Chapter 20: Histograms and frequency distribution diagrams

Key words

- Histogram
- Continuous
- Class interval
- Frequency
- Grouped
- Frequency tables
- Frequency density
- Modal class
- Cumulative frequency
- Cumulative frequency curve
- Quartiles
- Interquartile range
- Percentiles

In this chapter you will learn how to:

- construct and use histograms with equal intervals

EXTENDED

- construct and use histograms with unequal intervals
- draw cumulative frequency tables
- use tables to construct cumulative frequency diagrams
- identify the modal class from a grouped frequency distribution.

The diagram shown on the top left of the digital camera screen is a type of histogram which shows how light and shadows are distributed in the photograph. The peaks show that this photo (the bird) is too dark (underexposed).

You have already collected, organised, summarised and displayed different sets of data using pie charts, bar graphs and line graphs. In this section you are going to work with numerical data (sets of data where the class intervals are numbers) to learn how to draw frequency distribution diagrams called histograms and cumulative frequency curves.

Histograms are useful for visually showing patterns in large sets of numerical data. The shape of the graph allows you see where most of the measurements are located and how spread out they are.

RECAP

You should be familiar with the following work on organising and displaying data:

Averages and range for frequency data (Chapters 4 and 12)

You can find the mean, median and mode from a frequency table of discrete data values:

Add a row or column for 'frequency × values' or 'fx'

$$\text{Mean} = \frac{\text{sum of } fx}{\text{total frequency}}$$

The table shows the number of passengers carried by a sedan taxi each trip for one weekend.

Number of passengers (x)	Frequency (f)	fx
1	12	$1 \times 12 = 12$
2	5	$2 \times 5 = 10$
3	6	$3 \times 6 = 18$
4	9	$4 \times 3 = 12$
	total frequency = 26	Sum of $fx = 52$

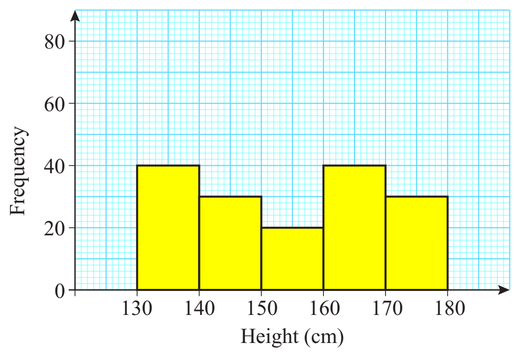
$$\text{Mean} = \frac{\text{sum of } fx}{\text{total frequency}} = \frac{52}{26} = 2$$

Mode = most frequently occurring value = 1

Median falls between 13^{th} and 14^{th} value, so median = 2

Frequency diagrams (Year 9 Mathematics)

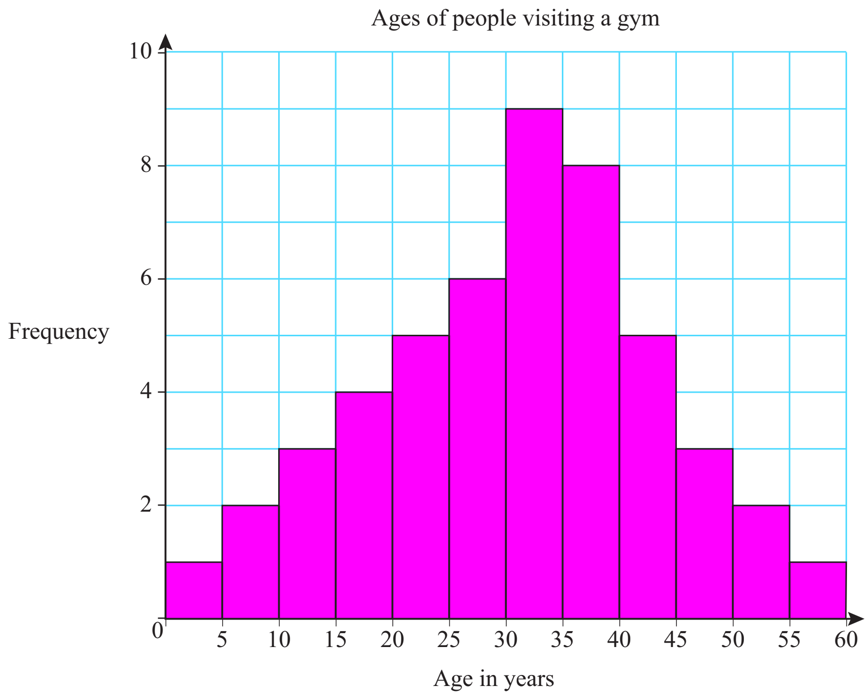
Data grouped in class intervals with no overlap can be shown on a histogram.



20.1 Histograms

A **histogram** is a specialised graph that looks a lot like a bar chart but is normally used to show the distribution of continuous or grouped data.

Look at this histogram showing the ages of people visiting a gym.



Notice that:

- The horizontal scale is **continuous** and each column is drawn above a particular **class interval**.
- The frequency of the data is shown by the *area* of the bars.
- There are no spaces between the bars on the graph because the horizontal scale is continuous. (If the **frequency** relating to a class interval is 0, you won't draw a bar in that class, so there will be a gap in the bars in that case.)

REWIND

Continuous data was introduced in chapter 4. ◀

REWIND

The grouping of data into classes was covered in chapters 4 and 12. ◀

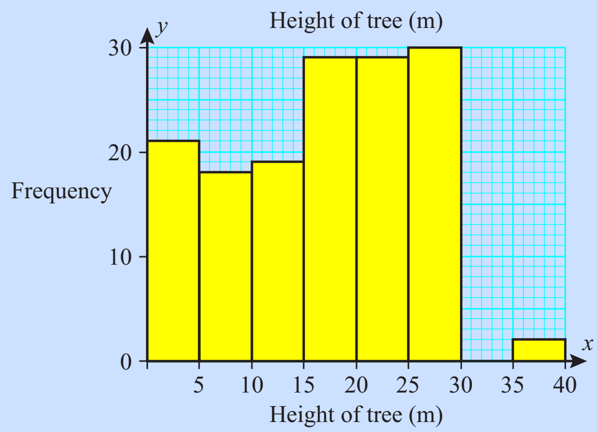
Histograms with equal class intervals

When the class intervals are equal, the bars are all the same width. Although it should be the area of the bar that tells you the frequency of the class, it is common practice when class intervals are all equal to just let the vertical scale show the frequency per class interval (and so it is just labelled 'frequency', as in the diagram above).

Worked example 1

The table and histogram below show the heights of trees in a sample from a forestry site.

Height (h) of trees in metres	Frequency
$0 \leqslant h < 5$	21
$5 \leqslant h < 10$	18
$10 \leqslant h < 15$	19
$15 \leqslant h < 20$	29
$20 \leqslant h < 25$	29
$25 \leqslant h < 30$	30
$30 \leqslant h < 35$	0
$35 \leqslant h < 40$	2



a How many trees are less than 5 m tall?
b What is the most common height of tree?
c How many trees are 20 m or taller?
d Why do you think the class intervals include inequality symbols?
e Why is there a gap between the columns on the right-hand side of the graph?

a	21	Read the frequency (vertical scale on the histogram) for the bar 0 – 5.
b	$25 \leqslant h < 30$ m	Find the tallest bar and read the class interval from the horizontal scale.
c	61	Find the frequency for each class with heights of 20 m or more and add them together.

d The horizontal scale of a histogram is continuous, so the class intervals are also continuous. The inequality symbols prevent the same height of tree falling into more than one group. For example, without the symbols a tree of height 5 m could go into two groups and thus be counted twice.

e The frequency for the class interval $30 \leqslant h < 35$ is zero, so no bar is drawn.

REWIND

You learned how to draw **grouped frequency tables** in chapter 4.

Worked example 2

Joy-Anne did an experiment in her class to see what mass of raisins (in grams) the students could hold in one hand. Here are her results.

18	18	20	22	22	22	22	23
23	24	24	25	25	25	25	25
25	26	26	27	30	30	31	35

a Using the class intervals 16–20, 21–25, 26–30 and 31–35 draw a grouped frequency table.
b What is the modal class (the mode) of this data?
c Draw a histogram to show her results.

a

Mass of raisins	Frequency
16–20	3
21–25	14
26–30	5
31–35	2

Count the number in each class to fill in the table.

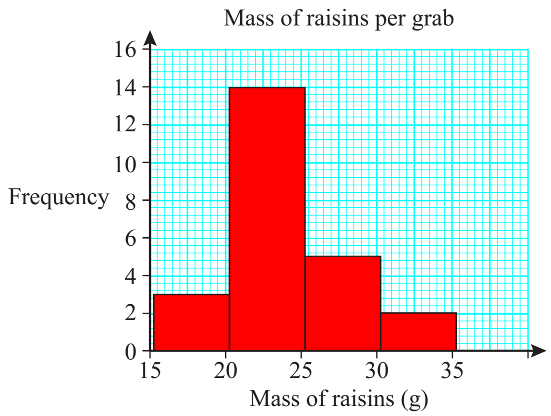
b The modal class is 21–25.

It is actually not possible to find the mode of grouped data because you do not have the individual values within each group. Instead, you find the class interval that has the greatest frequency. This is called the '**modal class**' (Extended students learned this in chapter 12).

REWIND

You saw in chapter 12 that mode is the most frequent result.

c



Although the data is in discrete groups, the raw data is actually continuous (it is mass). When Joy-Anne grouped the data, she rounded each mass to the nearest gram. This means that some raisins will have an *actual* mass that is between two of the discrete groups. To take this into account, each bar is plotted according to its upper and lower bound. So, the group 16 – 20 is drawn from $15.5 \leqslant h < 20.5$ and so on, such that a handful of raisins with a mass of 20.56g, would be in the class interval 21 – 25 because the group's boundaries are 20.5 – 25.5g.

REWIND

You learned about upper and lower bounds in chapter 13. ◀

Exercise 20.1

Applying your skills

1 Maria is a midwife who recorded the mass of the babies she delivered in one month.

Mass (kg)	$0.5 \leqslant m < 1.5$	$1.5 \leqslant m < 2.5$	$2.5 \leqslant m < 3.5$	$3.5 \leqslant m < 4.5$	$4.5 \leqslant m < 5.5$
No. of babies	1	12	31	16	0

a What is the modal class?
b How many babies have a mass of 2.5kg or less?
c Draw a histogram to show this distribution.

2 Annike did a breakdown of the length of telephone calls (t) on her mobile phone account. These are her results.

Length of time (t) in minutes	**Frequency**
$0 \leqslant t < 2$	15
$2 \leqslant t < 4$	43
$4 \leqslant t < 6$	12
$6 \leqslant t < 8$	19
$8 \leqslant t < 10$	15
$10 \leqslant t < 12$	10
$12 \leqslant t < 14$	11
$14 \leqslant t < 16$	17

a How many calls did she make altogether?
b What is the most common length of a call?
c Draw a histogram to show this distribution.
d Make a new frequency table of these results using the class intervals given in the following table.

Class interval	$0 \leqslant t < 4$	$4 \leqslant t < 8$	$8 \leqslant t < 12$	$12 \leqslant t < 16$
Frequency				

e Draw a histogram to show the new distribution.
f Write a few sentences comparing the distribution shown on the two histograms.

3 Shamiela cut 30 pieces of ribbon, which she estimated were each about 30 cm long. Her sister measured them and got the following actual lengths in centimetres:

29.1	30.2	30.5	31.1	32.0	31.3	29.8	29.5	31.6	32.4
32.1	30.2	31.7	31.9	32.1	29.9	32.1	31.4	28.9	29.8
31.2	31.2	30.5	29.7	30.3	30.4	30.1	31.1	28.8	29.5

a Draw a suitable frequency distribution table for this data. Use an equal class interval.
b Construct a histogram to show your distribution.
c How accurately did Shamiela estimate? Give a reason for your answer.

Be careful of discrete groups of continuous data; the raw data is continuous so can take any value between the groups.

4 The French Traffic Police recorded the number of vehicles speeding on a stretch of highway on a Friday night. Draw a histogram to show this data.

Speed over the limit (km/h)	1–10	11–20	21–30	31–40	41–50
Frequency	47	21	32	7	4

In other books, you might see histograms being used for grouped *discrete* data. Question 5 is a common example. In these cases, draw the histogram by extending the boundaries of each class interval to make them continuous, e.g. change 95 – 99 and 100 – 104 to $94.5 \leqslant m < 99.5$, $99.5 \leqslant m < 104.5$ etc. To draw a bar chart from this data, you would treat each group as a 'category' and draw the bar chart with gaps as normal.

5 Here are the IQ-test scores of a group of students.

IQ	Frequency
95–99	3
100–104	8
105–109	21
110–114	24
115–119	6
120–124	3
125–129	5
130–134	2
135–140	1

Draw a histogram to show this distribution.

Histograms with unequal class intervals

E

When the class intervals are not the same, using the height to give the frequency can be misleading. A class that is twice the width of another but with the same frequency covers twice the area. So, if the height is used to represent the frequency, the initial impression it gives is that it contains more values, which is not necessarily the case (see worked example 3). To overcome this, when the class intervals are unequal a new vertical scale is used called the **frequency density**.

$$\text{frequency density } (fd) = \frac{\text{frequency } (f)}{\text{class width } (cw)}$$

Frequency density takes into account the frequency relative to the size of the class interval, making it more fair when comparing different sized intervals.

Tip

Notice that $(f) = fd \times cw$ = area of a bar. You can use this to help you read frequencies from the histogram. Many questions are based on this principle.

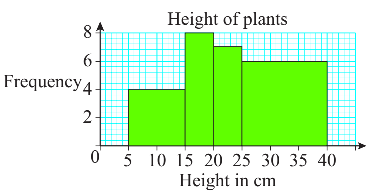
Exercise 20.2

1 140 people at a school fund-raising event were asked to guess how many sweets were

E

The heights in cm are the class intervals. The number of plants is the frequency.

If the data was plotted against frequency instead of frequency density (see below), it looks as though there are more plants in the class 25 – 40 compared to the class 5 – 10 but actually, their frequency densities are the same (see histogram in Worked example 3). The larger size of interval is misleading here, so we use frequency density as it is a fairer way to compare frequencies in classes of different sizes.



Worked example 3

Here is a table showing the heights of 25 plants. Draw a histogram to show these results.

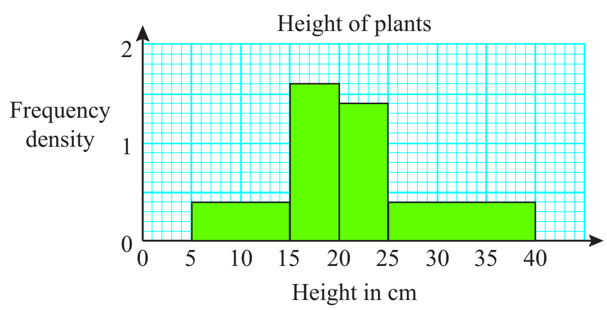
Height in cm	Number of plants
$5 \leqslant h < 15$	4
$15 \leqslant h < 20$	8
$20 \leqslant h < 25$	7
$25 \leqslant h < 40$	6

First work out the frequency density by adding columns to your frequency distribution table like this:

Height (h) in cm	Number of plants (f)	Class width (cw)	Frequency density ($= f \div cw$)
$5 \leqslant h < 15$	4	10	$\frac{4}{10} = 0.4$
$15 \leqslant h < 20$	8	5	$\frac{8}{5} = 1.6$
$20 \leqslant h < 25$	7	5	$\frac{7}{5} = 1.4$
$25 \leqslant h < 40$	6	15	$\frac{6}{15} = 0.4$

Next draw the axes. You will need to decide on a suitable scale for both the horizontal and the vertical axes. Here, 1 cm has been used to represent 10 cm on the horizontal axis (label height in cm) and 2 cm per unit on the vertical axis (label frequency density).

Once you have done this, draw the histogram, paying careful attention to the scales on the axes.



E

in a large glass jar. Those who guessed correctly were put into a draw to win the sweets as a prize. The table shows the guesses.

No. of sweets (n)	Frequency (f)
$100 \leqslant n < 200$	18
$200 \leqslant n < 250$	18
$250 \leqslant n < 300$	32
$300 \leqslant n < 350$	31
$350 \leqslant n < 400$	21
$400 \leqslant n < 500$	20

a Use the table to calculate the frequency density for each class.
b Construct a histogram to display the results. Use a scale of 1 cm = 100 sweets on the horizontal axis and a scale of 1 cm = 0.2 units on the vertical axis.

2 The table shows the mass of young children visiting a clinic (to the nearest kg). Draw a histogram to illustrate the data.

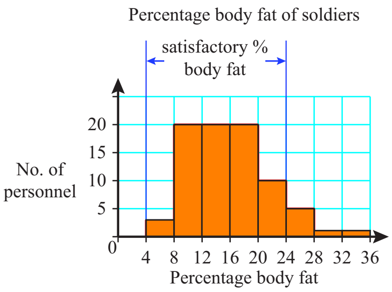
Mass in kilograms (m)	Frequency
$6 \leqslant m < 9$	9
$9 \leqslant m < 12$	12
$12 \leqslant m < 18$	30
$18 \leqslant m < 21$	15
$21 \leqslant m < 30$	18

3 The table shows the distribution of the masses of the actors in a theatre group. Draw a histogram to show the data.

Mass in kilograms (m)	Frequency
$60 \leqslant m < 63$	9
$63 \leqslant m < 64$	12
$64 \leqslant m < 65$	15
$65 \leqslant m < 66$	17
$66 \leqslant m < 68$	10
$68 \leqslant m < 72$	8

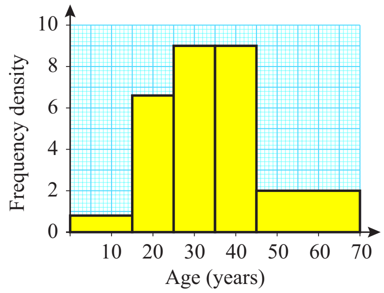
Applying your skills

4 A group of on-duty soldiers underwent fitness tests in which their percentage body fat was calculated. The fitness assessor drew up this histogram of the results.



a How many soldiers were tested?
b How many soldiers had body fat levels within the healthy limits?
c How many soldiers had levels which were too high?
d Why do you think there no bar in the 0–4 category?
e Would you expect a similar distribution if you tested a random selection of people in your community? Give reasons for your answers.

5 The histogram shows the ages of people using the fitness centre at the Sports Science



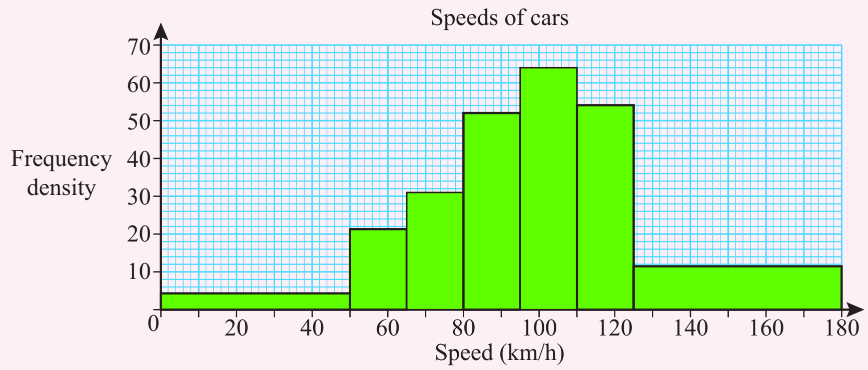
Institute after 5 pm in the evening.

a Copy and complete the frequency table for this data.

Age (a) in years	Frequency
$0 < a \leqslant 15$	
$15 < a \leqslant 25$	
$25 < a \leqslant 35$	
$35 < a \leqslant 40$	
$40 < a \leqslant 70$	

b How many people aged between 15 and 35 used the fitness centre after 5 pm?

6 A traffic officer used a computer program to draw this histogram showing the average speed (in km/h) of a sample of vehicles using a highway. The road has a minimum speed limit of 50 and a maximum speed limit of 125 km/h.



a Is it easy to see how many vehicles travelled above or below the speed limit? Give a reason for your answer.

b The traffic officer claims the graph shows that most people stick to the speed limit. Is he correct? Give a reason for your answer.

c His colleagues want to know exactly how many vehicles travel below or above the speed limit.

i Reconstruct this frequency table. Round frequencies to the nearest whole number.

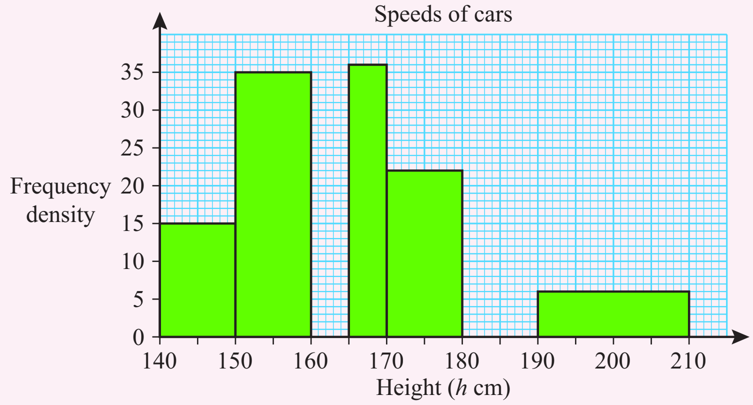
Speed in km/h (s)	Frequency	Class width	Frequency density
$0 \leqslant s < 50$			4.8
$50 \leqslant s < 65$			21.3
$65 \leqslant s < 80$			33.3
$80 \leqslant s < 95$			52
$95 \leqslant s < 110$			64
$110 \leqslant s < 125$			54.6
$125 \leqslant s < 180$			11.6

ii How many vehicles were below the minimum speed limit?

d What percentage of vehicles in this sample were exceeding the maximum speed limit?

E

7 The unfinished histogram and table give information about the heights, in centimetres, of the Senior students at a High School.



Height (h cm)	Frequency
$140 \leqslant h < 150$	15
$150 \leqslant h < 160$	
$160 \leqslant h < 165$	20
$165 \leqslant h < 170$	
$170 \leqslant h < 180$	
$180 \leqslant h < 190$	12
$190 \leqslant h < 210$	

a Use the histogram to complete the table
b Use the table to complete the histogram
c State the modal group
d Work out an estimate for the percentage of senior students at the High School above the height of 155 cm

LINK

Estimates of medians are very important in the study of psychology. When trying to understand how different conditions influence the response of the human brain it is often good to try to summarise data rather than use every number that you have. This makes the various trends that may appear much clearer.

20.2 Cumulative frequency

Sometimes you may be asked questions such as:

- How many people had a mass of less than 50 kilograms?
- How many cars were travelling above 100 km/h?
- How many students scored less than 50% on the test?

In statistics you can use a **cumulative frequency** table or a **cumulative frequency curve** to answer questions about data up to a particular class boundary. You can also use the cumulative frequencies to estimate and interpret the median and the value of other positions of a data set.

Cumulative means 'increasing as more is added'.

Cumulative frequency tables

Cumulative frequency is really just a 'running total' of the scores or results (the frequency in each group). The cumulative frequency gives the number of results which are less than, or equal to, a particular class boundary. This table shows how many students got a particular mark out of 10 (the frequency of each result) as well as the cumulative frequency.

E

Score out of 10	Frequency (f)	Cumulative frequency
3	4	4
4	5	$4 + 5 = 9$
5	3	$9 + 3 = 12$
6	3	$12 + 3 = 15$
7	5	20
8	7	27
9	2	29
10	1	30
Total	30	

- Each entry in the cumulative frequency column is calculated by adding the frequency of the current class to the previous cumulative frequency (or by adding all the frequencies up to and including the current class).
- The last figure in the cumulative frequency column must equal the sum of the frequencies because all results will be below or equal to the highest result.

Worked example 4

The heights of plants were measured during an experiment. The results are summarised in the table.

Height (h cm)	Frequency
$0 < h \leqslant 5$	20
$5 < h \leqslant 10$	40
$10 < h \leqslant 15$	60
$15 < h \leqslant 25$	80
$25 < h \leqslant 50$	50
Total	250

a Draw up a cumulative frequency table for this distribution.
b Determine which class interval contains the median height.

a

Height (h cm)	Frequency	Cumulative frequency
$0 < h \leqslant 5$	20	20
$5 < h \leqslant 10$	40	60
$10 < h \leqslant 15$	60	120
$15 < h \leqslant 25$	80	200
$25 < h \leqslant 50$	50	250
Total	250	

b $15 < h \leqslant 25$

The heights are given for 250 flowers, so the median height must be the mean of the height of the 125th and 126th flower. If you look at the cumulative frequency you can see that this value falls into the fourth height class (the 125th and 126th are both greater than 120 but less than 200).

REWIND

In chapter 12, median classes were introduced for grouped data. You will see that cumulative frequency curves will enable you to estimate the median when the number of data is large.

E

Cumulative frequency curves

When you plot the cumulative frequencies against the upper boundaries of each class interval you get a cumulative frequency curve.

Cumulative frequency curves are also called ogive curves or ogives because they take the shape of narrow pointed arches (called ogees) like these ones on a mosque in Dubai.

In mathematics, arches like these are seen as two symmetrical s-curves.

Tip
You must plot the cumulative frequency at the upper end point of the class interval. Do not confuse this section with the mid-point calculations you used to estimate the mean in frequency tables.

Worked example 5

The examination marks of 300 students are summarised in the table.

Mark	Frequency
1–10	3
11–20	7
21–30	13
31–40	29
41–50	44
51–60	65
61–70	70
71–80	49
81–90	14
91–100	6

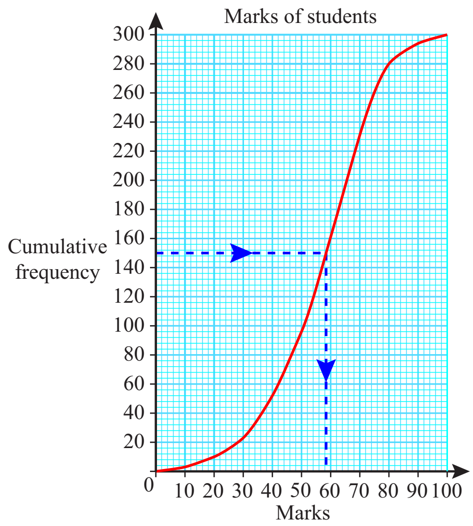
a Draw a cumulative frequency table.
b Construct a cumulative frequency graph to show this data.
c Calculate an *estimate* for the median mark.

E

a

Mark	Frequency	Cumulative frequency
1–10	3	3
11–20	7	10
21–30	13	23
31–40	29	52
41–50	44	96
51–60	65	161
61–70	70	231
71–80	49	280
81–90	14	294
91–100	6	300

b



c The median is the middle value. For continuous data, the middle value can be found by dividing the total frequency by 2.

$\frac{300}{2} = 150$, so the median mark is the 150th result. Draw a line from the 150th student (on the vertical axis) parallel to the marks (horizontal) axis. Drop a perpendicular from where this line cuts the graph. Read the value from the horizontal axis.

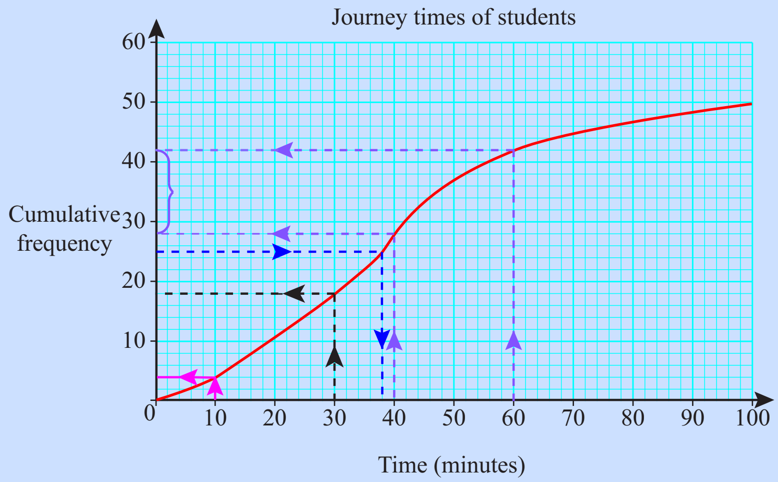
The median mark is 58.

REWIND

You learned in chapter 12 how to work out the median for discrete data. Note that the cumulative frequency graph allows you to find an *estimate* for the actual value rather than a class interval.

Worked example 6

This cumulative frequency curve shows the journey times to school of different students.



E

Use the curve to find:

a the total number of students
b an estimate of the median journey time
c the number of students who took less than 10 minutes to get to school
d the number of students who had journey times greater than 30 minutes
e the number of students who took between 40 minutes and one hour to get to school.

a	50	The top of the curve is at 50, so this is the total frequency.
b	38	$\frac{50}{2} = 25$ so its the 25th result; drop a perpendicular from where the line cuts the graph.
c	4	Read off the cumulative frequency at 10 minutes.
d	$50 - 18 = 32$	Subtract the cumulative frequency at 30 minutes, 18, from the total frequency.
e	$42 - 28 = 14$	Subtract the cumulative frequency at 40 minutes, 28, from that at 60 minutes, 42.

Worked example 7

Twenty bean seeds were planted for a biology experiment. The heights of the plants were measured after three weeks and recorded as below.

Heights (h cm)	$0 \leqslant h < 3$	$3 \leqslant h < 6$	$6 \leqslant h < 9$	$9 \leqslant h < 12$
Frequency	2	5	10	3

a Find an estimate for the mean height.
b Draw a cumulative frequency curve and find an estimate for the median height.

a You will need the mid-points of the classes to help you find an estimate of the mean, and the cumulative frequency to help find an estimate of the median, so more columns need to be added to the table. Don't forget to label the new columns.

Heights (h cm)	**Mid-point (x)**	**Frequency (f)**	**Frequency × mid-point (fx)**	**Cumulative frequency**
$0 \leqslant h < 3$	1.5	2	3	2
$3 \leqslant h < 6$	4.5	5	22.5	7
$6 \leqslant h < 9$	7.5	10	75	17
$9 \leqslant h < 12$	10.5	3	31.5	20
Total		20	132	

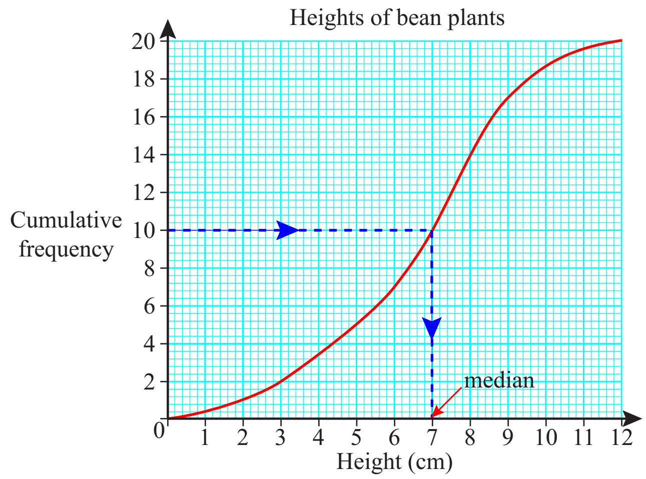
Mean height $= \frac{132}{20} = 6.6\,\text{cm}$ $\qquad \left(\text{mean} = \frac{\text{total } fx}{\text{total } f}\right)$

REWIND

You learned how to find an estimate for the mean of grouped data in chapter 12. Revise this now if you have forgotten it.

E

b



$\frac{20}{2} = 10$, so the median height is the 10th value.
Median height = 7.0 cm

Exercise 20.3

1 The heights of 25 plants were measured to the nearest centimetre.
The results are summarised in the table.

Height in cm	6–15	16–20	21–25	26–40
Number of plants	3	7	10	5

a Draw a cumulative frequency table for this distribution.
b In which interval does the median plant height lie?
c Draw the cumulative frequency curve and use it to estimate, to the nearest centimetre, the median plant height.

2 The table shows the amount of money, $\$x$, spent on books by a group of students.

Amount spent	**No. of students**
$0 < x \leqslant 10$	0
$10 < x \leqslant 20$	4
$20 < x \leqslant 30$	8
$30 < x \leqslant 40$	12
$40 < x \leqslant 50$	11
$50 < x \leqslant 60$	5

a Calculate an estimate of the mean amount of money per student spent on books.
b Use the information in the table above to find the values of p, q and r in the following cumulative frequency table.

Amount spent	$\leqslant 10$	$\leqslant 20$	$\leqslant 30$	$\leqslant 40$	$\leqslant 50$	$\leqslant 60$
Cumulative frequency	0	4	p	q	r	40

c Using a scale of 1 cm to represent 10 units on each axis, draw a cumulative frequency diagram.
d Use your diagram to estimate the median amount spent.

E

3 This cumulative frequency table shows the distribution of the masses of the children attending a clinic.

Mass in kilograms (M)	Cumulative frequency
$0 < m \leqslant 10.0$	12
$0 < m \leqslant 20.0$	26
$0 < m \leqslant 30.0$	33
$0 < m \leqslant 40.0$	41
$0 < m \leqslant 50.0$	46
$0 < m \leqslant 60.0$	50

a Draw a cumulative frequency diagram. Use a horizontal scale of 1 cm = 10 kg and a vertical scale of 0.5 cm = 5 children.
b Estimate the median mass.
c How many children had a mass higher than the median mass?

Quartiles

REWIND

Revise the work you did on quartiles and the interquartile range in Chapter 12 if you need to. ◀

In chapter 12 you found the range (the biggest value – the smallest value) to see how dispersed various sets of data were. The range, however, is easily affected by outliers (extreme or unusual values), so it is not always the best measure of how the data is spread out.

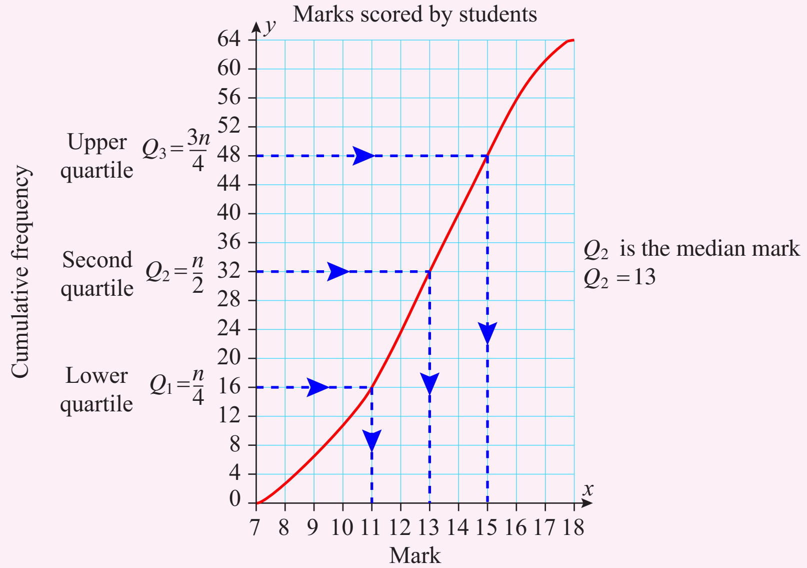
The data shown on a cumulative frequency curve can be divided into four equal groups called **quartiles** to find a measure of spread called the **interquartile range**, which is more representative than the range because it is not affected by extremes.

The cumulative frequency curve on the next page shows the marks obtained by 64 students in a test. These are listed below:

- 48 students scored less than 15 marks. 15 marks is the upper quartile or third quartile Q_3.
- 32 students scored less than 13 marks. 13 marks is the second quartile Q_2, or median mark.
- 16 students scored less than 11 marks. 11 marks is the lower quartile or first quartile Q_1.

Whole number values are being used in this example to make it easier to understand. Usually your answers will be estimates and they will involve decimal fractions.

When finding the positions of the quartiles from a cumulative frequency curve you do not use the $\frac{(n+1)}{4}$, $\frac{(n+1)}{2}$ and $\frac{3}{4}(n+1)$ rules that you met for discrete data in chapter 12. Instead you use: $\frac{n}{4}$, $\frac{n}{2}$ and $\frac{3n}{4}$.



E

The interquartile range

The interquartile range (IQR) is the difference between the upper and lower quartiles: $Q_3 - Q_1$.

In effect, this is the range of the middle 50% of the scores, or the median of the upper half of the values minus the median of the lower half of the values.

In the example above, the IQR = 15 − 11 = 4

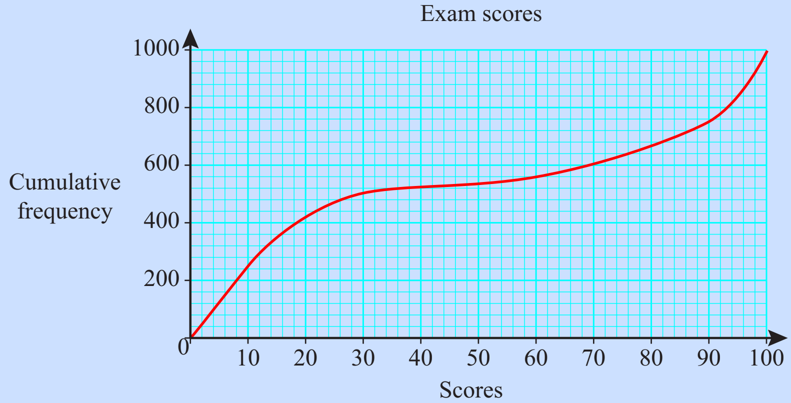
Because the interquartile range does not use any extreme small or large values it is considered a more reliable measure of spread than the range.

REWIND

The range was used to compare data sets in chapter 12.

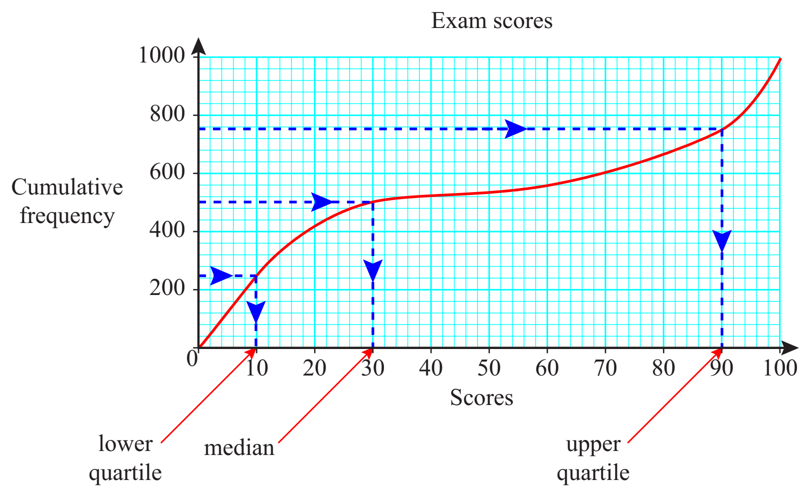
Worked example 8

The percentage scored by 1000 students on an exam is shown on this cumulative frequency curve.



Use the cumulative frequency curve to find an estimate for:

a the median score
b the lower quartile
c the upper quartile
d the interquartile range.



E

a $n = 1000$

So the position of Q_2 on the vertical axis is $\frac{n}{2} = \frac{1000}{2} = 500$

Draw the lines on the graph.

Estimate the median from the horizontal axis as 30 marks.

b $n = 1000$

So the position of Q_1 on the vertical axis is $\frac{n}{4} = \frac{1000}{4} = 250$

An estimate for the lower quartile, from the horizontal axis is 10 marks

c $n = 1000$

So the position of Q_3 on the vertical axis is $\frac{3n}{4} = \frac{3 \times 1000}{4} = 750$

An estimate for the upper quartile, from the horizontal axis is 90 marks.

d $\text{IQR} = Q_3 - Q_1$

$= 90 - 10$

$= 80$ marks

REWIND

Percentiles were briefly introduced in chapter 12.

Percentiles

When you are dealing with large amounts of data, such as examination results for the whole country, or the average height and mass of all children in different age groups, it is useful to divide it into even smaller groups called **percentiles**.

Percentiles divide the data into 100 equal parts.

To find the position of a percentile use the formula $\frac{pn}{100}$, where p is the percentile you are looking for and n is how much data you have (the total frequency).

Using the data set in worked example 8:

The position of the 10th percentile on the cumulative frequency axis is $P_{10} = \frac{10 \times 1000}{100} = 100$

The position of the 85th percentile on the cumulative frequency axis is $P_{85} = \frac{85 \times 1000}{100} = 850$

(Don't forget that you need to move right to the curve and down to the horizontal axis to find the values of the percentiles.)

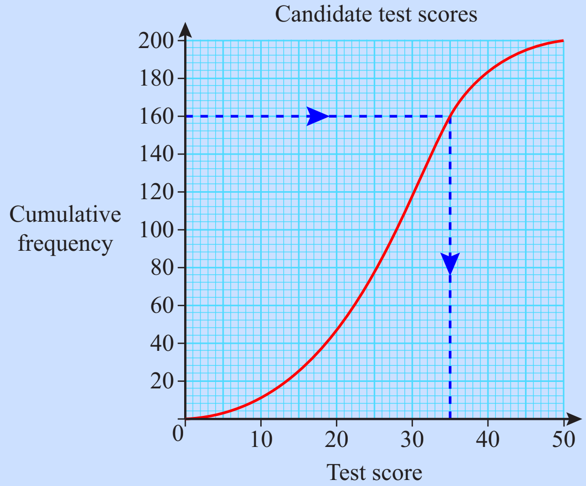
The percentile range is the difference between given percentiles. In the example above, this is $P_{85} - P_{10}$.

In chapter 12, percentiles were first introduced but only the 25th and 75th percentiles were used to introduce the interquartile range. A question was posed at the start of section 12.5 on page 244: 'All those candidates above the 80th percentile will be offered an interview. What does this mean?' The following worked example shows you how to answer this question.

E

Worked example 9

The cumulative frequency curve shows the test results of 200 candidates who have applied for a post at Fashkiddler's. Only those who score above the 80th percentile will be called for an interview. What is the lowest score that can be obtained to receive an interview letter?

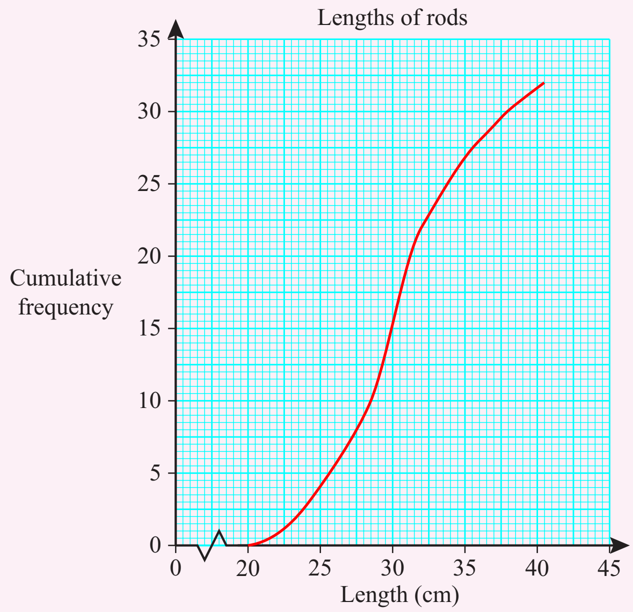


80% of 200 is 160.
So, the value of P_{80} is a test score of 35. (Read off the graph where the curve is 160)
Only those candidates who scored above 35 marks on the test will be called for an interview.

Exercise 20.4

1 The lengths of 32 metal rods were measured and recorded on this cumulative frequency curve. Use the graph to find an estimate for:

- **a** the median
- **b** Q_1
- **c** Q_3
- **d** the IQR
- **e** the 40th percentile.



E

2 This cumulative frequency curve compares the results 120 students obtained on two maths papers.

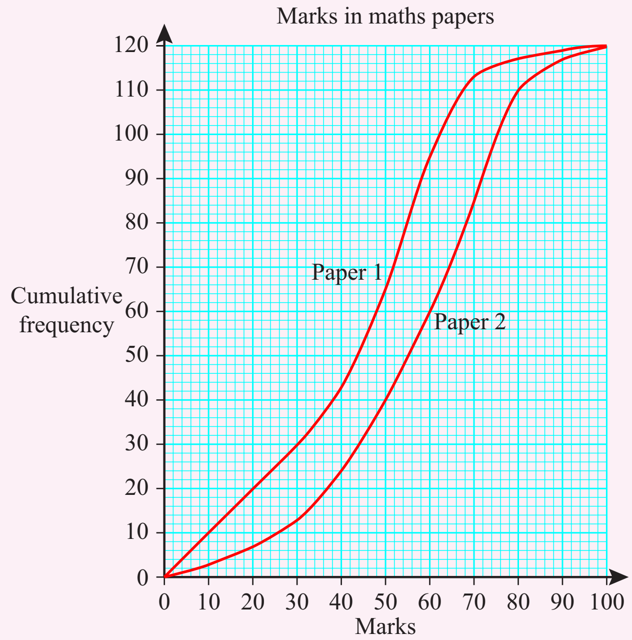
a For each paper, use the graph to find:

i the median mark

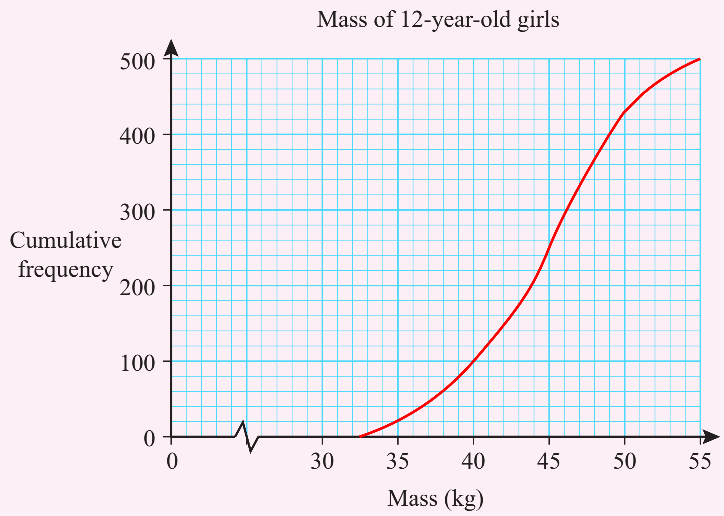
ii the IQR

iii the 60th percentile.

b What mark would you need to get to be above the 90th percentile on each paper?



3 This cumulative frequency curve shows the masses of 500 12-year-old girls (in kg).



a Use the graph to work out:

i the median mass of the 12-year-olds

ii how many girls have a mass between 40 and 50 kg.

b What percentage of girls are unable to go on an amusement park children's ride if the upper mass limit for the ride is 51 kg?

4 This cumulative frequency table gives the speeds of 200 cars travelling on the highway from Kuala Lumpur Airport into the city.

Speed in km/h (s)	Cumulative frequency
$s < 60$	2
$60 \leqslant s < 70$	8
$70 \leqslant s < 80$	24
$80 \leqslant s < 90$	45
$90 \leqslant s < 100$	96
$100 \leqslant s < 110$	123
$110 \leqslant s < 120$	171
$120 \leqslant s < 130$	195
$130 \leqslant s < 140$	200
Total	200

a Draw a cumulative frequency curve to show this data. Use a scale of 1 cm per 10 km/h on the horizontal axis and a scale of 1 cm per 10 cars on the vertical axis.
b Use your curve to estimate the median, Q_1 and Q_3 for this data.
c Estimate the IQR.
d The speed limit on this stretch of road is 120 km/h. What percentage of the cars were speeding?

Summary

Do you know the following?

- Histograms are specialised bar graphs used for displaying continuous and grouped data.
- There is no space between the bars of a histogram because the horizontal scale is continuous.
- When the class widths are equal the bars are equally wide and the vertical axis shows the frequency.
- If the class widths are unequal, the bars are not equally wide and the vertical axis shows the frequency density.
- Frequency density $= \dfrac{\text{frequency per class interval}}{\text{class width}}$
- Cumulative frequency is a running total of the class frequencies up to each upper class boundary.
- When cumulative frequencies are plotted they give a cumulative frequency curve or ogive.
- The curve can be used to estimate the median value in the data.
- The data can be divided into four equal groups called quartiles. The interquartile range is the difference between the upper and lower quartiles ($Q_3 - Q_1$).
- Large masses of data can be divided into percentiles which divide the data into 100 equal groups. They are used to compare and rank measurements.

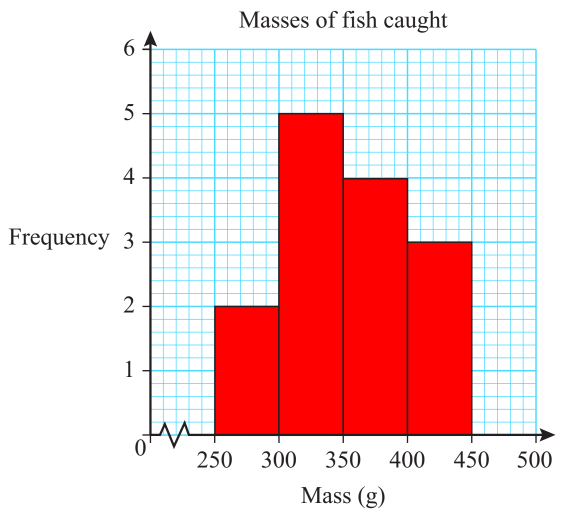
Are you able to …?

- read and interpret histograms with equal intervals
- construct histograms with equal intervals
- interpret and construct histograms with unequal intervals
- construct a table to find the frequency density of different classes
- calculate cumulative frequencies
- plot and draw a cumulative frequency curve
- use a cumulative frequency curve to estimate the median
- find quartiles and calculate the interquartile range
- estimate and interpret percentiles.

Examination practice

Exam-style questions

1 After a morning's fishing Imtiaz measured the mass, in grams, of the fish he had caught. The partially completed histogram represents his results.



a Imtiaz also caught four other fish. Their masses were 225 g, 466 g, 470 g and 498 g. Add this data to a copy of the graph to complete it.

b Use the completed graph to complete this table.

Mass (m) in grams	Number of fish	Classification
$m < 300$		Small
$300 \leqslant m < 400$		Medium
$m \geqslant 400$		Large

c Represent the information in the table as a pie chart. Show clearly how you calculate the angle of each sector.

E

2 A researcher took a questionnaire to 64 households. The grouped frequency table shows the time taken (t minutes) by various home owners to complete a questionnaire.

Time taken (t) in minutes	No. of home owners
$0 \leqslant t < 2$	2
$2 \leqslant t < 3$	18
$3 \leqslant t < 4$	25
$4 \leqslant t < 6$	12
$6 \leqslant t < 9$	5
$9 \leqslant t < 15$	2

Using a scale of 1 cm to represent 2 minutes, construct a horizontal axis for $0 \leqslant t < 15$.
Using a vertical scale of 1 cm per 2 units, draw a histogram to represent this data.

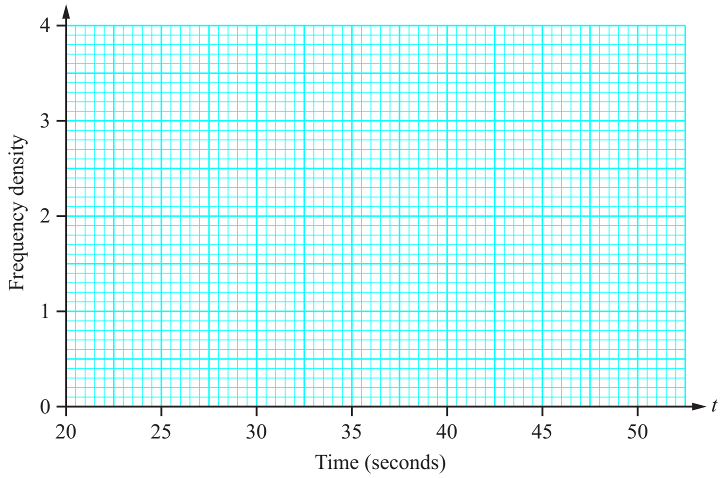
Past paper questions

1 A new frequency table is made from the results shown in the table below.

Time (t seconds)	$20 < t \leqslant 35$	$35 < t \leqslant 40$	$40 < t \leqslant 50$
Frequency			

a Complete the table. [1]

b On the grid, draw a histogram to show the information in this new table.



[3]

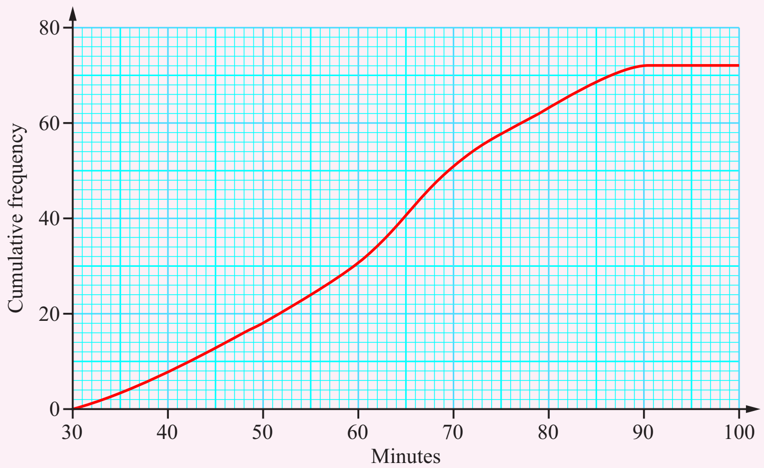
Cambridge IGCSE Mathematics 0580 Paper 42, Q3c, November 2014

E

2 72 students are given homework one evening.

They are told to spend no more than 100 minutes completing their homework.

The cumulative frequency diagram shows the number of minutes they spend.



a How many students spent more than 48 minutes completing their homework? [2]

b Find

i the median, [1]

ii the inter-quartile range. [2]

Cambridge IGCSE Mathematics 0580 Paper 22, Q18, November 2014