Median

The median is an average that is not influenced by extreme values.
The median is the $\frac{1}{2}(n+1)^{th}$ value.

Examples

(i) $n = 7$ 2 3 8 **9** 10 14 23

median = 9

When there are 7 numbers, the median is the 4th in the list.

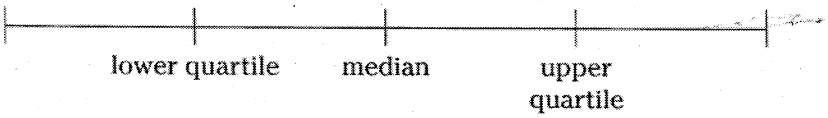
(ii) $n = 6$ 145 152 160 164 173 185

$$\text{median} = \frac{160 + 164}{2} = 162$$

When there are 6 numbers, the median is the 3.5th value, i.e. the mean of the 3rd and 4th values.

Quartiles

The lower and upper quartiles are not influenced by extreme values. They are values such that, together with the median, they split a distribution into four equal parts.



Finding Median and Quartiles – Raw Data/Ungrouped Data

Example 1

In this set of 15 numbers, the median is 35.

23 25 25 **27** 29 30 32 **35** 36 38 40 **42** 43 45 47

median

Notation:
The lower quartile is often denoted by Q_1, the median by Q_2 and the upper quartile by Q_3.

For data arranged in order,

- the **lower quartile**, Q_1 is the median of all the values **before** the median
- the **upper quartile**, Q_3 is the median of all the values **after** the median.

Interquartile range

The difference between the quartiles gives the **interquartile range**. It tells you the range of the middle 50% of the data. It is a particularly useful measure of spread when there are outliers, as it is unaffected by one or two extreme values.

$$\begin{aligned}\text{Interquartile range} &= \text{upper quartile} - \text{lower quartile} \\ &= Q_3 - Q_1\end{aligned}$$

Don't forget to subtract.

Example 2

Find the interquartile rang for the following sets of data.

(i) 7, 7, 2, 3, 4, 2, 7, 9, 31

(ii) 36, 41, 27, 32, 29, 39, 39, 43

Finding Median and Quartiles - Stemplots

Example 3

To investigate hand-eye coordination in reacting to a stimulus, students took part in an experiment where a ruler was dropped and the distance it travelled before the student caught it was measured. The results of 21 girls and 27 boys are shown in the back-to-back stem-and-leaf diagram.

Distance, in cm

	Girls		Boys	
		4	8 9	(2)
		5	0 5 9	(3)
(1)	8	6	2 4	(2)
(5)	5 5 3 2 2	7	1 4 5	(3)
(2)	7 6	8	0 2 5 7 8 9	(6)
(6)	9 9 8 4 1 1	9	3 3 6	(3)
(3)	9 6 3	10	0 0 5 7 8	(5)
(3)	7 5 3	11	2 3	(2)
(1)	4	12	7	(1)

Key:
8 | 6 | 2 represents a distance of 6.8 cm for the girls and 6.2 cm for the boys.

Find the median and interquartile range for both sets of data. Comment on your answers.

	Distance (cm)	
	Girls	Boys
Median	9.4	8.7
Interquartile range	3.25	3.6

The distance the ruler travelled was further on average for the girls, indicating that the girls were slower to react.

However, the larger interquartile range for the boys indicates that their results were more variable. The girls' results were more consistent.

Cumulative Frequency for Grouped Data

When data have been grouped, cumulative frequencies are calculated up to each **upper class boundary**, as in the following example.

The frequency table shows the heights of broad bean plants six weeks after planting.

Height, x cm	$3 \leqslant x < 6$	$6 \leqslant x < 9$	$9 \leqslant x < 12$	$12 \leqslant x < 15$	$15 \leqslant x < 18$	$18 \leqslant x < 21$	
Frequency	1	2	11	10	5	1	Total **30**

The upper class boundaries are 6, 9, 12, 15, 18 and 21.

The lower boundary of the first interval is 3. It is a good idea to include this in the cumulative frequency table.

This is the cumulative frequency table

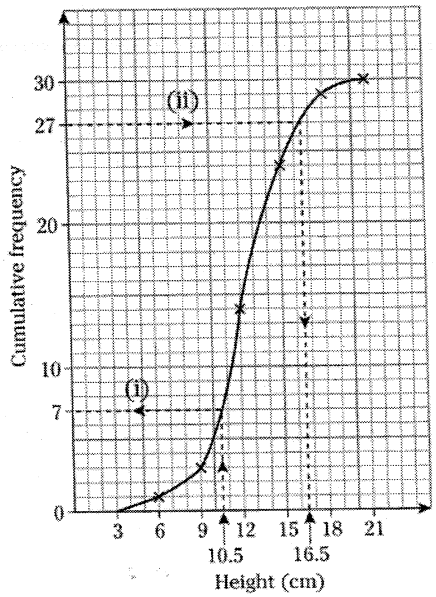
Height, x cm	< 3	< 6	< 9	< 12	< 15	< 18	< 21
Cumulative frequency	0	1	3	14	24	29	**30**

1 + 2 | 1 + 2 + 11 | This must be the same as the total frequency.

Cumulative frequencies can be thought of as 'running totals' of the frequencies.

The cumulative frequency table can be illustrated on a **cumulative frequency graph** in which the cumulative frequencies are plotted against the **upper class boundaries**. The points are joined either with a curve or with straight lines. Often curves are drawn, but either method is acceptable in the examination.

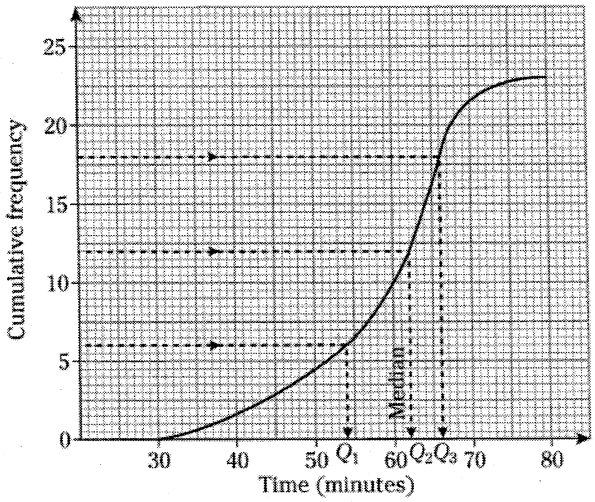
Cumulative frequency curve to show heights of 30 broad bean plants



Finding Median and Quartiles - Grouped Data

When data have been grouped into intervals, the original information has been lost, so it is only possible to make **estimates** of the median and quartiles. One way of doing this is to read off the appropriate values from a cumulative frequency graph.

The following cumulative frequency curve shows the times spent by a group of 23 students on th statistics assignment.



There are 23 students.

Median $Q_2 = \frac{1}{2}(23 + 1)^{th}$ value $= 12^{th}$ value.

Q_1 is the median of the 11 values before the median $= 6^{th}$ value.

Q_3 is the median of the 11 values after the median $= (12 + 6)^{th}$ value $= 18^{th}$ value.

Using the graph, estimates are as follows:

median $Q_2 = 62$ minutes,

lower quartile $Q_1 = 54$ minutes,

upper quartile $Q_3 = 66$ minutes

Example 5

The arrival times of 204 trains were noted and the number of minutes, t, that each train was late was recorded.

Number of minutes late (t)	$-2 \leqslant t < 0$	$0 \leqslant t < 2$	$2 \leqslant t < 4$	$4 \leqslant t < 6$	$6 \leqslant t < 10$
Number of trains	43	51	69	22	19

(i) Explain what $-2 \leqslant t < 0$ means about the arrival times of the trains.

(ii) Draw a cumulative frequency graph, and from it estimate the median and interquartile range of the number of minutes late of these trains.

Cambridge Paper 6 Q5 N07

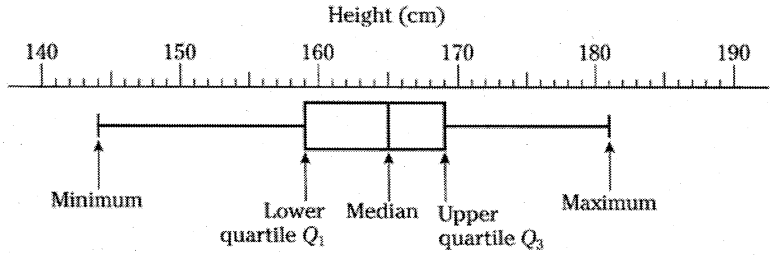
Box and Whiskers Plots/ Boxplots

In a **box-and-whisker plot** the median and quartiles are shown, as well as the minimum and maximum values of a distribution. It gives a very good visual summary of a distribution and is particularly useful when comparing sets of data.

A survey on the heights of all the girls in a particular year group in a school gave the following information.

Minimum height	144 cm
Lower quartile	159 cm
Median	165 cm
Upper quartile	169 cm
Maximum height	181 cm

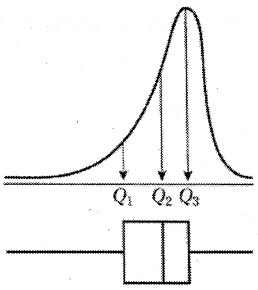
Box-and-whisker plot to show heights of 15-year-old girls



The shape a distribution

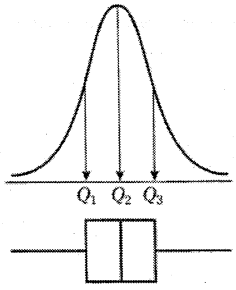
A box-and-whisker plot illustrates the spread of a distribution and also gives an idea of the shape of the distribution.

Negative skew



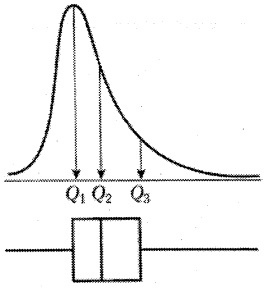
The whisker to the left is longer and the median is nearer to Q_3.

Symmetrical



The whiskers are of equal length and the median is in the middle of the box.

Positive skew



The whisker to the right is longer and the median is nearer to Q_1.

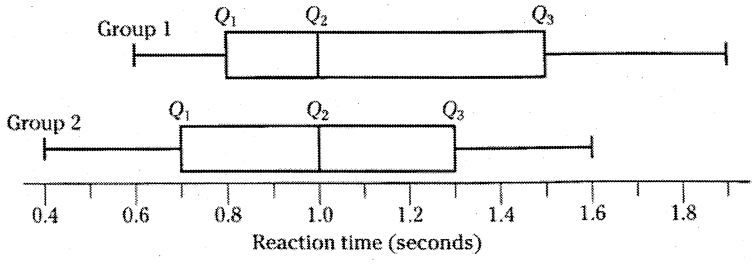
Example 6

Two groups of people played a computer game which tested how quickly they reacted to a visual instruction to press a particular key. The computer measured their reaction times in seconds, to the nearest tenth of a second. The following summary statistics were displayed for each group.

	Minimum	Lower quartile Q_1	Median Q_2	Upper quartile Q_3	Maximum
Group 1	0.6	0.8	1.0	1.5	1.9
Group 2	0.4	0.7	1.0	1.3	1.6

Draw two box-and-whisker plots and compare the reaction times of the two groups.

Box-and-whisker plots to show reaction times



The median reaction time is 1.0 seconds for both groups. However, the range of times for Group 2 is smaller than for Group 1 and the times are evenly distributed for Group 2.

There is a greater spread of times for Group 1. Their distribution has a long tail to the right, indicating that there are a few very high values.

In general, Group 2 has the faster reaction time.

Note:

A box-and-whisker plot can be drawn on the same diagram as a cumulative frequency graph.