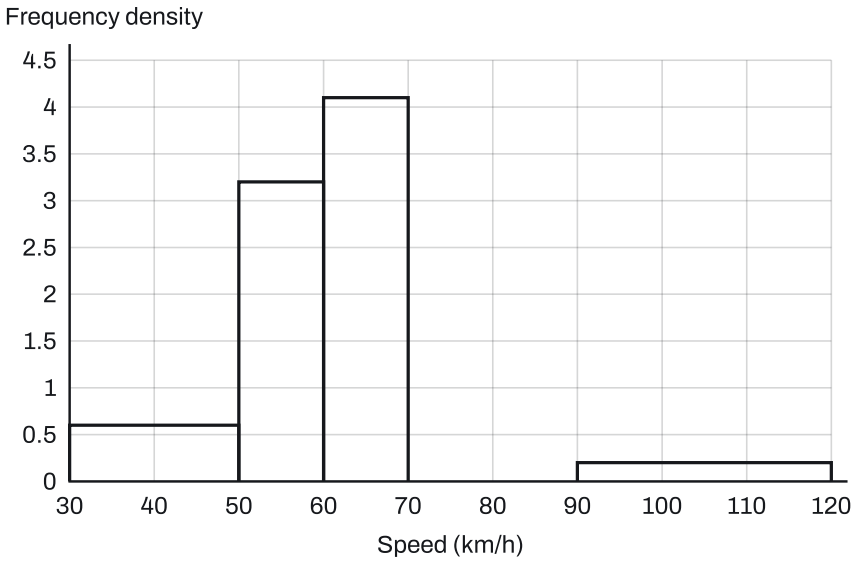
(iv) Up from 190 g: the curve reads about 109 apples with mass 190 g or less. **B1**

Greater than 190 g: $120 - 109 = 11$ apples. **B1**

A smooth curve drawn by hand will give readings a little different from these; the mark scheme accepts a range around each value.

Example 3

The histogram shows the speeds, v km/h, of some cars. The bar for $70 < v \leqslant 90$ has not been drawn.



(a) Show that 32 cars have a speed in the interval $50 < v \leqslant 60$. **[1]**

(b) There are 115 cars altogether. Complete the histogram. **[3]**

(c) Calculate an estimate of the number of cars with a speed greater than 80 km/h. **[2]**

(d) Calculate an estimate of the mean speed. **[4]**

(a) $3.2 \times (60 - 50) = 32$. **B1**

(b) Frequencies of the drawn bars: $0.6 \times 20 = 12$, 32, $4.1 \times 10 = 41$ and $0.2 \times 30 = 6$. **M1** for these.

Missing frequency $= 115 - (12 + 32 + 41 + 6) = 115 - 91 = 24$. **M1**

Height $= 24 \div 20 = 1.2$: draw a bar from 70 to 90 at height 1.2. **A1**

(c) From 80 to 90 is half of the 70–90 class, so about $\frac{1}{2} \times 24 = 12$ cars. **M1**

Add the 6 cars above 90: about $12 + 6 = 18$ cars. **A1**

(d) Midpoints $40, 55, 65, 80, 105$. **M1**

$\sum fx = 12 \times 40 + 32 \times 55 + 41 \times 65 + 24 \times 80 + 6 \times 105 = 480 + 1760 + 2665 + 1920 + 630 = 7455$. **M1**

Mean $\approx 7455 \div 115 = 64.826\ldots = 64.8$ km/h (3 s.f.). **M1 A1**

Common mistakes

- **Using frequency as the height of a histogram bar.** Examiners see bars drawn at the frequency, or at the frequency divided by the same width for every class. Divide each frequency by **its own** class width.
- **Dividing by the wrong thing.** Dividing by the upper boundary or the midpoint instead of the class width is a common error. The width of $125 < m \leqslant 140$ is 15, not 10 or 140.
- **Scaling bar heights from the frequencies.** When one bar's height in cm is given, the scale factor is that height divided by its **frequency density**, not by its frequency.
- **Freehand or misread bars.** Reports regularly mention bars drawn at 0.7 instead of 0.75, or at 7.4 instead of 7.5. Count what one small square is worth, and use a ruler.
- **Plotting at the midpoints, or leaving out the start.** Cumulative frequencies go at the **upper** boundaries. The first point, at the lowest boundary with cumulative frequency 0, is the one most often missed.
- **Drawing bars instead of a curve.** A cumulative frequency diagram is a curve or a set of joined points, never a bar chart.
- **Reading at the percentage itself.** For the 75th percentile of 200 values read across from 150, not from 75. For "30% of 80", read from 24, not 30.
- **Forgetting to subtract.** "More than 75 kg" needs the total minus the reading at 75, not the reading alone. "Between 250 and 400" needs two readings.
- **Writing the count instead of the value.** A percentile or quartile is a value on the horizontal axis (a mass or a time), not a cumulative frequency.
- **Using class widths or upper boundaries in the mean.** Multiply each frequency by its class **midpoint**, and divide by the total frequency, not by the number of classes.

How to get the marks

- **Write down the frequency densities.** A histogram drawn slightly wrong earns nothing on its own, but written densities can earn a mark.
- **Draw the reading lines on the graph** and write the count you read at (such as 96 for the 80th percentile). A method mark is often for that number alone, even if the reading is off.
- **Follow-through marks.** Readings are marked from your own curve, as long as it is increasing, so a wrong point earlier doesn't lose every later mark.
- **Estimates have a range.** Read as carefully as you can and give a single value; the mark scheme accepts values close to the right one.
- **"Show that"** means write the calculation that gives the stated number, such as $(160 - 120) \times 0.2 = 8$.
- **"Complete the table"** for cumulative frequencies: the last value must be the total. If it isn't, check your additions before you draw.
- **"You must show all your working"** means the answer alone scores nothing: show the frequencies, the midpoints and $\sum fx$.
- **Accuracy:** give answers exactly or to 3 significant figures; counts of people or objects are whole numbers.

Quick check

1. The class $15 < x \leqslant 40$ has frequency 30. Find its frequency density.
2. On a histogram the bar for $60 < t \leqslant 75$ has frequency density 2.4. How many values are in this class?
3. A table has classes $0 < m \leqslant 10$, $10 < m \leqslant 15$ and $15 < m \leqslant 30$ with frequencies 12, 20 and 27. On a histogram the first bar is 3 cm tall. Find the heights of the other two bars.
4. The frequencies for $0 < x \leqslant 10$, $10 < x \leqslant 20$, $20 < x \leqslant 30$ and $30 < x \leqslant 40$ are 5, 17, 26 and 12. Write down the cumulative frequencies, the points you would plot, and the cumulative frequency at which you would read the median.
5. A cumulative frequency diagram shows the times of 200 runners. The curve reads 146 at 40 minutes. (a) How many runners took more than 40 minutes? (b) Which percentile is 40 minutes? (c) At which cumulative frequency would you read the 90th percentile?

Answers

1. Width $= 40 - 15 = 25$, so frequency density $= 30 \div 25 = 1.2$.
2. Frequency $= 2.4 \times (75 - 60) = 36$.
3. Frequency densities: $12 \div 10 = 1.2$, $20 \div 5 = 4$ and $27 \div 15 = 1.8$. Scale factor $= 3 \div 1.2 = 2.5$, so the heights are $4 \times 2.5 = 10$ cm and $1.8 \times 2.5 = 4.5$ cm.
4. Cumulative frequencies $5, 22, 48, 60$. Plot $(0, 0)$, $(10, 5)$, $(20, 22)$, $(30, 48)$ and $(40, 60)$. Read the median across from $60 \div 2 = 30$.
5. (a) $200 - 146 = 54$ runners. (b) $146 \div 200 = 73\%$, so 40 minutes is the 73rd percentile. (c) 90% of $200 = 180$.

Past questions to try

- **0580/42/O/N/25 Q16**: frequencies read from a histogram, then an estimate of the mean.
- **0580/43/M/J/25 Q24**: a table from a histogram, and a count in part of a class.
- **0580/43/M/J/23 Q6**: a cumulative frequency curve with the median, IQR and a count, then bar heights in cm.
- **0580/42/O/N/22 Q3**: the mean and bar heights from a given bar, with unequal widths.