

MEDIAN AND CUMULATIVE FREQUENCY

THE MEDIAN TEMPERATURES

The median is the third type of statistical average you will use in his course. You met the other two, the mean and the mode in pack MS4.

THE MEDIAN is a set of numbers which is in the middle when the numbers have been arranged in size order. Look at the following examples.

Example 1

Find the median value of the numbers 3, 7, 2, 9, 10, 6, 4.

Answer

First arrange the numbers in order of size:

2, 3, 4, 6, 7, 9, 10
 |
 median

There are several numbers so in this case the median is the fourth one.

Median= 6

If it should happen that there are an even number of values the set of numbers then the median is the average if the two centre numbers.

Example 2

Find the median value of the following lengths

12 cm, 19 cm, 4 cm, 13 cm, 20 cm, 39 cm

Answer

Arrange the lengths in size order:

4 cm, 12 cm, 13 cm, 19 cm, 20 cm, 39 cm
 ↑
 median

There are six items in the set (an even number) so there are two middle values – the third and the fourth.

The median is the average of these.

$$\text{Median} = \frac{13 + 19}{2} = \frac{32}{2} = 16 \text{ cm}$$

In general, to **find the median value of distribution**, if there are n terms in the distribution the median is given by the $\frac{n+1}{2}$ th term

(When the terms have been arranged in size order)

If there were 35 values the median would be the $\frac{35+1}{2} = 18^{\text{th}}$ value

If there were 36 values the median would be the $\frac{36+1}{2} = 18\frac{1}{2}^{\text{th}}$ value

In other words the average of the 18th and 19th values.

Exercise 1

- Find the median of: 38, 19, 2, 8, 15, 22, 13, 6, 5.
- What is the median of: 9, 8, 17, 47, 3, 4, 15, 40, 18, 28.

TO FIND THE MEDIAN OF FREQUENCY DISTRIBUTION

Example 3

Look at the following distribution. It refers to the number of apples found in a per-packed 2 kg bags.

Number of apples in bag	Number of bags (frequency)
16	3
17	5
18	6
19	5
20	$\frac{2}{21}$

Answer

The median of 21 values is given by the $\frac{21+1}{2} = 11^{\text{th}}$ term

To find out which is the 11th term we could list the number of apples in each bag in size order as we have been done before:

Number of apples: 16, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 18,.....and

|
median

As you can see this method will be long and tedious for large distribution. A better method is to use cumulative frequency distribution.

THE CUMULATIVE FREQUENCY DISTRIBUTION

Look at the cumulative frequency distribution for our example.

Number of apples in bag	Number of bags (frequency)
Less than 16	0
Less than 17	3
Less than 18	8
Less than 19	14
Less than 20	19
Less than 21	21

Note:

- The total frequency of the distribution is the same as the cumulative frequency of the last class.
- This cumulative frequency distribution give you the following information.
 - No bags contain less than 16 apples; 3 bags have less than 17 apples; 8 bags have less than 18 apples and so on.
 - The median value of this distribution is 18 apples. We know this because the median is the 11th value and this must lie in the class 'less than 19'.

FINDING THE MEDIAN OF A GROUPED CUMULATIVE FREQUENCY DISTRIBUTION

Suppose that we want to find the median height of the class of the first school children. This is the cumulative frequency distribution.

Example 4

Height (cm)	Frequency (f)
80 – 89	2
90 – 99	10
100 – 109	16
110 – 119	24
120 – 129	16
130 – 139	11
140 – 149	1
	$\Sigma f = 80$

Answer

In order to find the median we must first form the cumulative frequency distribution.

Height (cm)	Cumulative Frequency (cf)
Less than 89.5	2
Less than 99.5	12
Less than 109.5	28
Less than 119.5	52
Less than 129.5	68
Less than 139.5	79
Less than 149.5	80

Notice how the class limits of the cumulative frequency distribution are chosen. They are taken from the original distribution. The limit of each class is given by the lower boundary of the next higher class.

This is easier to understand if you look at what it means in this example.

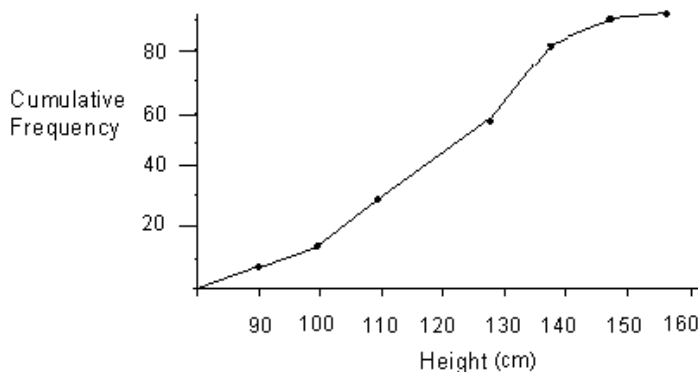
The heights are given to the nearest cm so that 80 – 89 cm really means 79.5 – 89.49 cm (in other words less than 89.5cm); 90 – 99cm means 89.5 – 99.49 and so on.

There are 80 children in the distribution so we can see that the median will come somewhere in the class 'less than 119.5 cm'. That is, somewhere between 109.5 cm and 119.5 cm. We can get a better estimate of the median than this by using the cumulative frequency curve. This is explained in the next section.

THE CUMULATIVE FREQUENCY CURVE (sometimes called an ogive)

The ogive is formed by plotting points from the cumulative frequency distribution and joining them with a smooth curve. The median is then the value of the variable which corresponds to $\frac{1}{2}(n + 1)^{\text{th}}$ value. If n is reasonably large, we can say this is approximately equal to half to the total frequency. We can find out many things about a distribution by looking at the cumulative frequency curve/ the median is one of them, we shall look at some others in the next example.

Ogive to show all the heights of 80 children



Notice that this is a 'less than' curve. Cumulative frequency curves usually have this 'flattened S' shape and that was the reason why they were called ogives (an architectural term). The horizontal scale measures the variable of the distribution. The vertical scale tells us how many are 'less than' a particular value. Some examples of the problem we could solve by using the ogive are on the following pages.

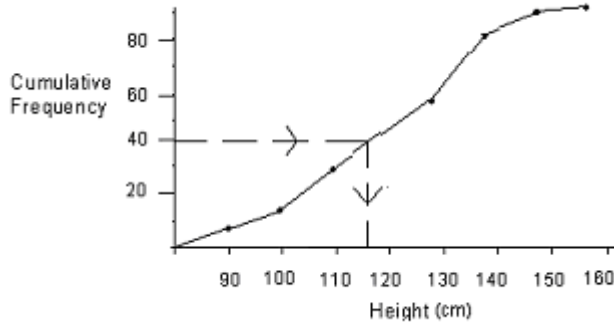
Always use graph paper when drawing cumulative frequency curves.

Example 5

Use the cumulative frequency curve to find the median of the distribution of 80 first school children.

Answer

First draw the ogive.



The median is the value which divides the distribution into two halves, that is the 40th value, and this is read from the graph.

Median = 114 cm.

In this example we have used the ogive to find the median. Other statistics can also be obtained from it and some of these are explained briefly below. The most important of these for this course are the **quartiles**.

QUARTILES

The median divides a distribution into two equal parts. The quartiles divide it into four equal parts. The second quartile is the same as the median.

Here is a distribution already arranged in size order.

1, 2, 2, 2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 7, 8
 | | |
 QL M QU

There are 19 values so the median (M) is the $\frac{19+1}{2} = 10^{\text{th}}$ value.

The lower quartile (QL) is the $\frac{1}{4} (19 + 1) = 5^{\text{th}}$ value
 The upper quartile (QU) is the $\frac{3}{4} (19 + 1) = 15^{\text{th}}$ value
 Therefore, M=5, QL=3, QU=6.

For a continuous distribution or a grouped frequency distribution it is better to find the quartiles in the same way that we found the median in example 5. That is: first form a cumulative frequency distribution and from it draw a cumulative frequency curve: then find quartiles from this graph.

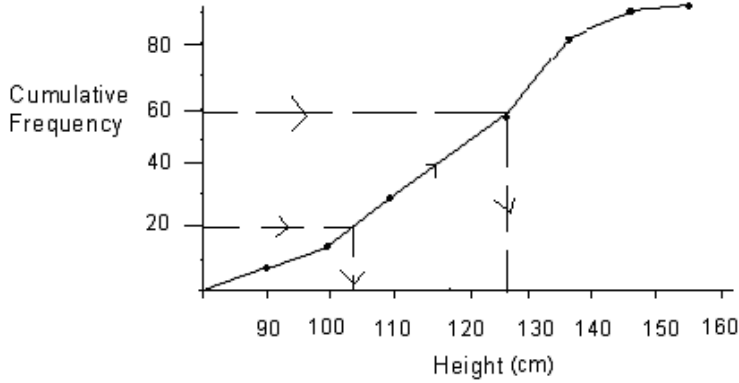
PERCENTILES AND DECILES

Quartiles divide distributions into four equal parts. Percentiles are the statistics which divide them into 100 equal parts and deciles into 10 equal parts. (The upper quartile could also be called the 75th percentile)

The collective name for quartiles, deciles and percales is **quantiles**.

Example 6

Using the ogive drawn for example 4 find the upper and lower quartile for the distribution of heights of first school children.



Answer

The lower quartile (QL) is given by the $\frac{1}{4}(80) = 20^{\text{th}}$ value
 The upper quartile (UP) is given by the $\frac{3}{4}(80) = 60^{\text{th}}$ value
 From the graph:

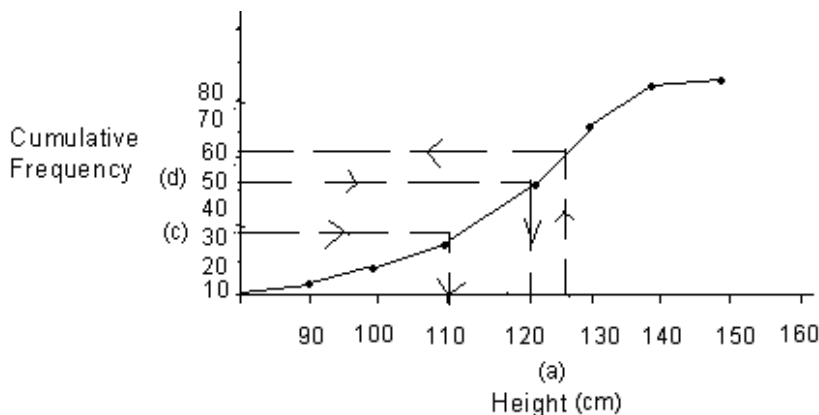
QL = 104 cm
 QU = 124 cm

Example 7

Looking again at the frequency distribution of the 80 first school children, example 4. We could use the cumulative frequency curve to answer the following questions.

- a) How many of the pupils are less than 125 cm tall?
- b) How many are greater than 125 cm?
- c) What height limit should we set if we wish to give some special vitamin supplements to the smallest 30 children?
- d) What should the minimum height be if we wanted to find the tallest 40% of the children?

Answer



The arrows show you which way to read from the graph.

All of these answers have been read from the ogive.

- a) 61
- b) The number greater than 125 cm is equal to the total minus the number less than 125 cm.
 $80 - 61 = 19$
- c) 111 cm
- d) $40\% \text{ of } 80 = \frac{40 \times 80}{100} = 32$

Therefore, 32 children will be taller than the limit set which means that $80 - 32 = 48$ will be less than this limit. We can see from the ogive that the limit should be set at 118 cm.

The **Semi – Interquartile Range, SIR**, = $\frac{\text{Upper Quartile} - \text{Lower Quartile}}{2}$

Exercise 2

1. Form the cumulative frequency distributions from the following

a)

SCORE	0	1	2	3	4	5
FREQUENCY	1	4	4	6	3	2

b)

WEIGHT (kg)	FREQUENCY
130-134	7
135-139	17
140-144	40
145-149	31
150-155	5

2. Find a) the median b) the quartiles of the following

MARK	1	2	3	4	5	6	7	8	9	10
FREQUENCY	2	1	3	5	6	8	4	3	2	1

3. The following table shows the frequency distribution of examination marks for 120 candidates.

Exam Marks	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Number of candidates	2	6	7	14	20	35	29	6	1

Construct a cumulative frequency table and use it to draw a cumulative frequency curve of the distribution. From the curve, estimate:

- a) the median mark
- b) the upper and lower quartiles
- c) If the pass mark is 45, what percentage pass the paper?
- d) If 75% of the candidates are to pass, what should the pass mark be?

4. A sample of 150 people were asked to give the number of hours they watched BBC 1 during the week. The results were as follows:

Number of hours	0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40
Frequency	2	8	17	52	38	26	5	2

Construct a cumulative frequency diagram and from it estimate:

- a) the median
- b) the percentage of people who watched 17 hours or more each week.

ANSWERS

Exercise 1

- a) Median = 13
 b) Median = $\frac{15 + 17}{2} = 16$

Exercise 2

1. a) **SCORE** **CUMULATIVE FREQUENCY**

Less than 1	1
“ 2	5
“ 3	9
“ 4	15
“ 5	18
“ 6	20

b) **SCORE** **CUMULATIVE FREQUENCY**

Less than 134.5	7
“ 139.5	24
“ 144.5	64
“ 149.5	95
“ 154.5	100

2. a) Total frequency = 35

$$\text{Median} = \frac{35 + 1}{2} = 18^{\text{th}} \text{ value}$$

$$\text{Lower Quartile} = \frac{1}{4}(35 + 1) = 9^{\text{th}} \text{ value}$$

$$\text{Upper Quartile} = 3\left(35 + \frac{1}{4}\right) = 27^{\text{th}} \text{ value}$$

Forming cumulative frequency

MARK	FREQ	CUM. FREQ
1	2	2
2	1	3
3	3	6
4	5	11
5	6	17
6	8	25 < --
7	4	29 etc.

- b) Median = 6
 Lower Quartile = 4
 Upper Quartile = 7

3. EXAM MARK CUMULATIVE FREQUENCY

Less than 20	2
" 30	8
" 40	15
" 50	29
" 60	49
" 70	84
" 80	113
" 90	119
" 100	120

From the graph:-

- a) The median mark = 63.5
- b) Lower Quartile = 50 Upper Quartile = 72.5
- c) Pass mark = 45 and 21 students score less than this.
Therefore $120 - 21 = 99$ pass $\rightarrow \frac{99}{120} \times 100 = 82 \frac{1}{2}\%$
- d) 75% are to pass $\frac{75}{100} \times 120 = 90$
Therefore 30 are to fail. Pass mark should be 50.

4. NUMBER OF HOURS CUMLATIVE FREQUENCY

Less than 6	2
" 11	10
" 16	27
" 21	79
" 26	117
" 31	143
" 36	148
" 41	150

- a) Median = 75th value = 20.5 hours
- b) 35 people watched 17 hours or less per week.
Therefore $150 - 35 = 115$ people watched 17 hours or more each week.

Expressed as a percentage,

$$\frac{115}{120} \times 120 = 95.8\%$$