Frequency Distributions and Cumulative Frequency

Cumulative frequency

The construction of a cumulative frequency table is best illustrated by example.

<u>Example</u>

The following data concern a random sample of 1000 men with heights in the given ranges.

Height (cm)	No. of men	Height (cm)	No. of men
164 –	2	178 –	130
166 –	6	180 –	105
168 –	28	182 –	108
170 –	74	184 –	78
172 –	110	186 –	52
174 –	126	188 –	31
176 –	137	190 – 192	13

Draw a cumulative frequency polygon to illustrate these data.

Solution

Firstly, we have to create a cumulative frequency table. Below 164 cm there are 0 men. Between 164 and 166 cm there are 2 men, so below 166 cm there are 2 men. Between 166 and 168 cm there are 6 men, so below 168 cm there are $2 + 6 = 8$ men, and so forth.

This enables us to draw the following cumulative frequency table

Height (cm)	Cumulative frequency
< 164	0
< 166	2
< 168	8
< 170	36
< 172	110
< 174	220
< 176	346
< 178	483
< 180	613
< 182	718
< 184	826
< 186	904
< 188	956
< 190	987
< 192	1000

From which we can plot the cumulative frequency polygon $(n = 1000)$.
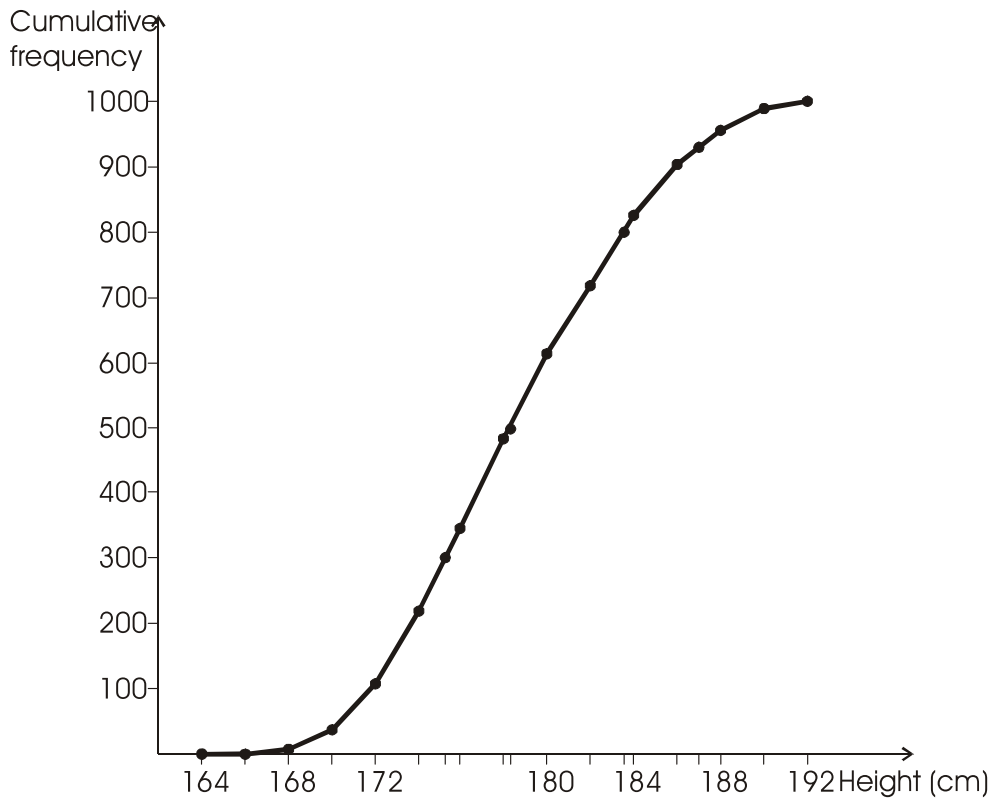
Median, quartiles and percentiles

The median is the value below which 50% of the cumulative frequency distribution lies.

The first quartile is the value below which 25% of the cumulative frequency distribution lies.

The third quartile is the value below which 75% of the cumulative frequency distribution lies.

The median is the same as the second quartile.

The range is the difference between the third and first quartiles. However, other ranges can be asked for.

The nth percentile is the value below which n% of the cumulative frequency distribution lies.

We illustrate these definitions by continuing the above example.

Example continued

Use your graph to estimate

(a)   the median height,

(b)   the range of heights for men who are between the thirtieth and eightieth percentiles,

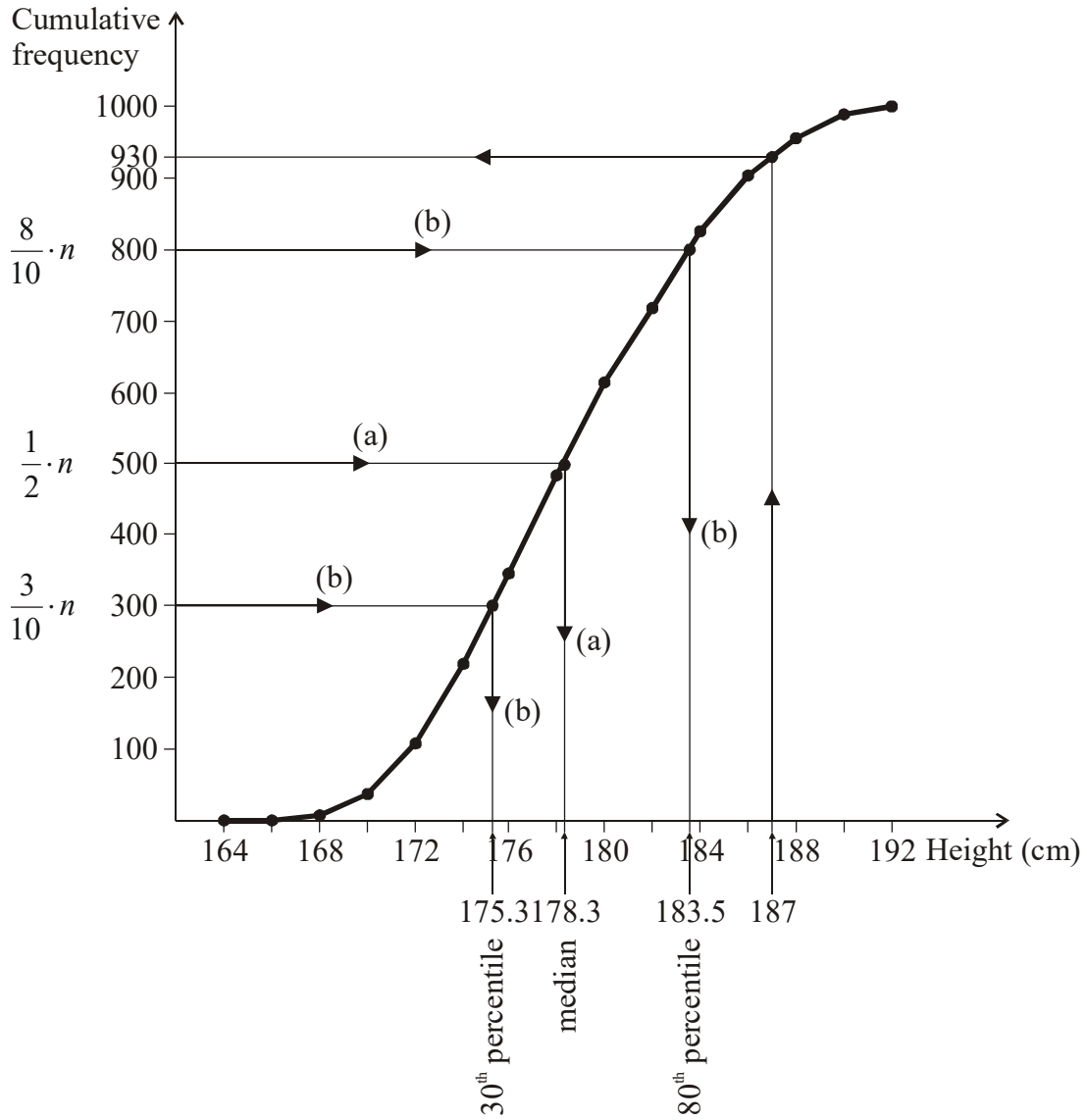(c)   the number of men in the sample with heights of at least 187 cm.

Solution

[The graph follows on the next page]

(a)   From the cumulative frequency polygon we estimate that the median height equals 178.3 cm.

(b)   From the cumulative frequency polygon we estimate that 30% of men had heights less than 175.3 cm and 80% of men had heights less than 183.5

© **blacksacademy.net**

cm. Thus, the range of heights for men who one between the 30th and 80th percentiles equals $183.5 - 175.3 = 8.2$ cm .

(c)   From the cumulative frequency polygon we estimate that 930 men in the sample had heights less than 187 cm. Hence, there were 70 men in the sample with heights of at least 187 cm.

Frequency polygons

A frequency polygon is another way to represent data graphically. It is an extension of a histogram. We will introduce it by means of example.

Example

The amounts spent by 180 shoppers at a supermarket were recorded and are tabulated below.

Amount spent, £	Number of shoppers
$x \le 15$	18
$15 < x \le 30$	48
$30 < x \le 40$	55
$40 < x \le 50$	39
$50 < x \le 70$	14
$70 < x \le 100$	6

Draw a histogram to represent this information.

Use this histogram to draw a frequency polygon for the distribution.

Estimate the 60th percentile.

Solution

The class boundaries are: 0, 15, 30, 40, 50, 70, 100
The class widths are: 15, 15, 10, 10, 20, 30

Notice that the class widths are not equal. If we were to draw a histogram with heights proportional to the frequencies (here the numbers of shoppers) we would distort the information and make it appear that there were more shoppers spending sums of money in those classes where the widths are largest. In order to avoid this we must calculate the *frequency density* of each class.
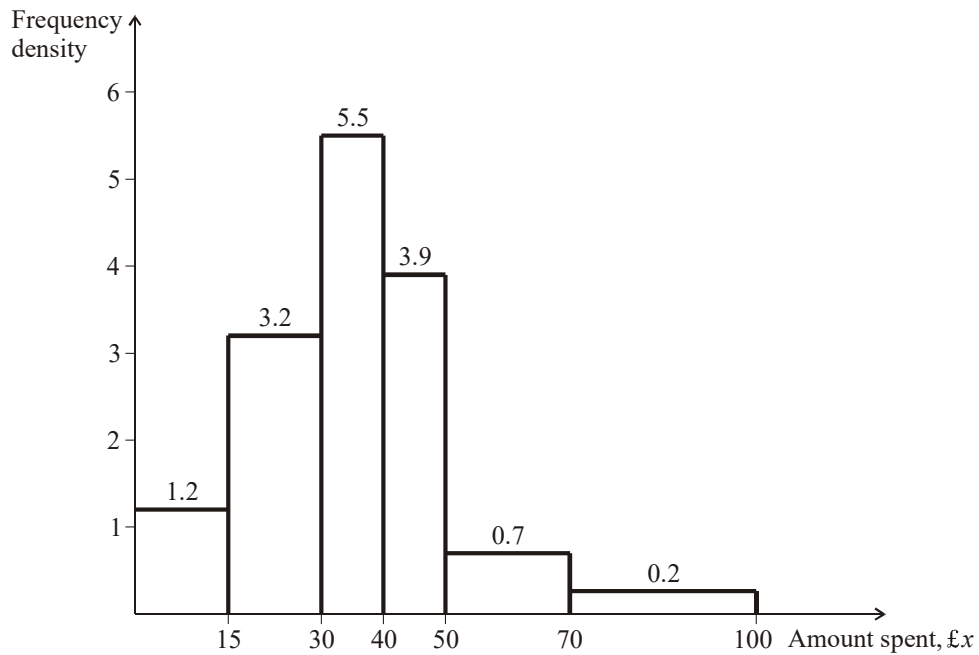
The frequency density is given by

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

This is illustrated by the following table.

Amount spent, £x	Class width	Frequency	Frequency density	Cumulative frequency
$0 < x \leq 15$	15	18	$\frac{18}{15} = 1.2$	18
$15 < x \leq 30$	15	48	$\frac{48}{15} = 3.2$	66
$30 < x \leq 40$	10	55	$\frac{55}{10} = 5.5$	121
$40 < x \leq 50$	10	39	$\frac{39}{10} = 3.9$	160
$50 < x \leq 70$	20	14	$\frac{14}{20} = 0.7$	174
$70 < x \leq 100$	30	6	$\frac{6}{30} = 0.2$	180

We can now draw the histogram.

The frequency polygon is drawn by joining the midpoints of the rectangles that make up the histogram.
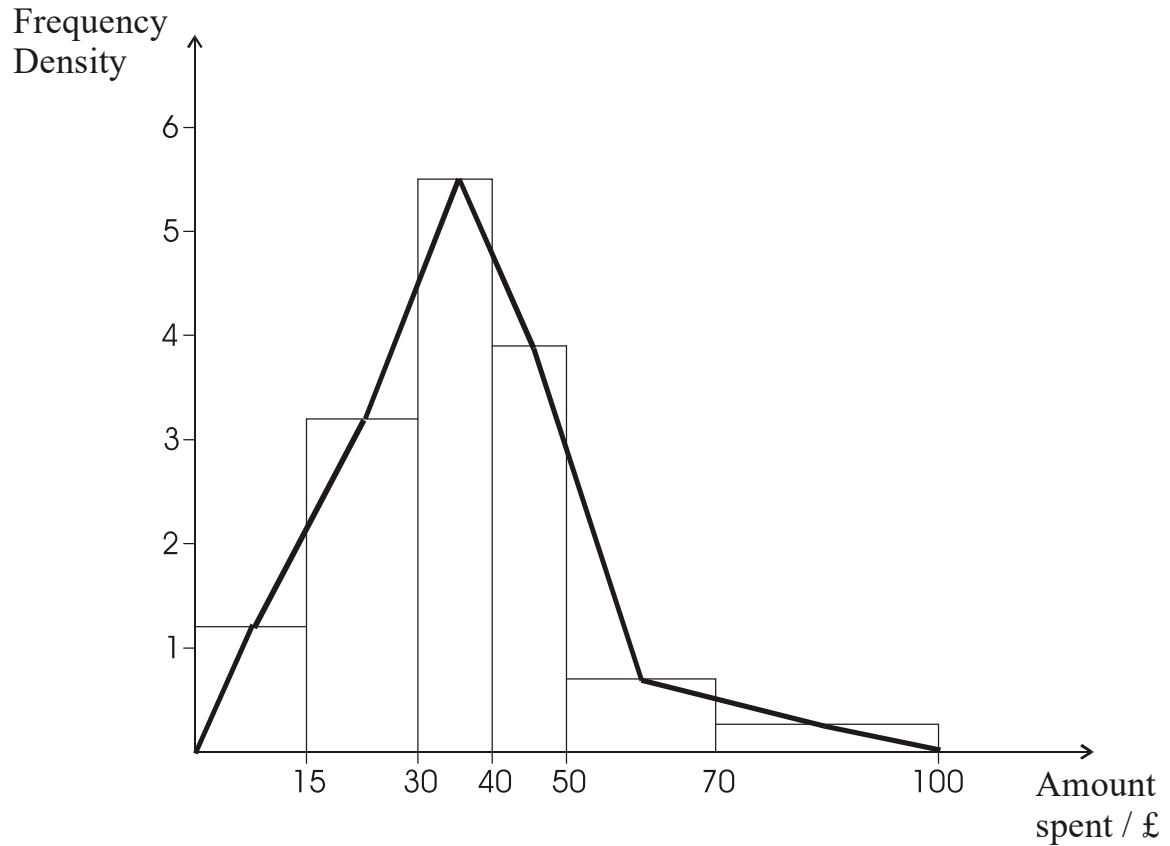
To estimate the 60$^{\text{th}}$ percentile we could draw the cumulative frequency polygon and work from that. However, here we use a more numerical form of reasoning.

The 60$^{\text{th}}$ percentile is the $0.6 \cdot n^{\text{th}}$ value $= 0.6 \cdot 180^{\text{th}}$ value $= 108^{\text{th}}$ value.

This value is in the interval is in the interval from £30 to £40.

As the width of this interval is 10, there are 55 values in this interval, and there are 66 values in the two intervals before it, the estimated value of the 60$^{\text{th}}$ percentile is

$$30 + \frac{(108 - 66)}{55} \times 10 = 30 + \frac{84}{11} = \text{£37.6} \ (1.\text{d.p})$$

© blacksacademy.net