16 Cumulative Frequency

16.1 Averages

In this section we revisit the three types of average, the *mean*, *median* and *mode*. We also use the *range* of a set of data.

Mean	=	$\dfrac{\text{sum of the values}}{\text{number of values}}$
Median	=	middle value (when the data is arranged in order); where there are two central values, the median is their mean
Mode	=	most common value
Range	=	difference between largest and smallest values

Example 1

1 7 8 2 3 6 5 10 3

For this sample,

(a) calculate the *mean*, (b) determine the *median*,

(c) state the *mode*, (d) calculate the *range*.

Solution

(a) $\text{Mean} = \dfrac{1+7+8+2+3+6+5+10+3}{9}$

$= \dfrac{45}{9}$

$= 5$

(b) To find the median, first write the numbers in order.

1 2 3 3 (5) 6 7 8 10

Median

As the number of data items is odd, the median will be the middle number, which is 5 in this case, so

Median = 5

(c) The mode is the most common value, which is 3 for this set of values.

(d) Range = 10 – 1

= 9

Example 2

Determine the *median* of the following set of values:

44 32 88 19 33 74 62 31 33 56

Solution

First write the numbers in order:

19 31 32 33 33 44 56 62 74 88

In this case, there are 2 middle numbers, 33 and 44. The median will be the mean of these.

$$\text{Median} = \frac{33+44}{2}$$

$$= \frac{77}{2}$$

$$= 38.5$$

Example 3

A class collected data on the number of people living in their home, which is shown in the following table:

Number of People Living in Home	*Frequency*
2	3
3	9
4	10
5	2
6	3
7	1
8	1
9	0
10	1

(a) Calculate the *mean* number of people living in each home.

(b) Determine the *median* of the data.

(c) State the *mode* of the data.

Solution

(a) The first step is to complete the table below:

Number of People Living in Home	*Frequency*	*Number of People* × *Frequency*
2	3	6
3	9	27
4	10	40
5	2	10
6	3	18
7	1	7
8	1	8
9	0	0
10	1	10
TOTALS	30	126

$$\text{Mean} = \frac{126}{30}$$
$$= 4.2 \text{ people per home}$$

(b) As there are 30 values, the median is the mean of the 15th and 16th values. From the first table we can see that both the 15th and 16th values are 4, so the median is 4 people per home.

(c) The most common value is 4 so the mode is 4 people per home.

Exercises

1. Calculate the *mean* and the *range* of each of the following sets of data:

   (a) 3 17 5 6 12

   (b) 30 42 19 21 33 62

   (c) 7 8 3 14 31 3 8 9 13 22

   (d) 114 115 110 119 114 118 123 133

2. Determine the *median* and the *mode* of each of the following sets of data:

   (a) 8 5 19 32 19

   (b) 33 14 16 19 22 33 16 33 22

   (c) 5 9 19 3 14 21 5 7

   (d) 11 21 19 11 13 16 11 19 22 20

3. In which of the following data sets is the *mean* the same as the *median*:

A 34 6 19 17 9

B 29 12 17 18 44 13 17 40

C 101 107 183 51 57 77 100 92

D 27 92 56 83 45

4. Which of the following data sets has the *largest* range:

A 14 27 88 73 56 61

B 374 521 628 314 729

C 888 912 897 907 887 893

5. The following table gives the results of a survey question asking people how many television sets they had in their home.

Number of Televisions	*Frequency*
0	3
1	18
2	64
3	73
4	22
5	14
6	6

For this data,

(a) calculate the *mean,*

(b) determine the *median,*

(c) state the *mode.*

6. A car park manager recorded the number of cars entering her car park each hour. The data she collected is listed below.

16	22	17	6	5	8	32	15	9	7	14	33
21	11	6	5	11	14	12	22	19	11	3	14
14	7	23	41	32	16	5	19	14	33	7	12

For this data:

(a) calculate the *mean,*

(b) determine the *median,*

(c) determine the *mode,*

(d) calculate the *range.*

Which of the 3 averages should the manager use to convince her employers that the car park is going to make a large profit?

7. John looks at the price of a computer game in 8 different shops. The prices he sees are:

£29.99	£25.00	£34.99	£29.00
£24.99	£29.99	£31.00	£29.95

(a) Calculate the *mean* of this data.

(b) State the *mode* of this data.

(c) Determine the *median.*

Which of these averages should he use to argue that the computer game is too expensive?

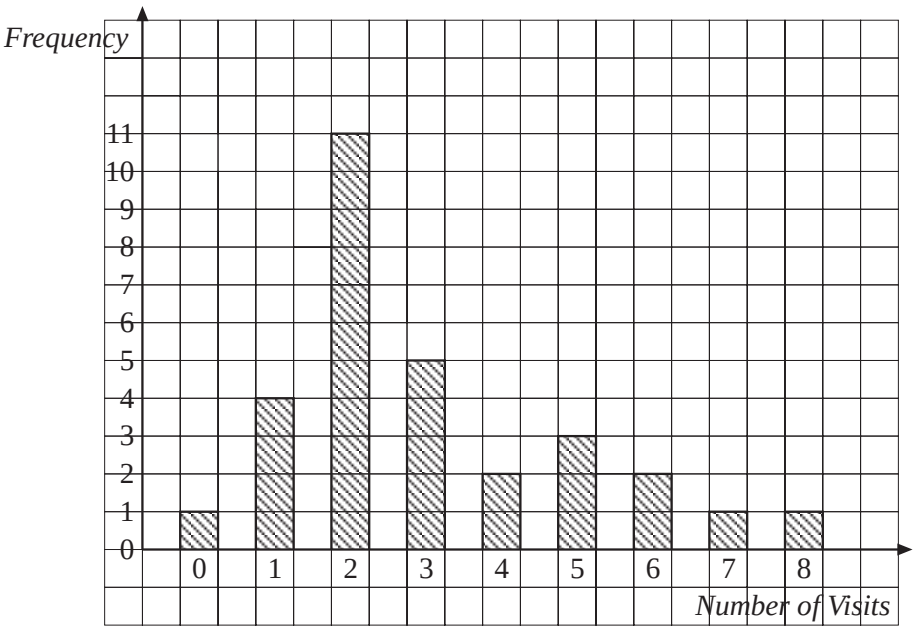
8. For the set of data given below, calculate the *mean* and determine the *median.*

4 7 3 9 5 6 142 3 7 11

Describe the advantages of using the median, rather than the mean in this case.

9. A student collected data on the number of visits to the dentist made by members of his class in one school year.

His results are shown in the following bar chart:



For the data:

(a) state the *mode,*

(b) calculate the *mean,*

(c) determine the *median.*

10. A set of three numbers has *mean* 11, *median* 12 and *range* 13. What are the 3 numbers?

16.2 Grouped Data

When dealing with grouped data it is important to think about the type of data that is being processed. You also have to decide the range of values that each group contains.

When calculating the mean of grouped data, we assume that all the values lie at the midpoint of the group.

These ideas are illustrated in the following examples.

Example 1

The table below shows the times taken by a group of walkers to complete a 15-mile walk. Their times have been recorded to the nearest hour.

Illustrate the data using a bar chart and a frequency polygon.

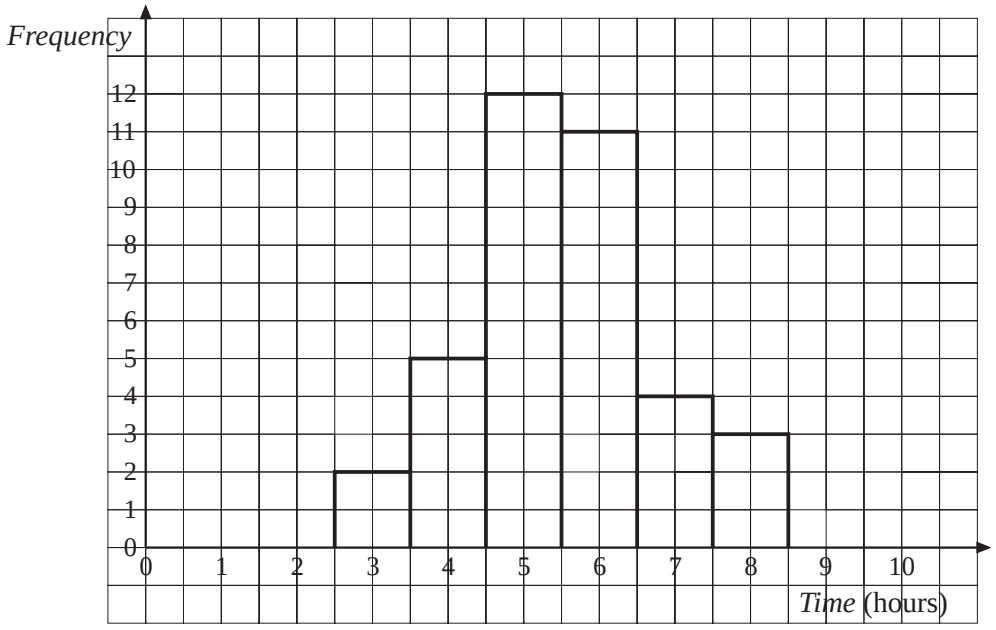
Time (hours)	3	4	5	6	7	8
Frequency	2	5	12	11	4	3

Solution

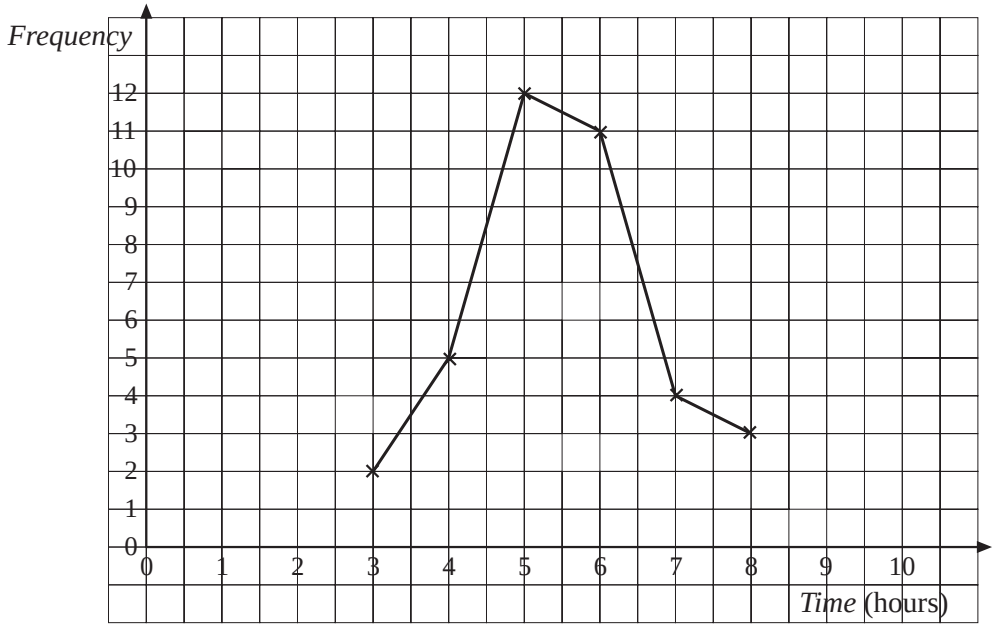
A time of 5 hours actually means a time that is greater than or equal to $4\frac{1}{2}$ hours but is less than $5\frac{1}{2}$ hours, so the bar representing this time on the bar chart will begin at 4.5 and end at 5.5.

Similarly, the bar for a time of 3 will begin at 2.5 and end at 3.5.

The bar chart is shown below:



The frequency polygon is shown below. We obtain it by joining the midpoints of the tops of the bars from the previous graph.



Example 2

At a school fair, visitors enter a 'Guess the weight of the cake' competition. Their guesses, rounded to the nearest 100 grams, were recorded in the following table:

Guess (kg)	0.5 - 0.7	0.8 - 1.0	1.1 - 1.3	1.4 - 1.6	1.7 - 1.9
Frequency	5	32	26	11	6

(a) Illustrate the data using a bar chart.

(b) Estimate the mean of the data.

(c) State the modal class.

Solution

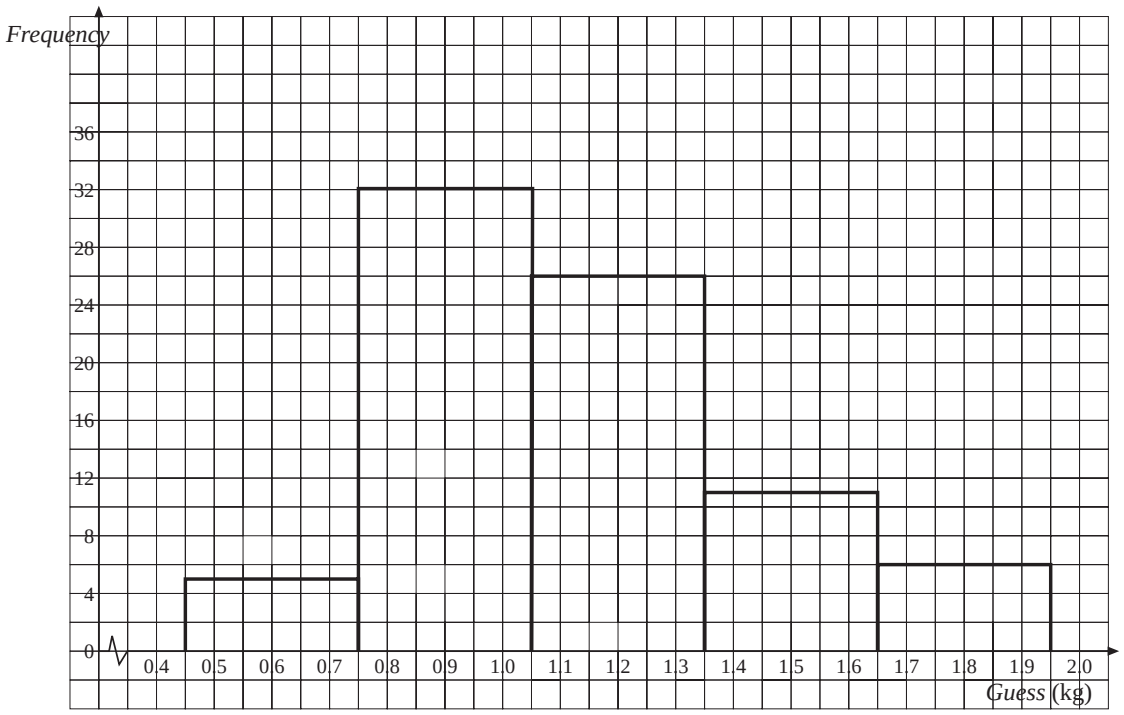
(a) The guesses have been recorded to one decimal place, in other words to the nearest 100 grams. This means that the first category, nominally described as '0.5 - 0.7 kg' actually includes guesses greater than or equal to 0.45 kg but less than 0.75 kg. The precise description of the first category is therefore

$$0.45 \text{ kg} \leq \text{guess} < 0.75 \text{ kg}$$

The nominal descriptions of the other classes must also be interpreted precisely if we are to represent the data accurately.

Guess (kg)	$0.45 \leq G < 0.75$	$0.75 \leq G < 1.05$	$1.05 \leq G < 1.35$	$1.35 \leq G < 1.65$	$1.65 \leq G < 1.95$
Frequency	5	32	26	11	6

The precise descriptions of the classes indicate how the bars should be drawn on the bar chart.



(b) The mean can be estimated by assuming that all the values in a class are equal to the midpoint of the class.

Class	*Midpoint*	*Frequency*	*Frequency* × *Midpoint*
$0.45 \le G < 0.75$	0.6	5	$5 \times 0.6 = 3$
$0.75 \le G < 1.05$	0.9	32	$32 \times 0.9 = 28.8$
$1.05 \le G < 1.35$	1.2	26	$26 \times 1.2 = 31.2$
$1.35 \le G < 1.65$	1.5	11	$11 \times 1.5 = 16.5$
$1.65 \le G < 1.95$	1.8	6	$6 \times 1.8 = 10.8$
TOTALS		80	90.3

$$\text{Estimate of mean} = \frac{90.3}{80} = 1.12875 \text{ kg}$$

$$= 1.1 \text{ kg} \quad \text{to 2 significant figures}$$

(c) The modal class is the one with the highest frequency. In this case, the modal class has nominal description '0.8 - 1.0 kg', which means guesses in the interval $0.75 \text{ kg} \le G < 1.05 \text{ kg}$, i.e. $750 \text{ grams} \le G < 1050 \text{ grams}$.

Exercises

1. The following table lists the results of a survey that recorded the heights of pupils in one year group. The heights have been given to the nearest 10 cm.

Height (cm)	140	150	160	170	180	190
Frequency	3	5	57	63	30	2

(a) Illustrate the data on a bar chart.

(b) Estimate the mean height of the pupils.

2. The following table lists the masses of a group of students, recorded to the nearest kg:

Mass (kg)	60	61	62	63	64	65	66	67	68	69	70
Frequency	3	7	9	11	10	22	17	23	11	9	5

(a) Illustrate the data using a frequency polygon.

(b) Estimate the mean mass for these students.

3. An English class looked at the number of words per sentence for an essay that one of them had written. Their results are summarised in the following table:

Number of Words	6 - 8	9 - 11	12 - 14	15 - 17	18 - 20
Frequency	13	10	8	4	3

(a) Estimate the mean number of words per sentence.

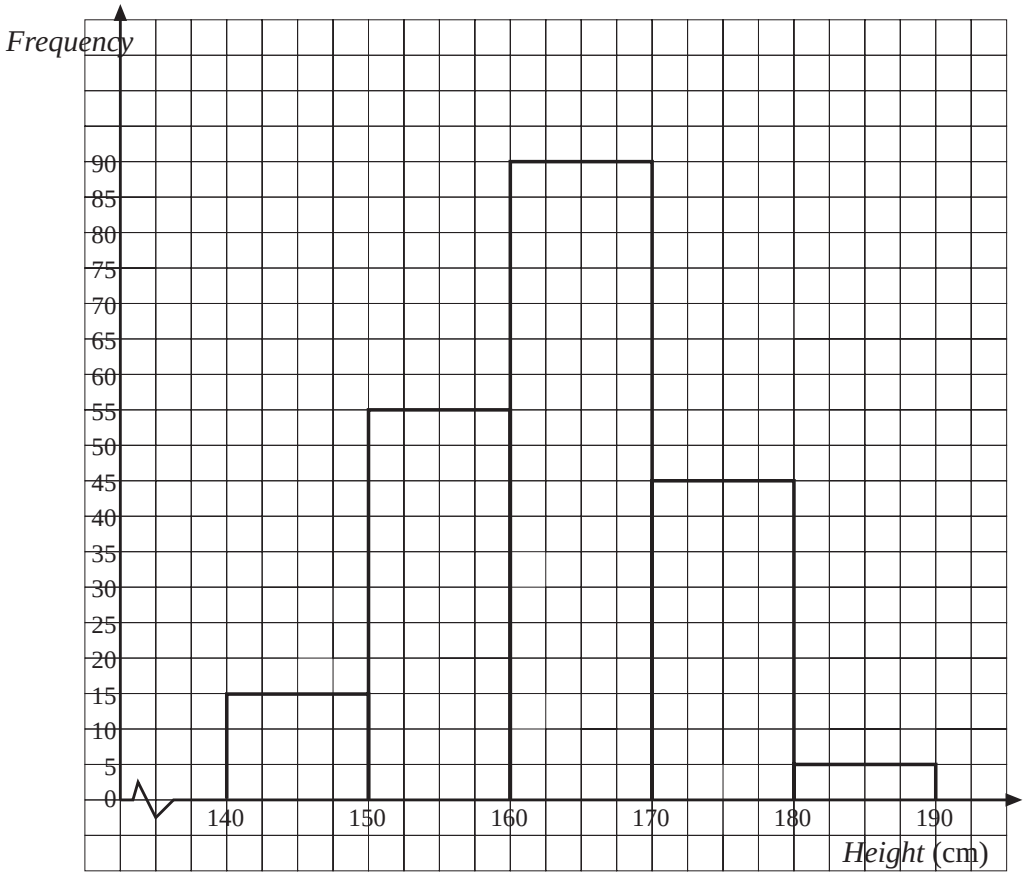
(b) What is the modal class?

4. The time taken for people to solve a puzzle is recorded, to the nearest minute, in the following table:

Time (mins)	2 - 5	6 - 9	10 - 13	14 - 17	18 - 21
Frequency	3	19	20	12	6

Estimate the mean time taken to solve the puzzle.

5. The bar chart shows the results of a survey into the height of 14-year-old pupils.



(a) State the modal class.

(b) Calculate an estimate of the mean height.

6. The heights of some plants grown in a laboratory were recorded after 4 weeks. The results are listed in the following table:

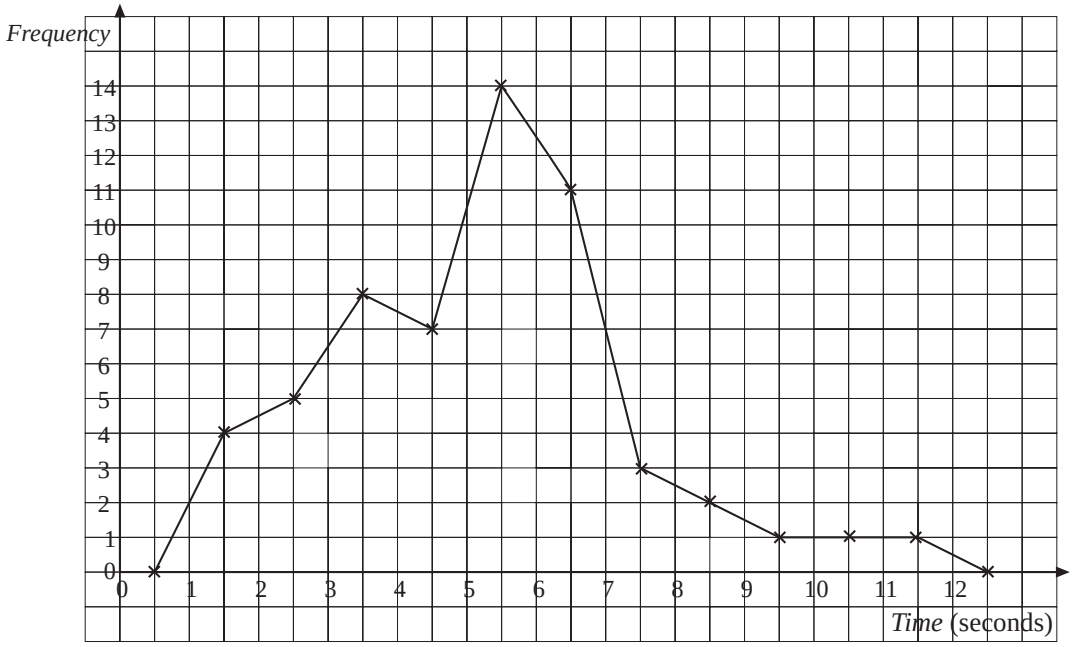
Height (cm)	11 - 15	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40
Frequency	3	7	19	20	11	2

(a) Draw a frequency polygon for the data.

(b) State the modal class.

(c) Calculate an estimate of the mean height.

7. Estimate the mean of the data illustrated in the following frequency polygon:



8. Children were asked to sell tickets for a school play. A record was kept of how many tickets each child sold.

Tickets Sold	0 - 10	11 - 20	21 - 50	51 - 100
Frequency	7	42	8	3

(a) Estimate the *mean* number of tickets sold.

(b) Estimate the *total* number of tickets sold.

9. A company owns a fleet of 20 vans. The mileage on each van is recorded. The results are given in the following table:

Mileage	$0 \le M < 5000$	$5000 \le M < 10\,000$	$10\,000 \le M < 15\,000$	$15\,000 \le M < 20\,000$
Frequency	1	4	8	7

(a) Illustrate the data with a bar chart.

(b) Estimate the mean mileage.

10. Joshua is given the data below and asked to estimate the mean.

Value	100 - 104	105 - 109	110 - 114	115 - 119
Frequency	5	16	32	7

(a) Calculate an estimate of the mean.

(b) Joshua also calculates that the mean must be greater than 107.9. Explain how he obtained this value.

(c) Determine a value that the mean must be less than.

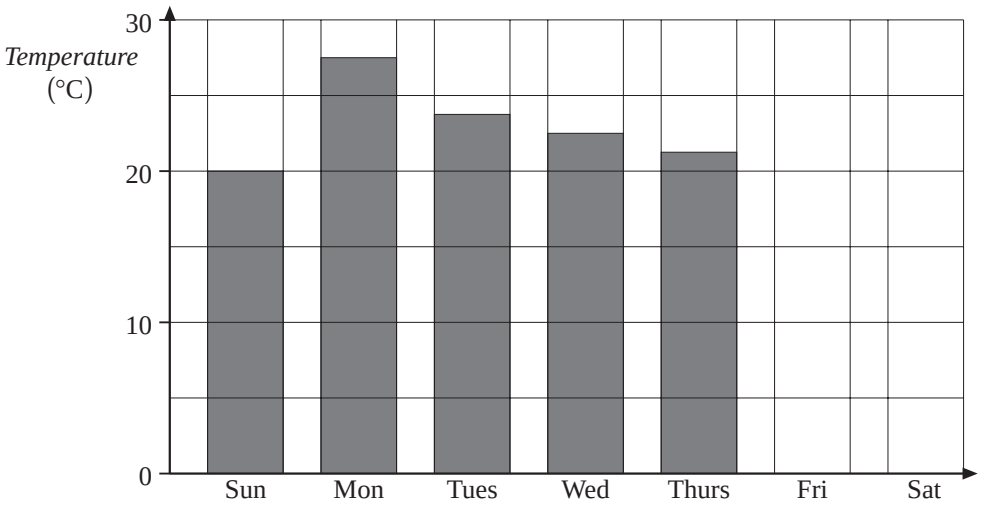
11. Lyn recorded the temperature at lunch time every day for a week.
She started to draw a bar chart to show her results.

(a) The temperature on *Friday* was 25 °C.

The temperature on *Saturday* was 19 °C.

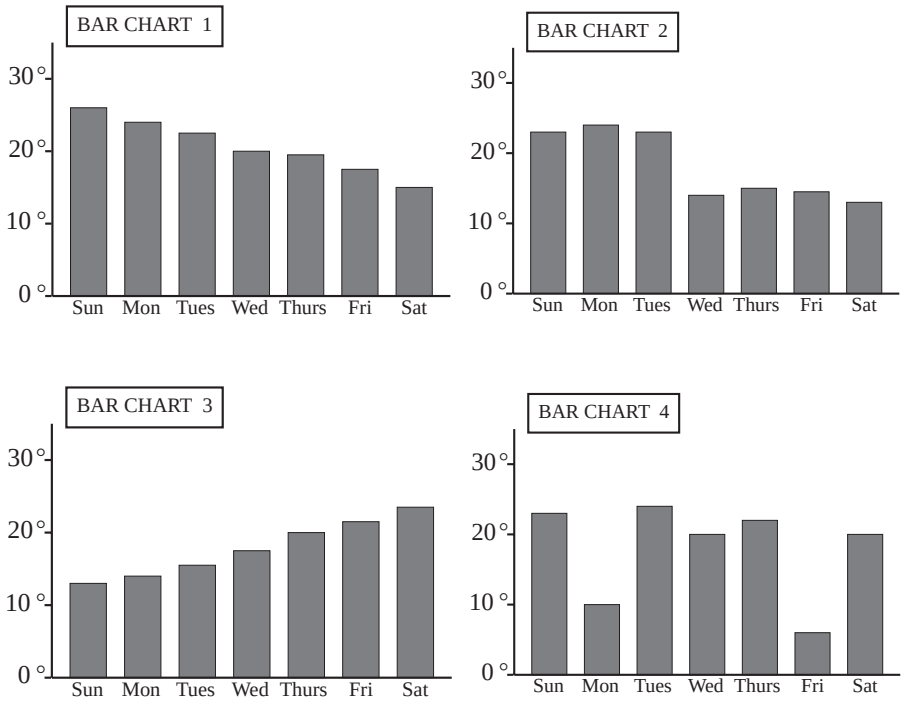
On a copy of Lyn's bar chart, draw the bars for *Friday* and *Saturday.*

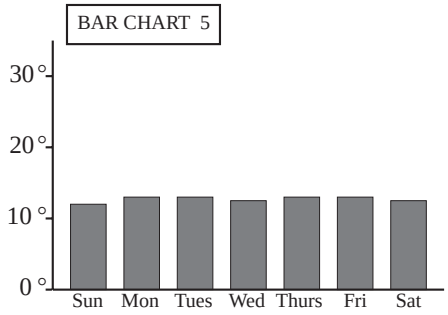
Lyn's Bar Chart



What was the temperature on *Monday*?

(b) Five more pupils recorded the temperature every day for different weeks in the year.





Match the pupils' comments to their bar charts. The first is done for you.

PUPILS' COMMENTS

Pupil A: *"It was very warm at first, then it suddenly got much colder."*

Pupil B: *"Each day was colder than the day before."*

Pupil C: *"The temperature was about the same all week."*

Pupil D: *"Each day was hotter than the day before."*

Pupil E: *"There were 5 warm days and 2 cold days."*

Pupil A: Bar Chart 2

Pupil B: Bar Chart .....

Pupil C: Bar Chart .....

Pupil D: Bar Chart .....

Pupil E: Bar Chart .....

(KS3/97/Ma/Tier 3-5/P2)

16.3 Cumulative Frequency

Cumulative frequencies are easy to calculate from a frequency table. Cumulative frequency graphs can then be used to estimate the median of a set of data. In this section we also look at the idea of *quartiles*, the *interquartile range* and the *semi-interquartile range*.

When you have a set of n values, in order,

$$\textit{Lower quartile} = \frac{n+1}{4}\text{th value}$$

$$\textit{Median} = \frac{n+1}{2}\text{th value}$$

$$\textit{Upper quartile} = \frac{3(n+1)}{4}\text{th value}$$

$$\textit{Interquartile range} = \text{upper quartile} - \text{lower quartile}$$

$$\textit{Semi-interquartile range} = \frac{\text{interquartile range}}{2}$$

If the data is arranged in an ordered list, and the number of data values, n, is odd then the $\frac{n+1}{2}$th value will be a single item from the list, and this will be the median. For example, if $n = 95$ the median will be the $\frac{95+1}{2} = 48\text{th}$ value.

However, if n is even then $\frac{n+1}{2}$ will determine the two central values that must be averaged to obtain the median. For example, if $n = 156$ then $\frac{156+1}{2} = 78.5$, which tells us that we must average the 78th and 79th values to get the median.

For large sets of data, we estimate the lower quartile, median and upper quartile using the $\frac{n}{4}$th, $\frac{n}{2}$th and $\frac{3n}{4}$th values. For example, if $n = 2000$, then we would estimate the lower quartile, median and upper quartile using the 500th, 1000th and 1500th values.

Example 1

For the following set of data,

4 7 18 3 9 5 10

(a) determine the *median*,

(b) calculate the *interquartile range*,

(c) calculate the *semi-interquartile range*.

Solution

First list the values in order:

3 4 5 7 9 10 18

(a) As there are 7 values, the median will be the $\frac{7+1}{2} = 4\text{th}$ value.

Median = 7.

(b) The lower quartile will be the $\frac{7+1}{4} = 2\text{nd}$ value.

Lower quartile = 4.

The upper quartile will be the $\frac{3(7+1)}{4} = 6\text{th}$ value.

Upper quartile = 10.

The interquartile range = upper quartile – lower quartile

$= 10 - 4$

$= 6$

The semi-interquartile range $= \frac{\text{interquartile range}}{2}$

$= \frac{6}{2}$

$= 3$

Example 2

(a) Draw a cumulative frequency graph for the following data:

Height (cm)	$150 \le h < 155$	$155 \le h < 160$	$160 \le h < 165$	$165 \le h < 170$	$170 \le h < 175$
Frequency	4	22	56	32	5

(b) Estimate the *median* from the graph.

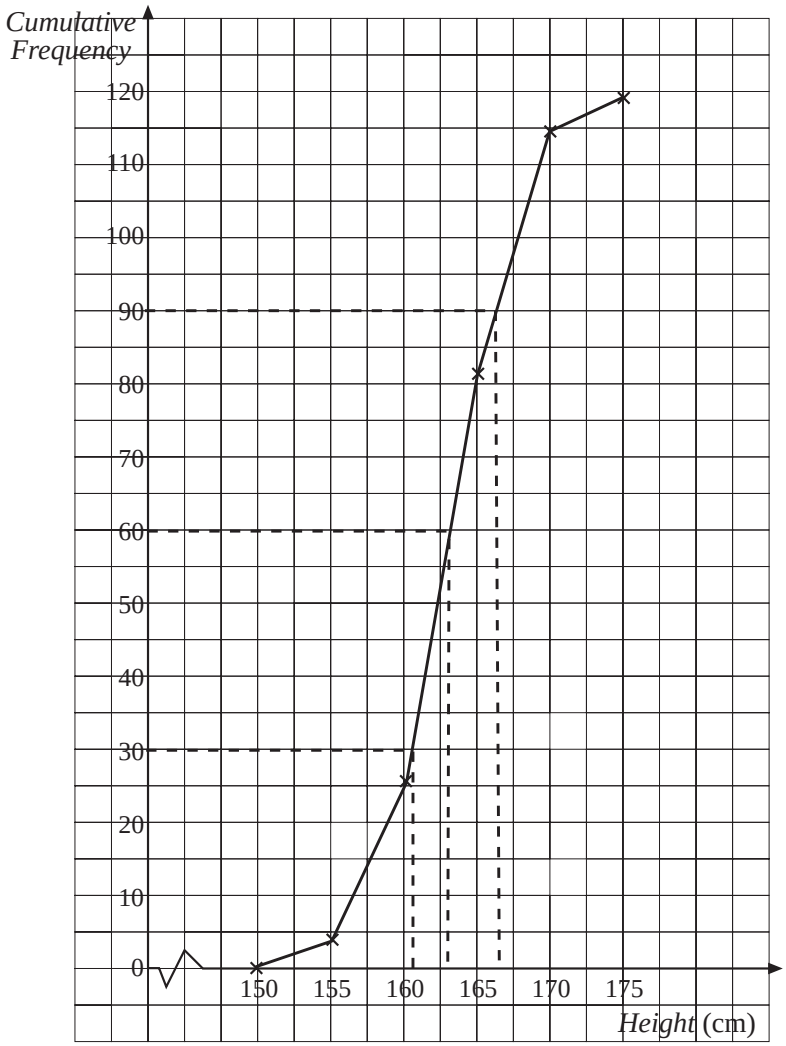
(c) Estimate the *interquartile range* from the graph.

Solution

(a) From the data table we can see that there are no heights under 150 cm.
Under 155 cm there are the first 4 heights.
Under 160 cm there are the first 4 heights plus a further 22 heights that are between 155 cm and 160 cm, giving 26 altogether.
Under 165 cm we have the 26 heights plus the 56 that are between 160 cm and 165 cm, giving 82 altogether.
Continuing this process until every height has been counted gives the following *cumulative frequency tabl*e.

Height (cm)	*Under 150*	*Under 155*	*Under 160*	*Under 165*	*Under 170*	*Under 175*
Cumulative Frequency	0	0 + 4 = 4	4 + 22 = 26	26 + 56 = 82	82 + 32 = 114	114 + 5 = 119

The cumulative frequency graph can now be plotted using the points in the table, (150, 0), (155, 4), (160, 26), (165, 82), (170, 114) and (175, 119).
To obtain the *cumulative frequency polygon,* we draw straight line sections to join these points in sequence.



(b) There are 119 values, so the median will be the $\frac{119+1}{2} = 60\text{th}$ value.

This can be read from the graph as shown above.

Median $\approx$ 163 cm.

The lower quartile will be given by the $\left(\frac{119+1}{4}\right)\text{th}$ value.

Lower quartile $\approx$ 160.5 cm.

The upper quartile will be given by the $\frac{3(119+1)}{4}\text{th}$ value.

Upper quartile $\approx$ 166.5 cm.

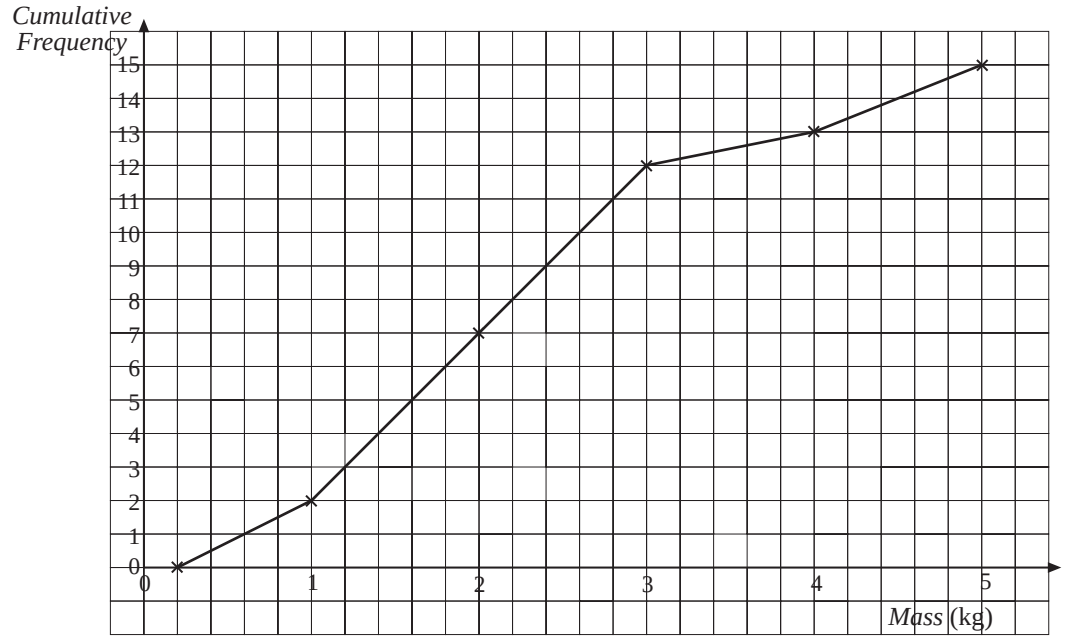
Using these values gives:

Interquartile range $= 166.5 - 160.5$

$= 6$ cm

Example 3

Estimate the semi-interquartile range of the data illustrated in the following cumulative frequency graph:

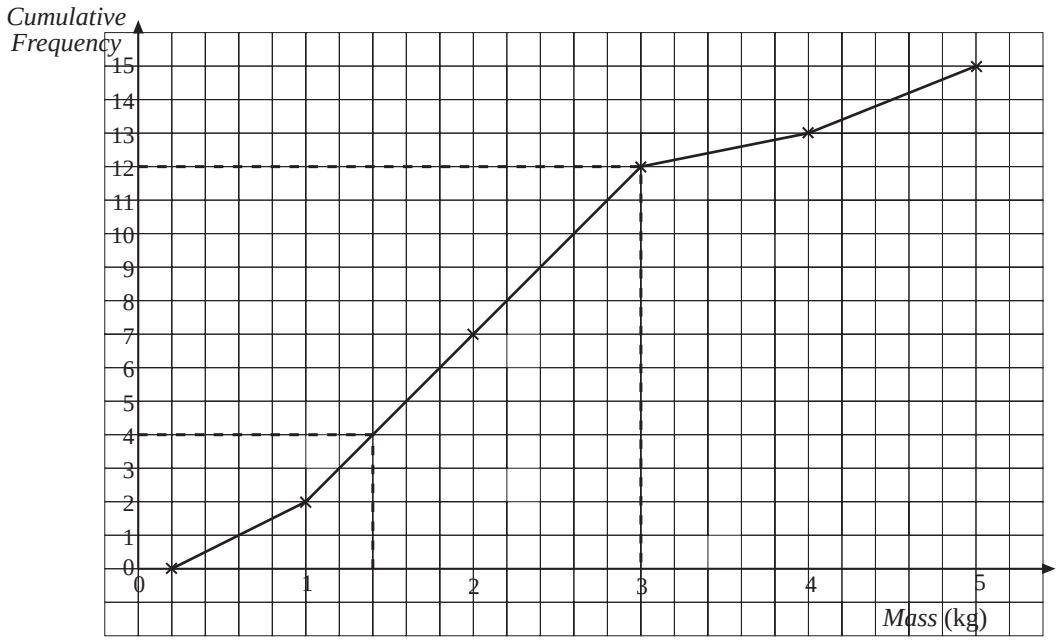


Solution

The sample contains 15 values, so the lower quartile will be the $\frac{15+1}{4} = 4\text{th}$ value.

Similarly, the upper quartile will be the 12th value.

These can be obtained from the graph, as follows:



Lower quartile = 1.4 kg

Upper quartile = 3 kg

Interquartile range = $3 - 1.4$

$= 1.6$ kg

Semi-interquartile range = 0.8 kg

Exercises

1. Determine the median and interquartile range of the following set of data:

   11 8 5 9 7 3 4 8 14 16 2

2. Calculate the semi-interquartile range of this sample:

   42 26 32 41 52 33 88 71 38 52 53 27 46 32 59

3. In a sample, the semi-interquartile range is 14. The lower quartile is 5 less than the median. Determine the median if the upper quartile is 91.

4. Below are the times, in minutes, spent on homework one evening by a group of students.

Time Spent (min)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$	$40 \le t < 50$
Frequency	3	7	10	15	4

(a) Copy and complete the following cumulative frequency table:

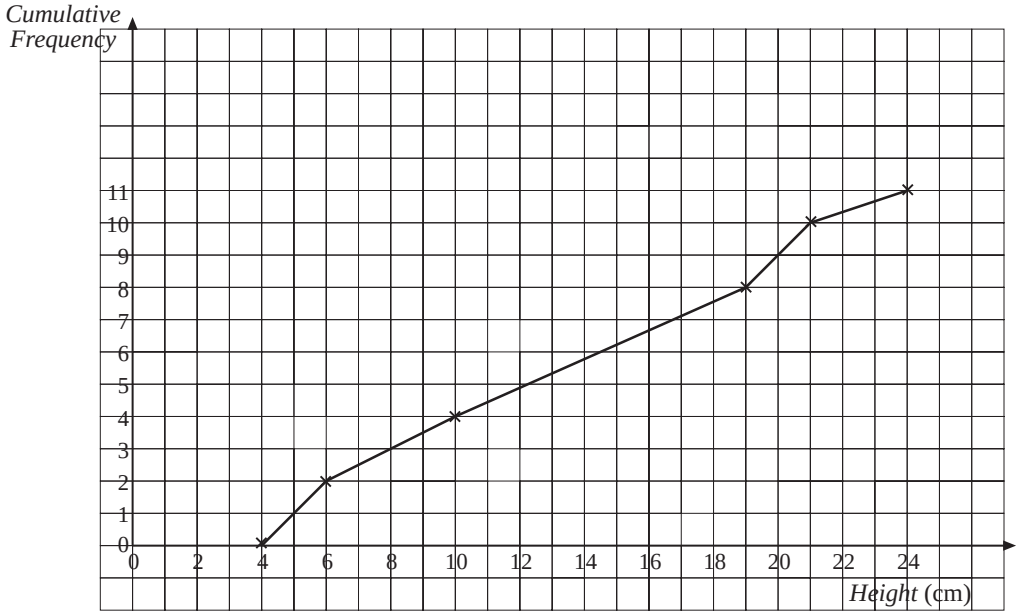
Time (min)	*Under 0*	*Under 10*	*Under 20*	*Under 30*	*Under 40*	*Under 50*
Cumulative Frequency						

(b) Draw a cumulative frequency polygon for this data.

(c) Use the polygon to estimate the median.

(d) Use the polygon to estimate the semi-interquartile range.

5. Estimate the median and interquartile range of the data illustrated in the following cumulative frequency graph:



6. (a) Gather data on the height of the pupils in your class.

(b) Draw a cumulative frequency graph for the data.

(c) Use the graph to estimate the median height and the semi-interquartile range.

7. Use a cumulative frequency graph to estimate the median and interquartile range of the following data:

Cost (£)	$10 \le c < 11$	$11 \le c < 12$	$12 \le c < 13$	$13 \le c < 14$	$14 \le c < 15$
Frequency	8	12	40	2	1

8. A factory collected data on the time for which a particular type of candle would burn. The data is summarised in the following table:

Time (mins)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$	$40 \le t < 50$
Frequency	1	2	12	15	5

(a) How do the mean and median compare?

(b) Determine the semi-interquartile range for the data.

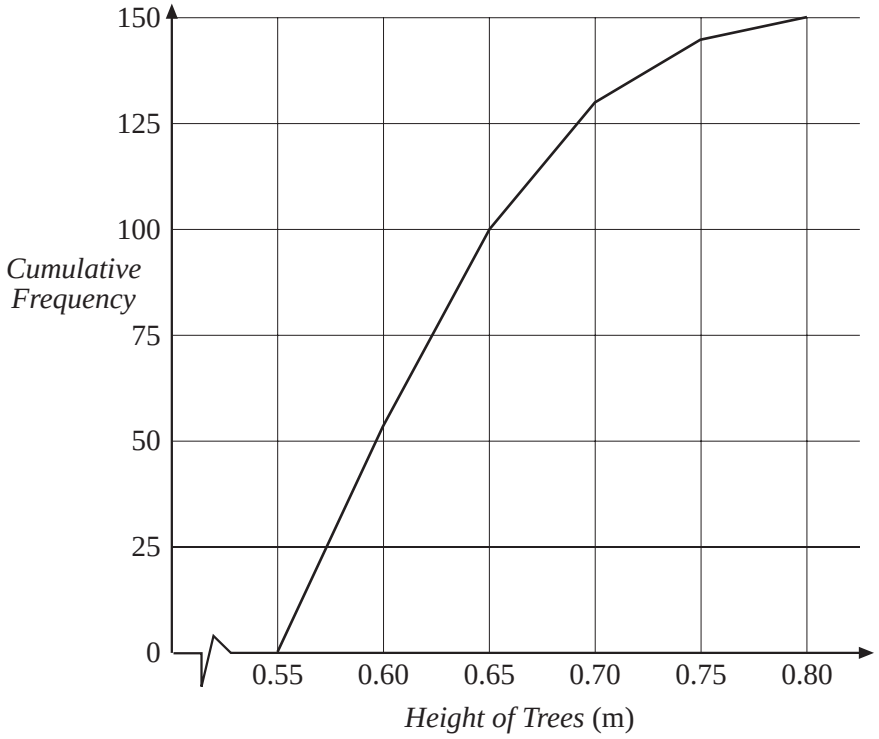
9. The number of passengers on a bus route was recorded over a period of time, to give the following data:

Number of Passengers	*Frequency*
$0 \le n < 10$	3
$10 \le n < 20$	7
$20 \le n < 30$	12
$30 \le n < 40$	13
$40 \le n < 50$	29
$50 \le n < 60$	27

Determine the median and semi-interquartile range of the data.

10. Give an example of a sample for which the semi-interquartile range is a quarter of the range of the sample.

11. The cumulative frequency graph shows the height of 150 Norway fir trees.



(a) Use the graph to estimate the *median* height and the *interquartile range* of the Norway firs.

(b) Which one of the following sketches of frequency diagrams shows the distribution of heights of the Norway firs?

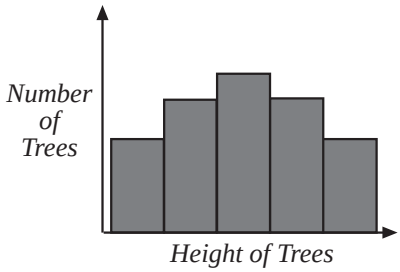


Diagram A

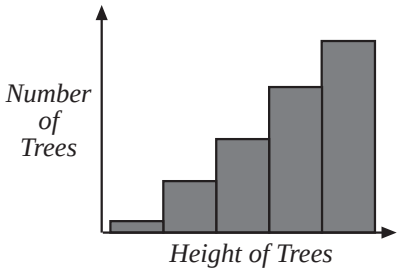


Diagram B

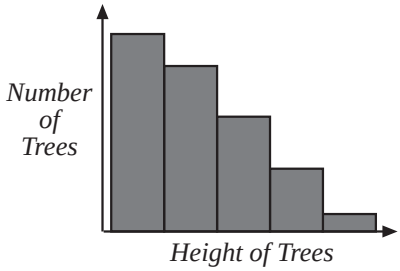


Diagram C

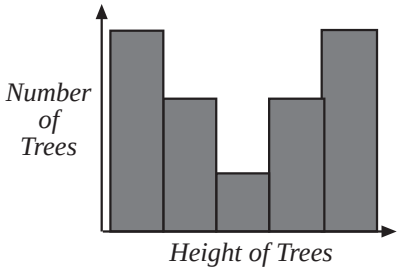
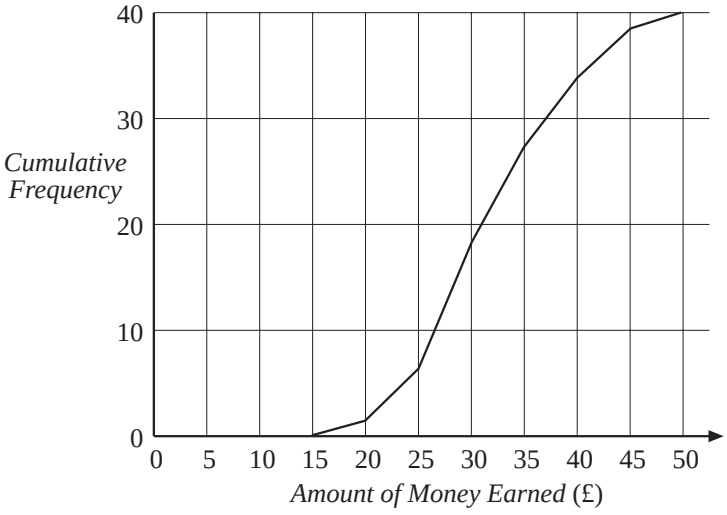


Diagram D

(KS3/98/Ma/Tier 6-8/P2)

12. 40 students worked on a farm one weekend. The cumulative frequency graph shows the distribution of the amount of money earned. No one earned less than £15.



(a) Read the graph to estimate the *median* amount of money earned.

(b) Estimate the *percentage* of students who earned *less than £40.*

(c) On a copy of the graph, show how to work out the *interquartile range* of the amount of money earned.

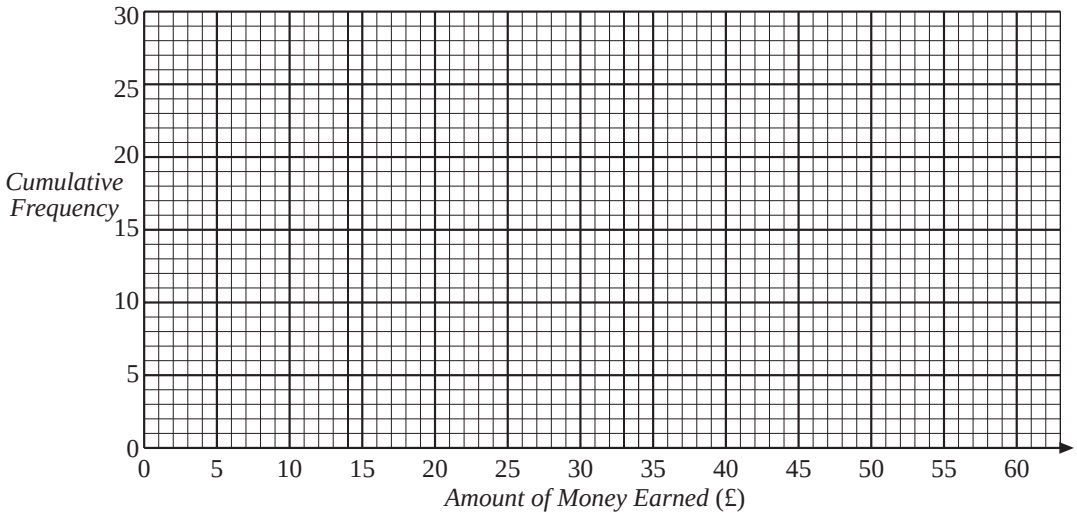
Write down the value of the interquartile range.

(d) 30 of the students work on the farm another weekend later in the year. The tables which follow show the distribution of the amount of money earned by the students.

Money Earned (£)	*No. of Students*
≥ 25 and< 30	1
≥ 30 and < 35	2
≥ 35 and < 40	3
≥ 40 and < 45	4
≥ 45 and < 50	10
≥ 50 and < 55	7
≥55 and < 60	3

Money Earned(£)	*No. of Students*
< 25	0
< 30	1
< 35	3
< 40	6
< 45	10
< 50	20
< 55	27
< 60	30

Draw a cumulative frequency graph using a copy of the axes below.



(e) State whether each of the following statements is *true* or *false.*

A. Three of the students earned less than £35 each.

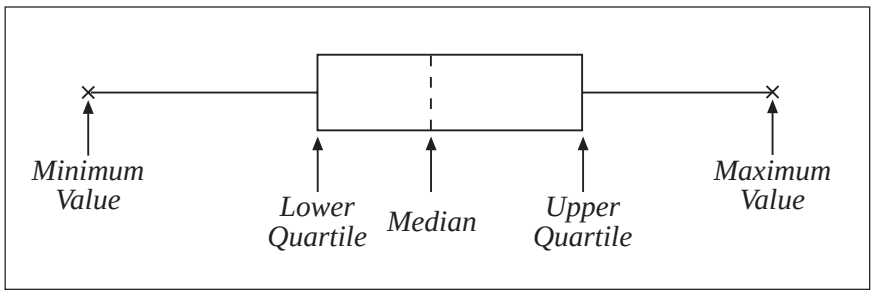
B. The median amount earned is between £40 and £45.

C. Most of the 30 students earned more than £50 each.

(KS3/97/Ma/Tier 6-8/P1)

16.4 Box and Whisker Plots

A box and whisker plot is based on the *minimum and maximum values*, the *upper and lower quartiles* and the *median*. This type of plot provides a good way to compare two or more samples.



Note: Box and whisker plots must always be drawn accurately to scale.

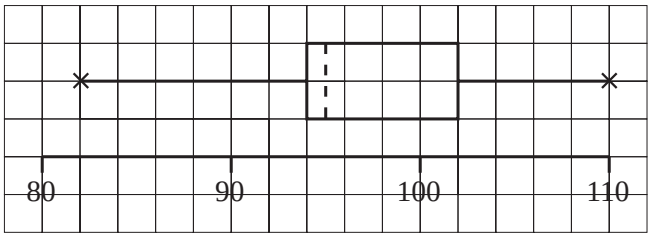
Example 1

Given the information below, draw a box and whisker plot.

Minimum	82
Lower quartile	94
Median	95
Upper quartile	102
Maximum	110

Solution

The box and whisker plot is shown below.



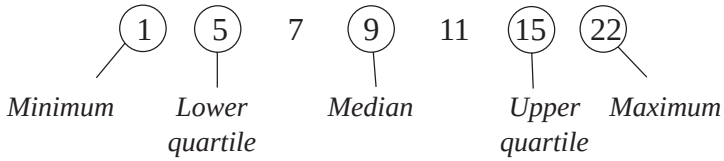
Example 2

Draw a box and whisker plot for this sample:

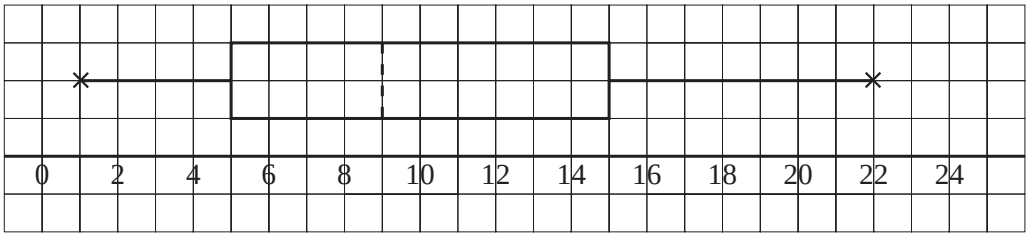
5 7 1 9 11 22 15

Solution

First list the sample in order, to determine the median and the quartiles.



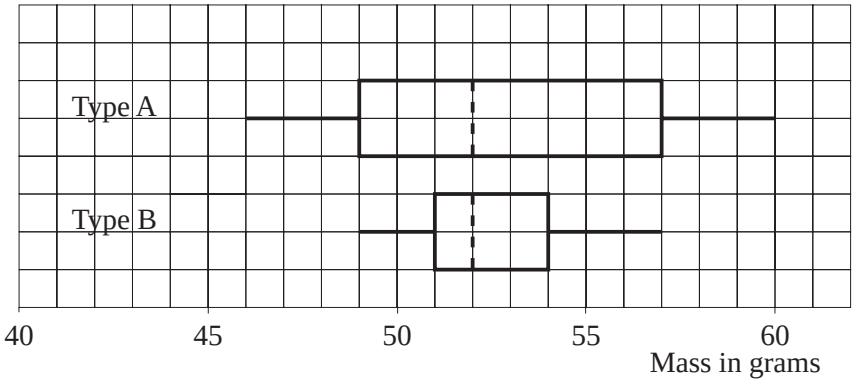
The box and whisker plot is shown below:



Example 3

A gardener collected data on two types of tomato. The box and whisker plot below shows data for the masses in grams of the tomatoes in the two samples.

Compare and contrast the two types and advise the gardener which type of tomato he should grow in future.



Solution

	Type A	*Type B*
Median	52 grams	52 grams
Lower Quartile	49 grams	51 grams
Upper Quartile	57 grams	54 grams
Range	14 grams	8 grams
Interquartile Range	8 grams	3 grams

From this table we can see that both types of tomato have the same average mass because their medians are the same.

Comparing the medians and interquartile ranges shows that there is far more variation in the masses of the type A tomatoes, which means that the masses of type B are more consistent than those of type A.

However, comparing the two box and whisker plots, and the upper quartiles, shows that type A tomatoes will generally have a larger mass than those of type B.

Nevertheless, there will be some type A tomatoes that are lighter than any of type B.

Taking all this together, the gardener would be best advised to plant type A tomatoes in future as he is likely to get a better yield from them than from type B.

Exercises

1. Draw a box and whisker plot for a sample that has:

Minimum	10
Lower quartile	14
Median	16
Upper quartile	20
Maximum	29

2. Draw a box and whisker plot for the following sample:

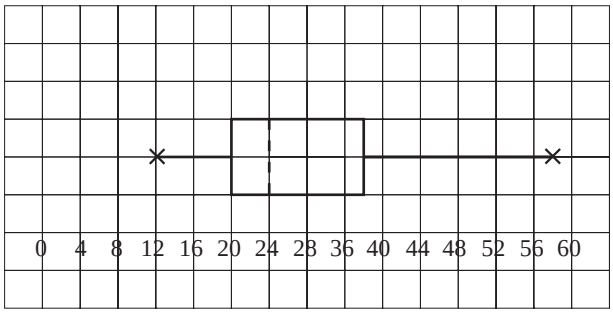
17 22 18 33 14 36 39 41
25 31 18 19 16 21 21

3. A sample has:

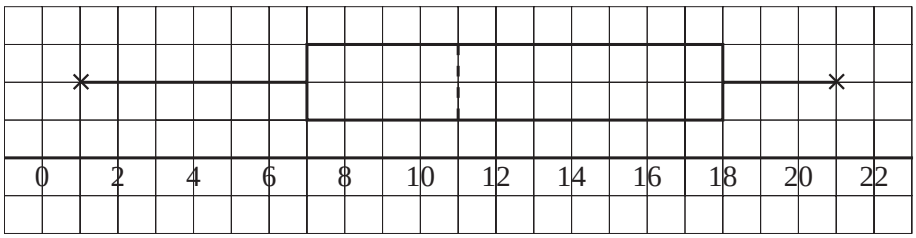
Minimum 3 Range 21
Semi-interquartile range 4 Median 17
Upper quartile 20

Draw a box and whisker plot for the sample.

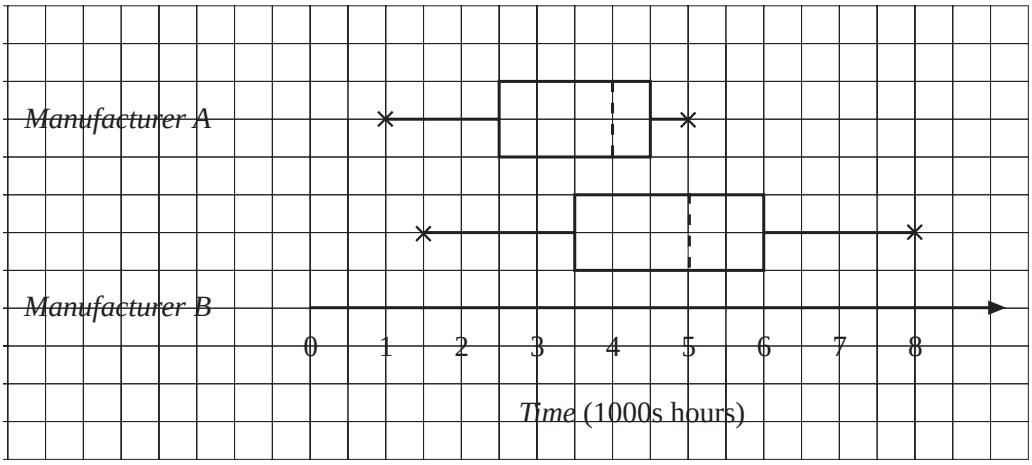
4. For the sample illustrated in the following box and whisker plot, determine:
(a) the range, (b) the semi-interquartile range.



5. What are the median and the semi-interquartile range of the following sample:

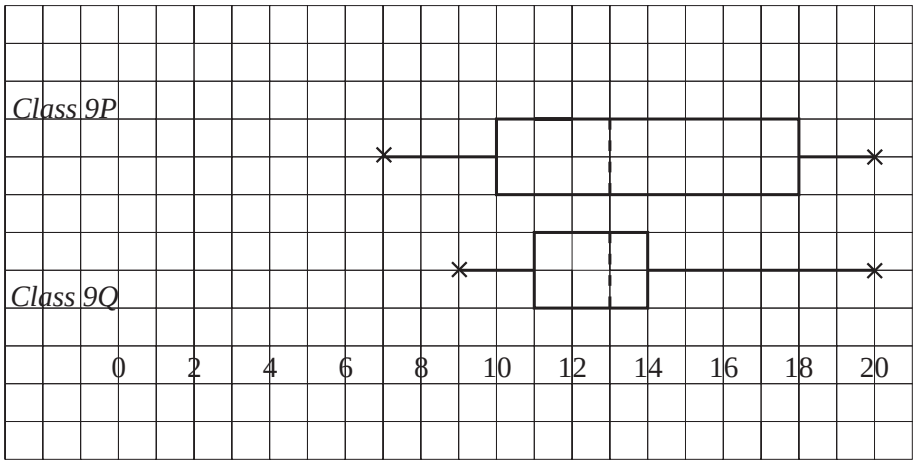


6. The two box and whisker plots show the data collected by the manufacturers on the life-span of light bulbs.

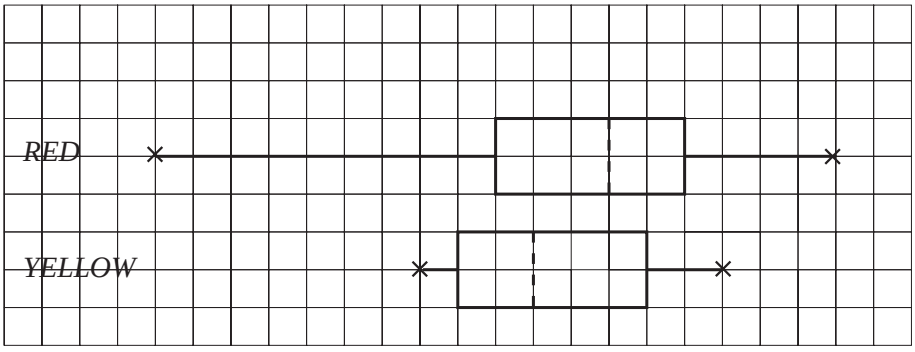


From this data, which manufacturer produces the better light bulb? Give reasons for your answer.

7. A maths test is given to two classes. The results are illustrated below. Compare and contrast the results.



8. A builder can choose between two different types of brick that are coloured *red* or *yellow*. The box and whisker plots below illustrate the results of tests on the strength of the bricks.

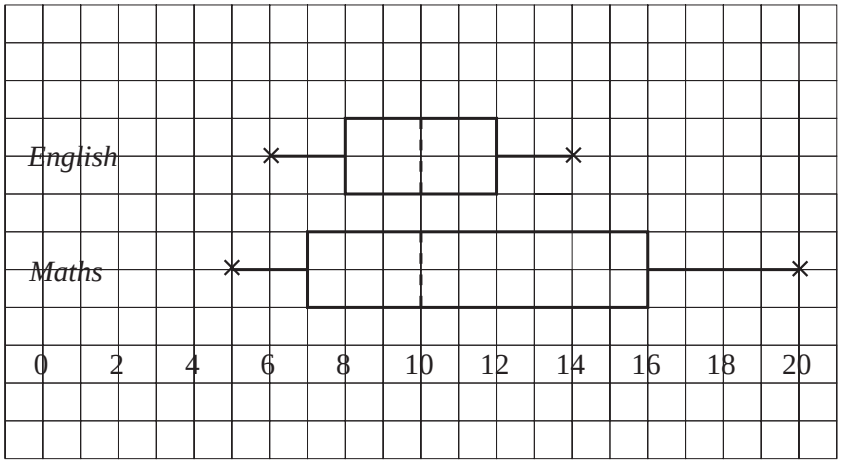


From the data illustrated in the box and whisker plots:

(a) give one reason why the builder might prefer to use *red* bricks.

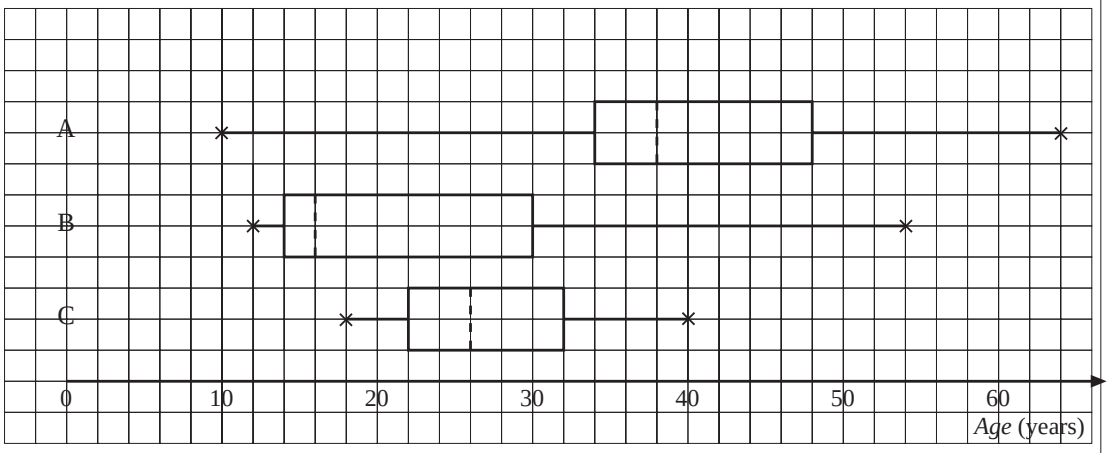
(b) give one reason why the builder might prefer to use *yellow* bricks.

9. A class took an English test and a Maths test. Both tests had a maximum possible mark of 25. The results are illustrated below.



Compare and contrast the results.

10. A cinema is showing 3 films, A, B and C. The ages of people watching the films are illustrated in the following box and whisker plots:



Answer the following questions, giving reasons to support your answers.

(a) Which film do you think you would *not* be allowed to watch?

(b) Which film would *you* probably enjoy most?

(c) Which film would *your parents* probably enjoy most?