Graphical Representation

Construction of Histograms when Class Size is Same

The frequency distribution table of the marks of 26 students in a particular subject is as follows.

Class interval (marks of students)	Frequency (number of students)
0 – 10	4
10 – 20	2
20 – 30	10
30 – 40	8
40 – 50	2

Can we represent this data graphically?

This data can be represented in the form of a histogram.

Let us now look at an example to understand this concept better.

Example 1:

The given tally table represents the total runs scored by 39 batsmen in 10 different test matches.

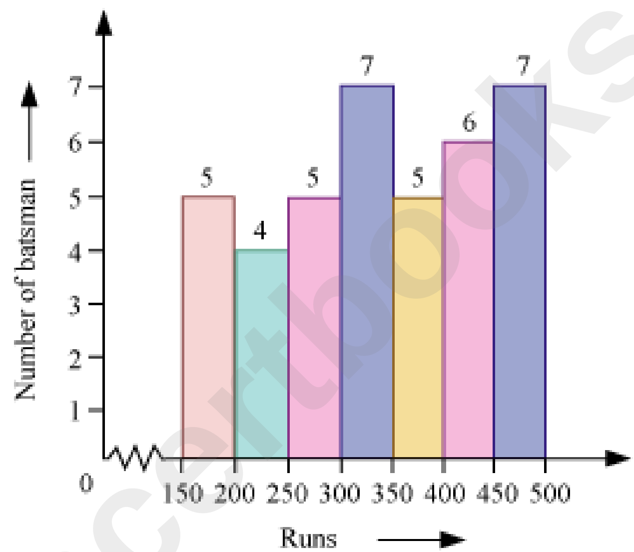
Runs	Tally marks	Frequency (Number of batsmen)
150 – 200	𝍸	5
200 – 250	||||	4
250 – 300	𝍸	5
300 – 350	𝍸 ||	7
350 – 400	𝍸	5
400 – 450	𝍸 |	6

450 – 500	卌 ||	7

Draw a histogram for the above given distribution table.

Solution:

In order to draw the histogram of the given frequency distribution table, we represent the runs on the horizontal axis and the number of batsmen on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.



Here, we will use a broken line ( ) to indicate that the values between 0 – 150 are not represented.

Example 2:

The given table represents the data related to intelligent quotient (IQ) of the students of a class.

IQ	Number of students
61 – 70	**4**
71 – 80	**3**
81 – 90	**5**
91 – 100	**8**

101 – 110	**15**
111 – 120	**12**
121 – 130	**13**
Total	**60**

Draw a histogram for the above given distribution table.

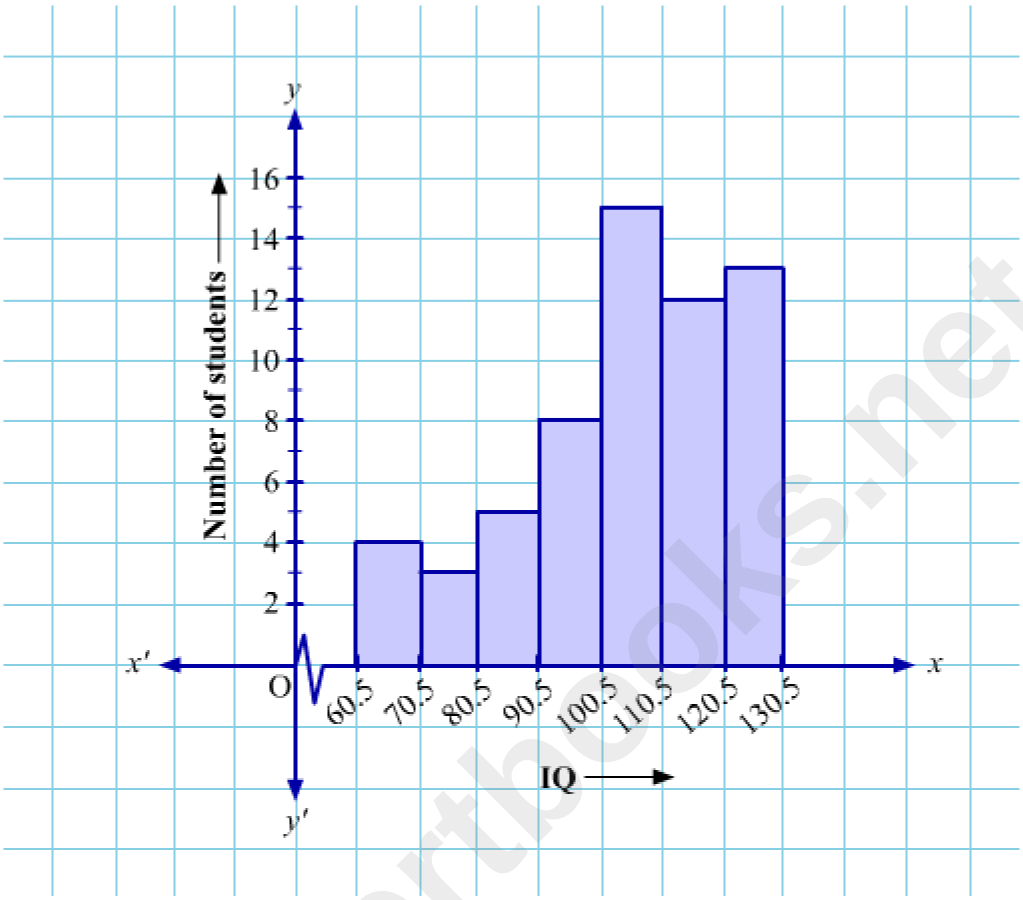
Solution:

In the given table the class intervals are of inclusive type, so we need to make them of exclusive type. Here, the difference between the upper limit of a class and lower limit of next class is 1. So, we need to subtract half of this i.e., 0.5 from lower limit and add 0.5 to upper limit of each class. Thus, we will get extended classes according to which, we can draw the histogram.

The modified table consisting extended classes is as follows:

Original Class	**Extended Class**	**Number of students (Frequency)**
61 – 70	**60.5 – 70.5**	**4**
71 – 80	**70.5 – 80.5**	**3**
81 – 90	**80.5 – 90.5**	**5**
91 – 100	**90.5 – 100.5**	**8**
101 – 110	**100.5 – 110.5**	**15**
111 – 120	**110.5 – 120.5**	**12**
121 – 130	**120.5 – 130.5**	**13**
Total		**60**

In order to draw the histogram of this frequency distribution table, we represent the IQ on the horizontal axis and the number of students on vertical axis. The height of each bar represents the frequency. The width of all the bars is same.



Here, we will use a broken line ( ) to indicate that the values between 0 – 60.5 are not represented.

Finding Mode of Data From a Histogram

We know that the mode of a given set of data is such a property, which gives us the observation that has the maximum frequency.

Consider the following frequency table.

Observation	10	20	30	40	50	60
Frequency	25	19	28	37	49	13

From the table, we can easily observe that the observation 50 has maximum frequency, which is 49.

Thus, the mode of the given data is 50.

Thus, the mode can be defined as:

“The observation which occurs the maximum number of times is called mode”.

Or, “The observation with maximum frequency is called mode”.

When the data is given in grouped form, we cannot find the mode by looking at the frequencies. Let us consider the following example.

Age of People	Number of People
0 − 10	4
10 − 20	2
20 − 30	3
30 − 40	6
40 − 50	5
50 − 60	19
60 − 70	25
70 − 80	22
80 − 90	10
90 − 100	4

From the above table, we can observe that the maximum frequency is 25 and it corresponds to the class interval of 60 − 70. However, we are not sure that exactly which value in the class interval 60 − 70 is the mode because we do not know the actual age of these 25 people.

In such situations, we find the mode of the data geometrically by drawing a histogram. For this, we have to follow the below given steps.

In order to understand this concept better, let us look at some more examples.

Example 1:

The following table shows the class intervals and the frequency corresponding to them.

Class Interval	**40 − 69**	**70 − 99**	**100 − 129**	**130 − 159**	**160 − 189**
Frequency	**12**	**15**	**24**	**16**	**21**

Find the mode of the given data geometrically.

Solution:

The given frequency distribution is discontinuous. To convert it into continuous distribution, we have to subtract $\frac{1}{2} = 0.5$ from the lower limits and add 0.5 to higher limits of each class interval.

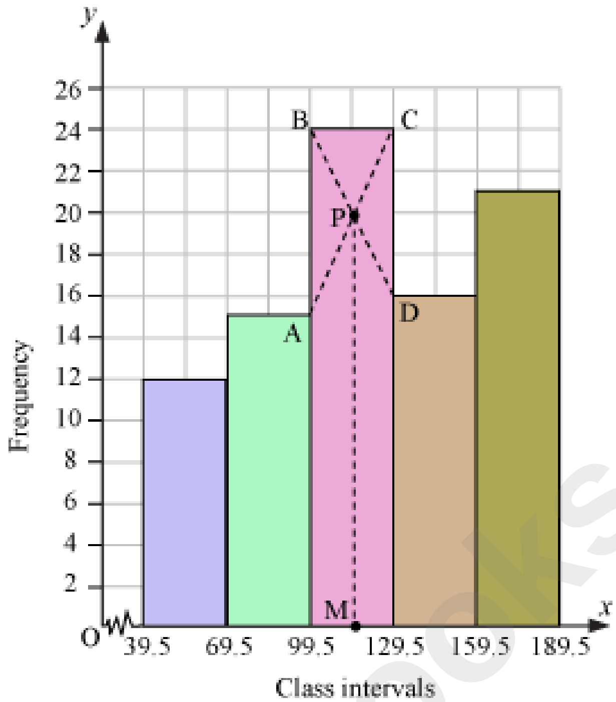
Now, the continuous frequency distribution table of the given data is as follows:

Class Interval	**Frequency**
39.5 − 69.5	**12**
69.5 − 99.5	**15**
99.5 − 129.5	**24**
129.5 − 159.5	**16**
159.5 − 189.5	**21**

To find the mode of the above data geometrically, first of all we have to draw its histogram by choosing 1 cm along *x*-axis = 30 (class-intervals) and 1 cm along *y*-axis = 2 (frequencies).

In the highest rectangle (class interval 99.5 − 129.5), we will draw two straight lines AC and BD from corners of the rectangles on either side of the highest rectangle to the opposite corners of the highest rectangle.

Let P be the intersection of the lines AC and BD. Now, we will draw a vertical line through the point P that cuts the *x*-axis at M.



The point M represents the value 115 (approximately) on *x*-axis. Therefore, the mode of the given data is 115 (approximately).

Example 2:

The following table shows the class intervals and the frequency corresponding to them.

Class Interval	**0 − 40**	**40 − 80**	**80 − 120**	**120 − 160**	**160 − 200**
Frequency	**15**	**35**	**22.5**	**10**	**30**

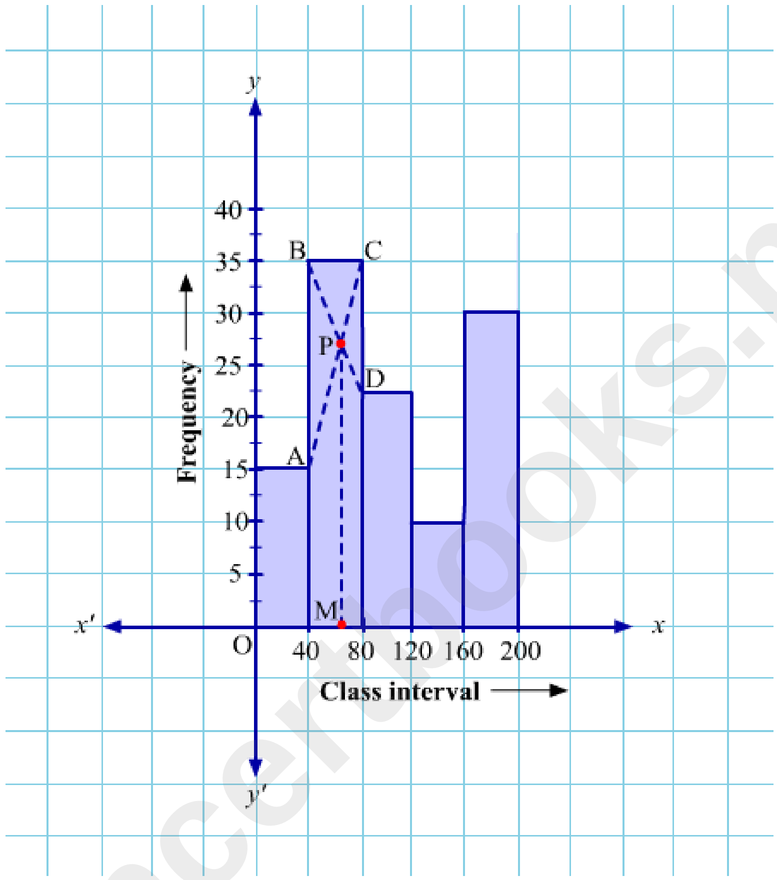
Find the mode of the given data geometrically.

Solution:

To find the mode of the given data geometrically, first of all we have to draw its histogram by choosing 1 unit along *x*-axis = 40 (class-intervals) and 1 unit along *y*-axis = 5 (frequencies).

In the highest rectangle (class interval 40 − 80), we will draw two straight lines AC and

BD from corners of the rectangles on either side of the highest rectangle to the opposite corners of the highest rectangle. Let P be the intersection of the lines AC and BD. Now, we will draw a vertical line through the point P that cuts the *x*-axis at M.



The point M represents the value 60 (approximately) on *x*-axis. Therefore, the mode of the given data is 60 (approximately).

Median Of A Grouped Data By Constructing Ogive

A pictorial representation always gives a better understanding than a written statement. A graphical representation helps us in understanding of a given data in an easier and detailed manner. In order to understand the median of a grouped data properly, we have to draw an ogive.

Firstly, let us discuss what is an ogive?

"**When the data is given as 'less than' or 'more than' type and a graph is plotted between either of the limits and the cumulative frequency, the smooth curve so obtained is known as ogive or cumulative frequency curve".**

Let us discuss with an example that how an ogive is helpful to find out the median of a grouped data.

Let us consider that 80 students of a class appeared in a Geography test. The marks obtained (out of 100) by them are given in the following frequency distribution table.

Table - 1

Marks obtained (out of 100)	Number of students
0 − 10	2
10 − 20	1
20 − 30	3
30 − 40	5
40 − 50	9
50 − 60	15
60 − 70	22
70 − 80	12
80 − 90	7
90 − 100	4

We can write the above table in following two ways.

1. Less than type
2. More than type

1. For less than type

We can write the given table in less than type as follows.

Marks obtained (out of 100)	Number of students (Cumulative frequency)
Less than 10	2
Less than 20	3
Less than 30	6
Less than 40	11
Less than 50	20
Less than 60	35
Less than 70	57

Less than 80	69
Less than 90	76
Less than 100	80

Construction of Ogive of less than type

The smooth curve drawn between the **upper limits of class intervals and cumulative frequency** is called cumulative frequency curve or ogive (of less than type). The upper limits of the intervals and cumulative frequency are shown in the above table.

The method of drawing an ogive of less than type is as follows.

(1) Firstly we draw two perpendicular lines, one is horizontal (x-axis) and the other is vertical (y-axis), on a graph.

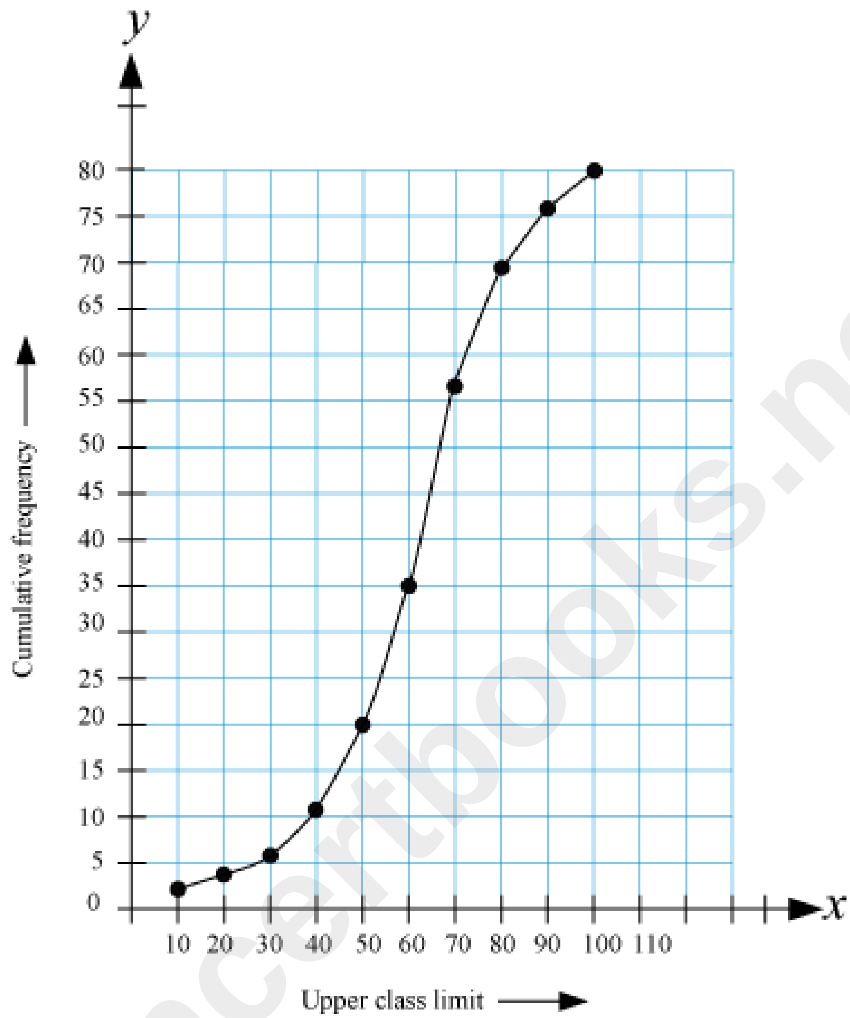
(2) Now, we mark the upper limits on the horizontal line and the cumulative

frequencies on the vertical line by taking suitable scale.

(3) After this, we plot the points (10, 2), (20, 3), (30, 6), (40, 11), (50, 20), (60, 35), (70, 57), (80, 69), (90, 76), (100, 80). These are the points corresponding to the upper limit and the cumulative frequency.

(4) Now, we join these points to obtain a smooth curve.

After following these steps, we obtain the following graph.



The smooth curve obtained in this figure is the ogive of less than type of the given data.

To find the median with the help of ogive

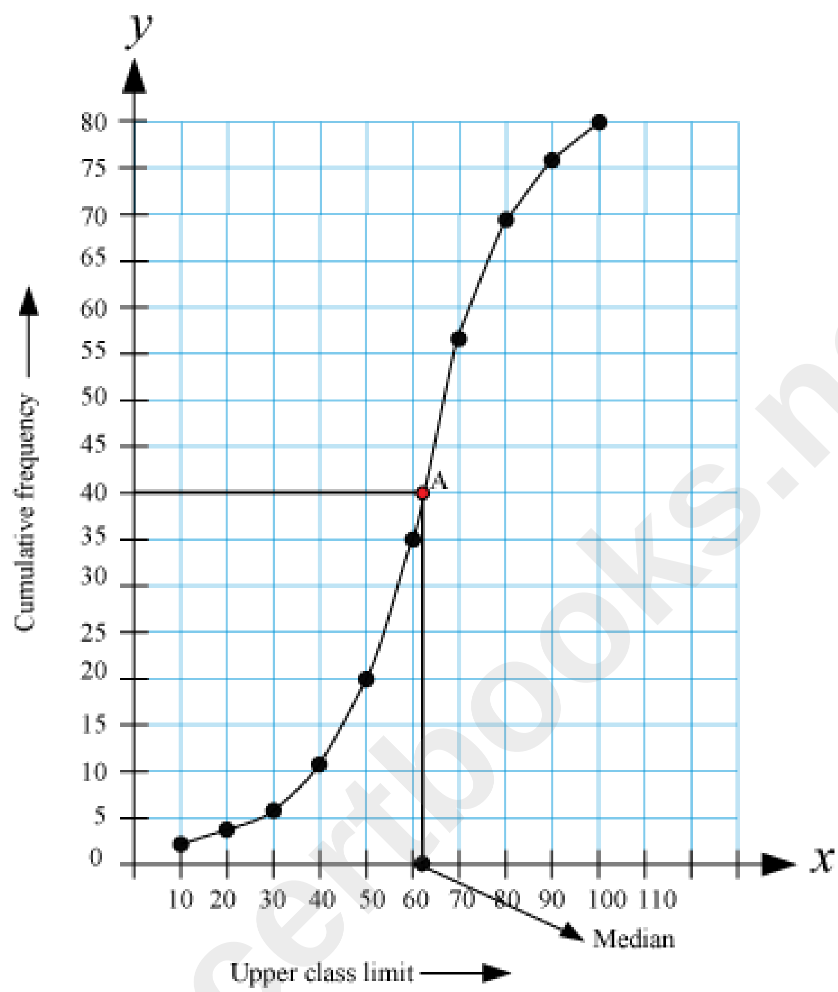
An important application of ogive in statistics is to find the median.

Let us see the method of finding the median through the same example.

In the previous example, number of observations, n = 80

$$\therefore \frac{n}{2} = 40$$

Mark the point 40 on the vertical line and then draw a horizontal line through this point. Let this horizontal line intersect the ogive at point A. Now, draw a vertical line through A. Median is the point at which this vertical line intersects the horizontal line.



In this example, the median is 62 (approximately).

2. For more than type

We can write the given table for more than type as follows.

Marks obtained	Number of students (Cumulative frequency)
More than or equal to 0	80
More than or equal to 10	78
More than or equal to 20	77
More than or equal to 30	74

More than or equal to 40	69
More than or equal to 50	60
More than or equal to 60	45
More than or equal to 70	23
More than or equal to 80	11
More than or equal to 90	4

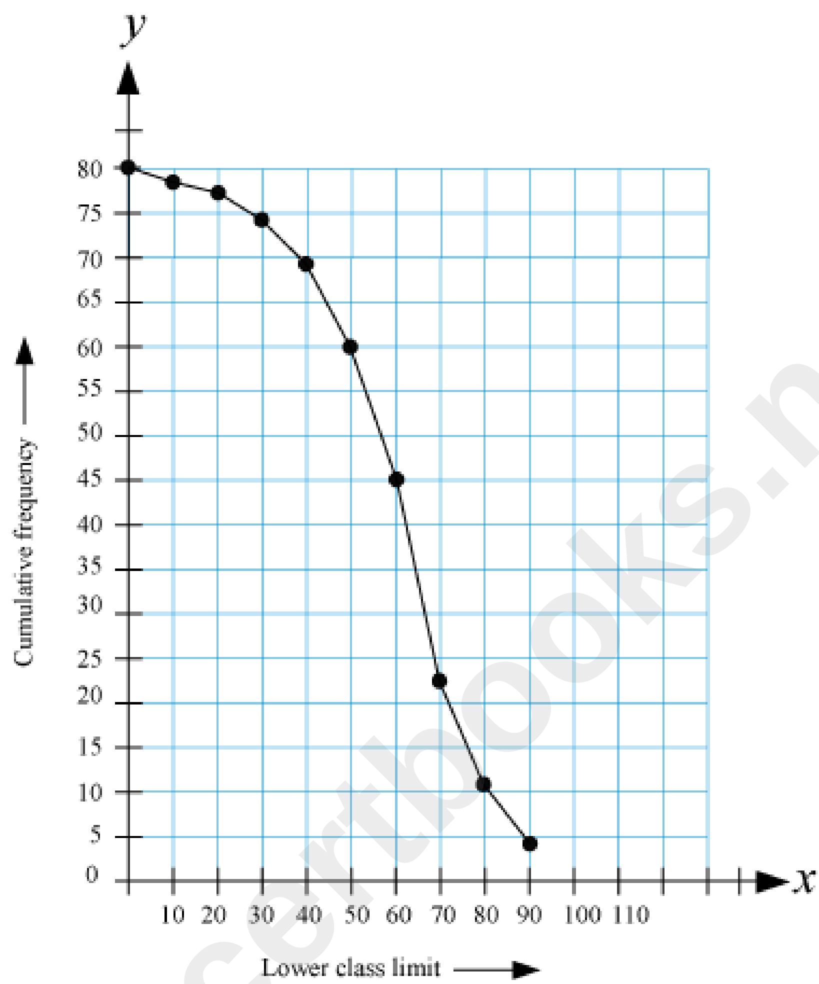
Construction of Ogive of more than type

The smooth curve drawn between **the lower limits of class intervals and cumulative frequencies** is called cumulative frequency curve or ogive (for more than type).

The method of construction of more than type ogive is same as the construction of less than type. For **more than type** ogive, we take the lower limits on *x*-axis and the cumulative frequencies on the *y*-axis. In the above table, the lower limits and the cumulative frequencies have been represented.

The ogive of more than type is obtained by plotting the points (0, 80), (10, 78), (20, 77), (30, 74), (40, 69), (50, 60), (60, 45), (70, 23), (80, 11), (90, 4).

The ogive of more than type of the previous table has been shown below.

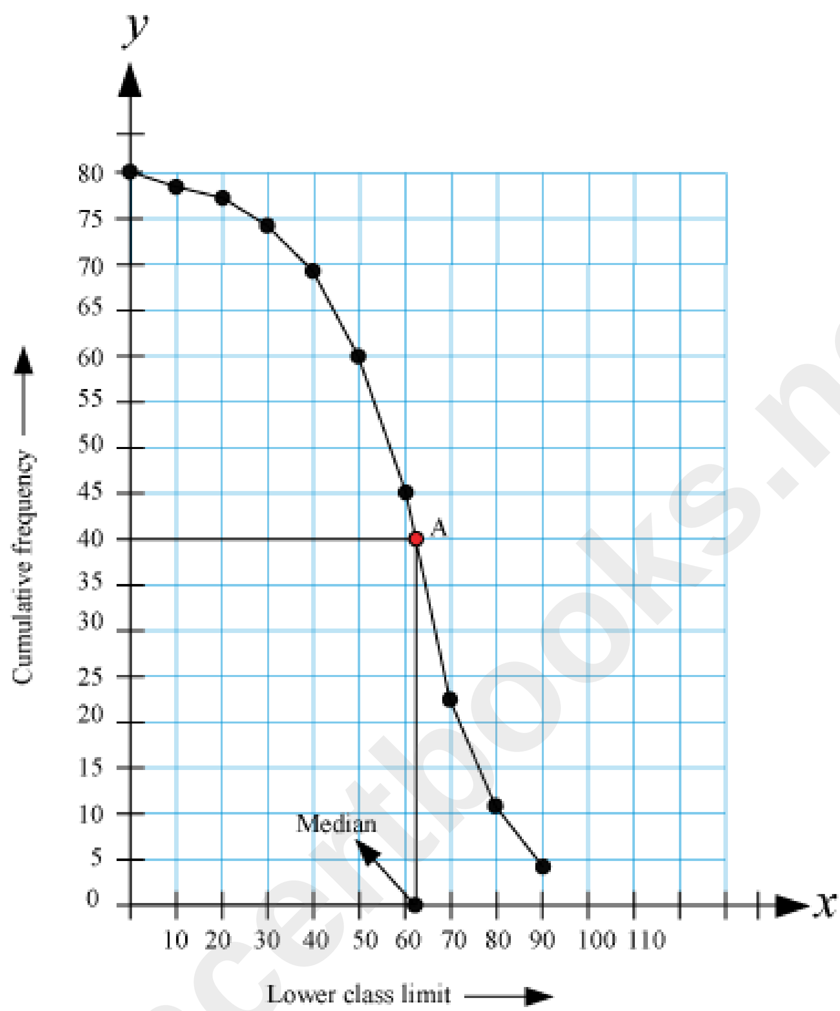


To find the median with the help of ogive

Number of observations, n = 80

$$\therefore \frac{n}{2} = 40$$

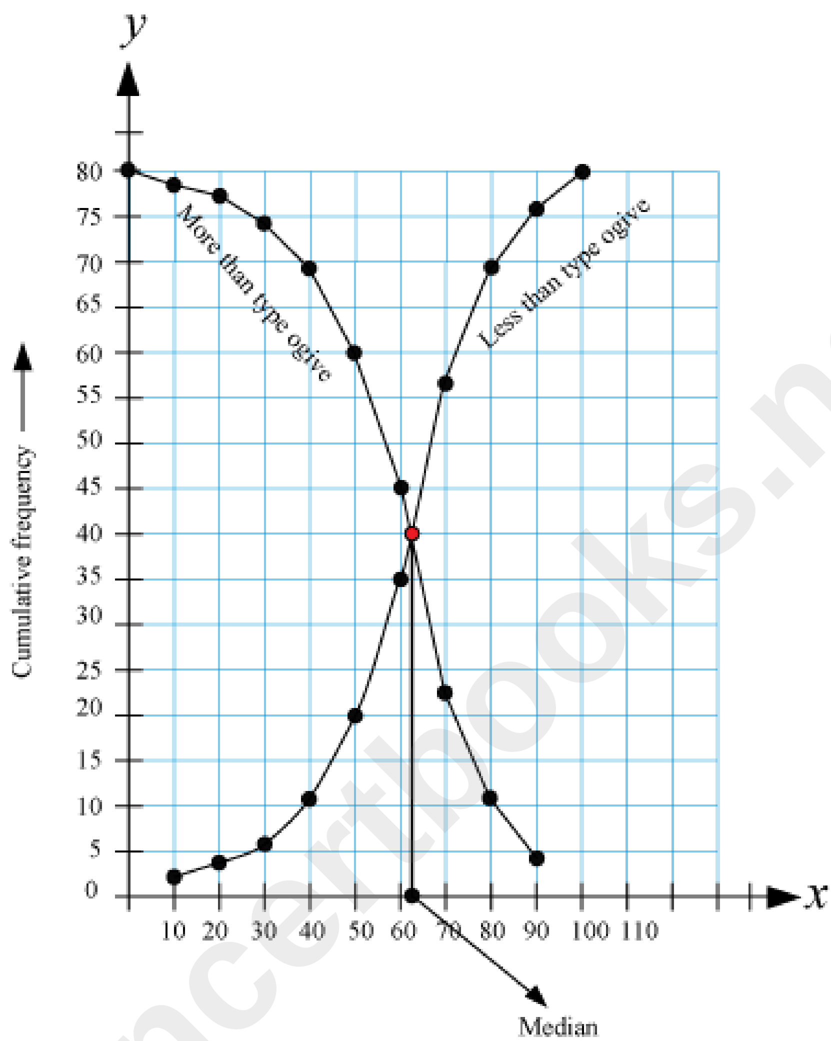
Mark the point 40 on the vertical line and then draw a horizontal line from this point. Let this line intersect the ogive at point A. Now, draw a vertical line through A. Median is the point at which this vertical line intersects the horizontal line.



Hence, the median is 62(approximately).

Relation between the ogive of less than type and more than type

Let us draw the graph of both less than type and more than type on the same graph paper.



We observe from the above graph that,

"**If we construct a line parallel to *y*-axis through the point of intersection of both the ogives i.e., of less than type and more than type, then the point at which this line intersects *x*-axis represents the median of the given data".**

Let us solve some more examples to understand the concept better.

Example 1:

The following distribution table gives the daily wages of 50 workers in a factory.

Wage (in Rs)	Number of workers
50 − 100	5

100 − 150	25
150 − 200	10
200 − 250	7
250 − 300	3

Convert it into a less than type distribution and draw its ogive.

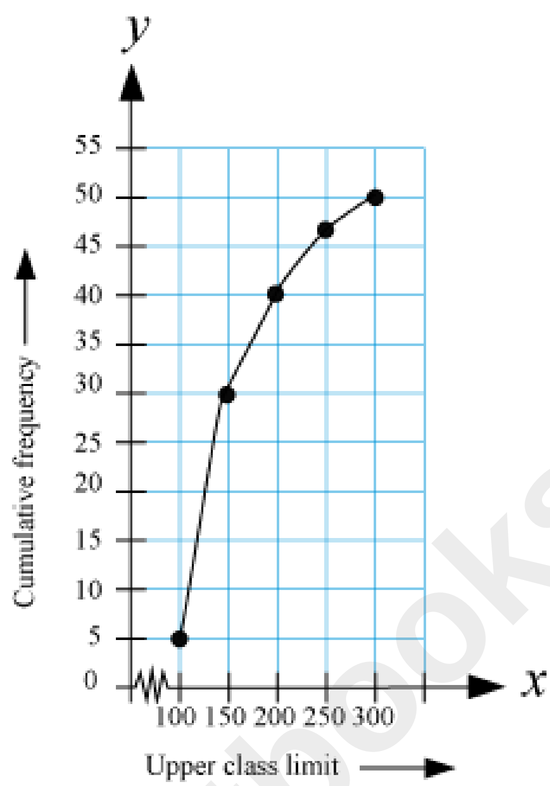
Solution:

We can write the given distribution table as a less than type distribution as follows.

Wage (in Rs)	Number of workers (cumulative frequency)
Less than 100	5
Less than 150	5 + 25 = 30
Less than 200	30 + 10 = 40
Less than 250	40 + 7 = 47
Less than 300	47 + 3 = 50

To obtain the ogive, we have to plot the points (100, 5), (150, 30), (200, 40), (250, 47), (300, 50) on a graph paper taking the upper limits on *x*-axis and the cumulative frequencies on *y*-axis.

The ogive obtained has been represented in the following figure.

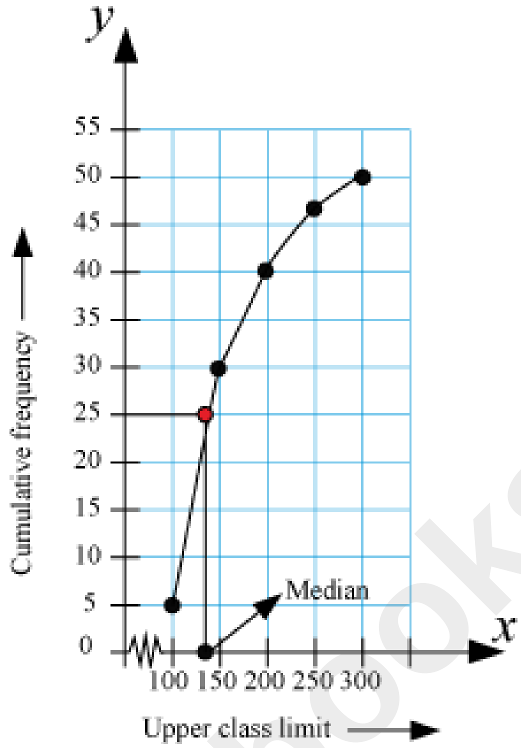


Here, n = 50

$\therefore \frac{n}{2}=25$

First, we mark the point 25 on the vertical line and then draw a horizontal line through this point. Let this horizontal line intersect the ogive at point A.

Now, draw a vertical line through point A. Median is the point where this vertical line intersects the x-axis. In this case, the value comes out to be 140. Hence, 140 is the median of the given distribution table.



Example 2:

The following table shows the total monthly household expenditure of 100 families in a city.

Expenditure	Number of families
0 – 1000	**22**
1000 – 2000	**30**
2000 – 3000	**24**
3000 – 4000	**14**
4000 – 5000	**10**

Convert it into a more than type of distribution and draw its ogive.

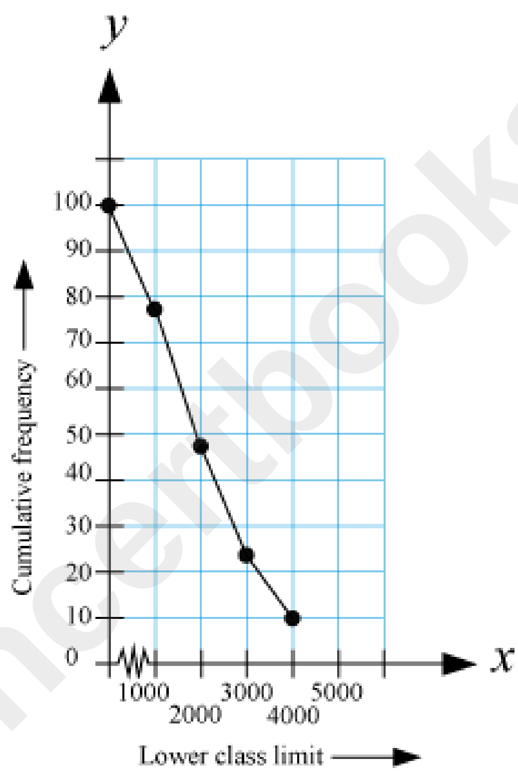
Solution:

We can write the given table in more than type as follows.

Expenditure (in Rs)	Number of families (cumulative frequency)
More than or equal to 0	100

More than or equal to 1000	100 − 22 = 78
More than or equal to 2000	78 − 30 = 48
More than or equal to 3000	48 − 24 = 24
More than or equal to 4000	24 − 14 = 10

Now, we plot the points (0, 100), (1000, 78), (2000, 48), (3000, 24), (4000, 10) on a graph paper to obtain the ogive of more than type distribution.



Example 3:

The following table shows the rainfall (in mm) in 60 cities on a particular day.

Rainfall (in mm)	Number of cities
25 − 29	16
30 − 34	12
35 − 39	10

40 − 44	14
45 − 49	8

Draw both the ogives for this data on the same graph paper and find the median.

Solution:

In the given table, the class intervals are of inclusive type, so we need to make them of exclusive type first. Now, a common table can be prepared for less than type and more than type cumulative frequencies to draw both the ogives on the same graph paper.

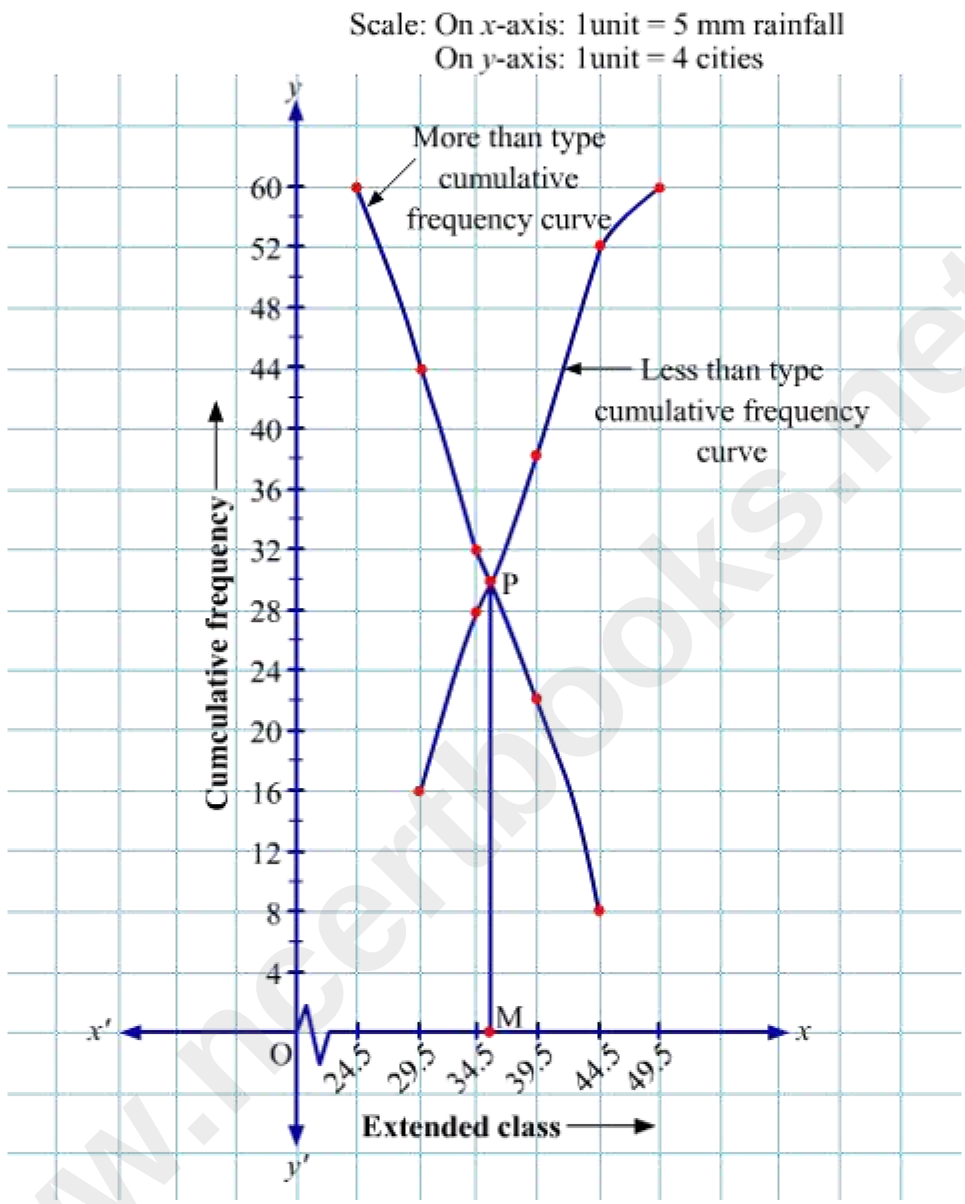
The table is given below.

Original classes	**Extended classes**	**Frequencies**	**Less than type**	**c.f. (Less than type)**	**More than type**	**c.f. (More than type)**
25 − 29	24.5 − 29.5	16	Less than 29.5	16	More than 24.5	60
30 − 34	29.5 − 34.5	12	Less than 34.5	28	More than 29.5	44
35 − 39	34.5 − 39.5	10	Less than 39.5	38	More than 34.5	32
40 − 44	39.5 − 44.5	14	Less than 44.5	52	More than 39.5	22
45 − 49	44.5 − 49.5	8	Less than 49.5	60	More than 44.5	8

Now, we plot the points (29.5, 16), (34.5, 28), (39.5, 38), (44.5, 52), (49.5, 60) on a graph paper to obtain the ogive of less than type distribution.

Also, we plot the points (24.5, 60), (29.5, 44), (34.5, 32), (39.5, 22), (44.5, 8) on the same graph paper to obtain the ogive of more than type distribution.

The graph is shown below:



Point M represents the median which is approximately 36.

Upper Quartiles and Lower Quartiles of a Grouped Data

We know that the median divides a given data into two equal parts. Here, we will study about a concept, which divides a given data into four equal parts.

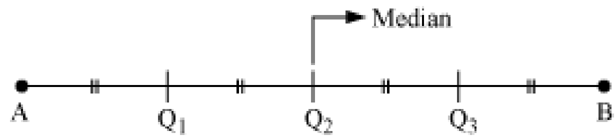
The values of a statistical data which divide the whole set of observations (variate) into four equal parts are known as quartiles.

We will now understand how to find the quartiles of a given data.

In order to find quartiles, firstly, the data is arranged in ascending order of their magnitudes.

Let us suppose they are arranged along a line segment AB.

The points Q_1, Q_2, and Q_3 divide AB into four equal parts.



Now, observe that Q_2 divides AB into 2 equal parts. Therefore, we can conclude that Q_2 is the median for the data. Q_2 is also known as **middle quartile**. The values Q_1 and Q_3 are respectively called the **lower quartile** and the **upper quartile**.

Let us study about them one by one in more detail.

Lower quartile (Q_1):

When the lower half, before the median, is divided into two equal parts, the value of the dividing variate is called lower quartile.

The formula for calculating the lower quartile is as follows:

$$Q_1 = \begin{cases} \dfrac{n+1}{4}\text{th observation, if } n \text{ is odd} \\ \dfrac{n}{4}\text{th observation, if } n \text{ is even} \end{cases}$$

Upper quartile (Q_3):

When the upper half, after the median, is divided into two equal parts, the value of the dividing variate is called upper quartile.

The formula for calculating the upper quartile is as follows:

$$Q_3 = \begin{cases} \dfrac{3(n+1)}{4}\text{th observation, if } n \text{ is odd} \\ \dfrac{3n}{4}\text{th observation, if } n \text{ is even} \end{cases}$$

Let us now find Q_1, Q_2, and Q_3 for the following data:

11, 4, 5, 1, 9, 6, 10

Here, number of observations, n = 7 (odd)

The given data can be arranged in ascending order as:

1, 4, 5, 6, 9, 10, 11

$\therefore$ Q_2, median = 6 (middle observation)

$$Q_1 = \left(\frac{n+1}{4}\right)\text{th observation}$$

= 2^{nd} observation

= 4

$$Q_3 = \frac{3(n+1)}{4}\text{th observation}$$

= 6^{th} observation

= 10

Now, what is $Q_3 - Q_1$?

We have $Q_3 - Q_1 = 10 - 4 = 6$

In this case, we say that the inter-quartile range of the data is 6.

The difference between the upper quartile (Q_3) and the lower quartile (Q_1) is called the inter-quartile range and $\frac{Q_3 - Q_1}{2}$ is the semi-quartile range.

Now, we can observe that $Q_3 > Q_1$ always. Therefore, $Q_3 - Q_1 > 0$

Thus, we can say that inter-quartile range is always positive.

Now that we have understood how to find the quartiles when an ungrouped data is given, let us now understand how to find the quartiles, when the data is given in the form of frequency distribution table.

Let us consider the following frequency distribution table:

Variate	120	130	140	150	160	170	180
Frequency	2	4	6	8	15	3	2

Here, we will calculate the cumulative frequency first. Therefore, let us find that out.

Variate	120	130	140	150	160	170	180
Frequency	2	4	6	8	15	3	2
Cumulative frequency	2	6	12	20	35	38	40

Here, n = 40, which is even.

$$\text{Median} = \frac{\left(\frac{n}{2}\right)\text{th observation} + \left(\frac{n}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{20\text{th observation} + 21\text{st observation}}{2}$$

$$= \frac{150+160}{2} = \frac{310}{2} = 155$$

Lower quartile (Q_1) = $\frac{n}{4}$th observation

= 10th observation

= 140

Upper quartile (Q_3) = $\frac{3n}{4}$th observation

= 30th observation

= 160

In this way, we calculate quartiles when data is given in a frequency distribution table.

Now, what if we are given data in the form of grouped frequency distribution table? How can we then calculate the three quartiles?

Let us now understand how to calculate the quartiles of a data given in grouped frequency distribution table.

We know that a pictorial representation always gives a better understanding than a written statement. A graphical representation helps us in understanding of a given data in an easier and detailed manner. In order to understand the quartiles of a grouped data properly, we have to draw an ogive.

Let us solve some more examples to understand the concept better.

Example 1:

The following distribution table gives the daily wages of 50 workers in a factory.

Wage (in Rs)	Number of workers
50 − 100	5
100 − 150	25
150 − 200	10
200 − 250	7
250 − 300	3

Convert it into a less than type distribution and draw its ogive. From the ogive, find the median, upper quartile, and lower quartile.

Solution:

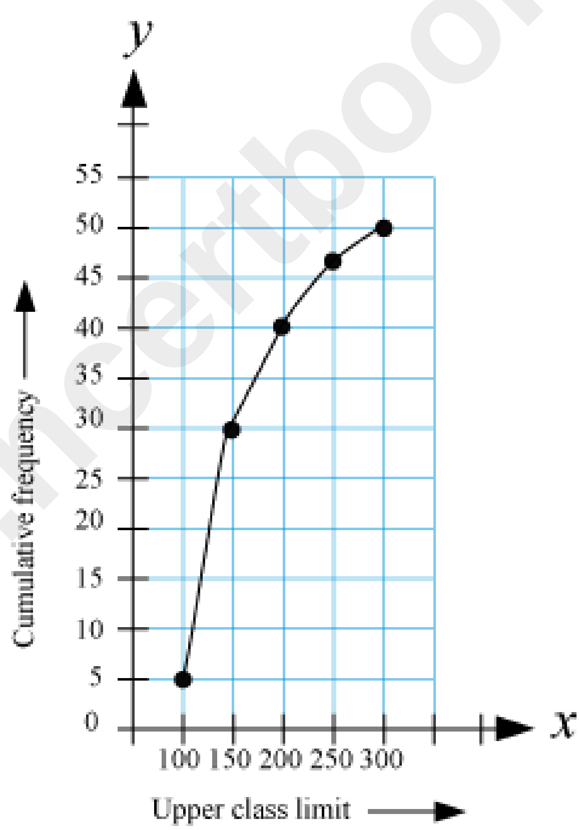
We can write the given distribution table as a less than type distribution as follows.

Wage (in Rs)	Number of workers (cumulative frequency)
Less than 100	5

Less than 150	5 + 25 = 30
Less than 200	30 + 10 = 40
Less than 250	40 + 7 = 47
Less than 300	47 + 3 = 50

To obtain the ogive, we have to plot the points (100, 5), (150, 30), (200, 40), (250, 47), (300, 50) on a graph paper, taking the upper limits on *x*-axis and the cumulative frequencies on *y*-axis.

The ogive obtained is represented in the following figure.



Here, $n = 50$

Therefore, we have

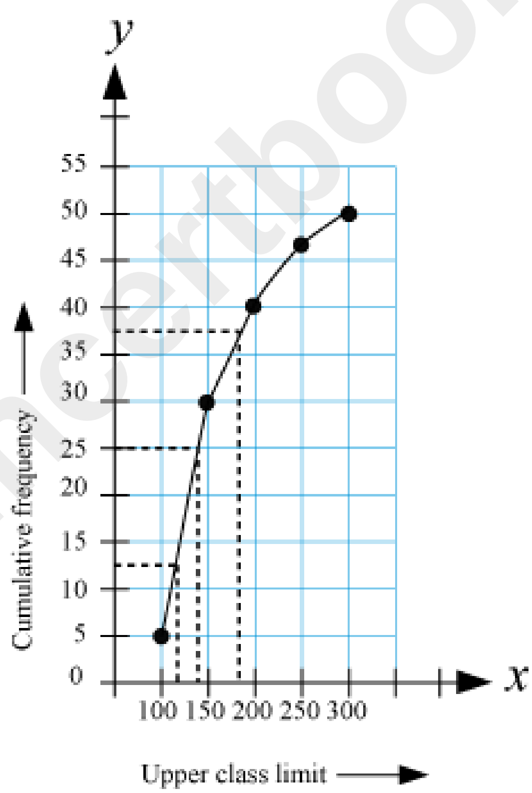
$$\text{Median} = \frac{n}{2}^{\text{th}} \text{ term} = 25^{\text{th}} \text{ term}$$

$$\text{Upper quartile} = \frac{3n}{4}^{\text{th}} \text{ term} = 37.5^{\text{th}} \text{ term}$$

$$\text{Lower quartile} = \frac{n}{4}^{\text{th}} \text{ term} = 12.5^{\text{th}} \text{ term}$$

Now, first we mark the point 25 on the vertical line and then draw a horizontal line through this point. Let this horizontal line intersect the ogive at a point. Now, draw a vertical line through this point. Median is the point where this vertical line intersects the *x*-axis. In this case, the value comes out to be 140. Hence, 140 is the median of the given distribution table.

Similarly, we find the lower and upper quartiles.



We have:

Upper quartile = 190 (approximately)

Median = 140 (approximately)

Lower quartile = 120 (approximately)

Example 2:

The following table shows the total monthly household expenditure of 100 families in a city.

Expenditure	Number of families
0 − 1000	22
1000 − 2000	30
2000 − 3000	24
3000 − 4000	14
4000 − 5000	10

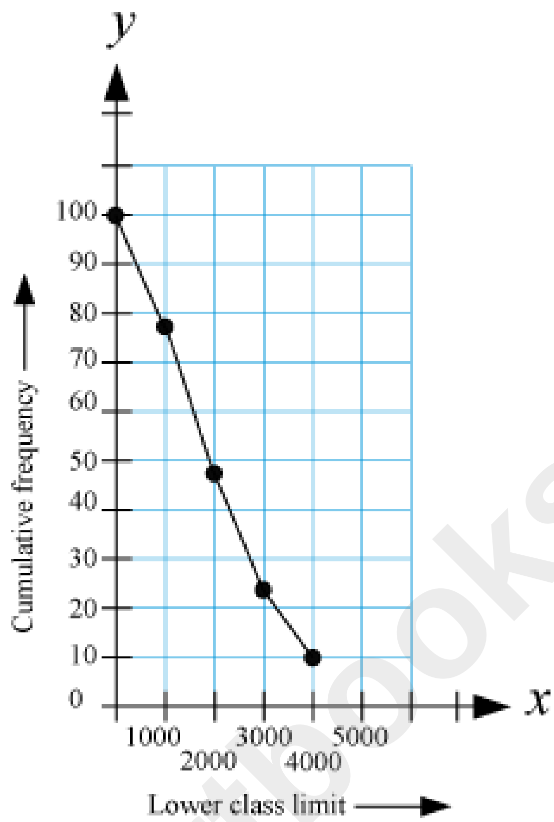
Convert it into a more than type of distribution and draw its ogive. From the ogive, find the median, upper quartile, and lower quartile.

Solution:

We can write the given table in more than type as follows.

Expenditure (in Rs)	Number of families (cumulative frequency)
More than or equal to 0	100
More than or equal to 1000	100 − 22 = 78
More than or equal to 2000	78 − 30 = 48
More than or equal to 3000	48 − 24 = 24
More than or equal to 4000	24 − 14 = 10

Now, we plot the points (0, 100), (1000, 78), (2000, 48), (3000, 24), (4000, 10) on a graph paper to obtain the ogive of more than type distribution.



Here, n = 100

Therefore, we have

$$\text{Median} = \frac{n}{2}^{\text{th}} \text{ term} = 50^{\text{th}} \text{ term}$$

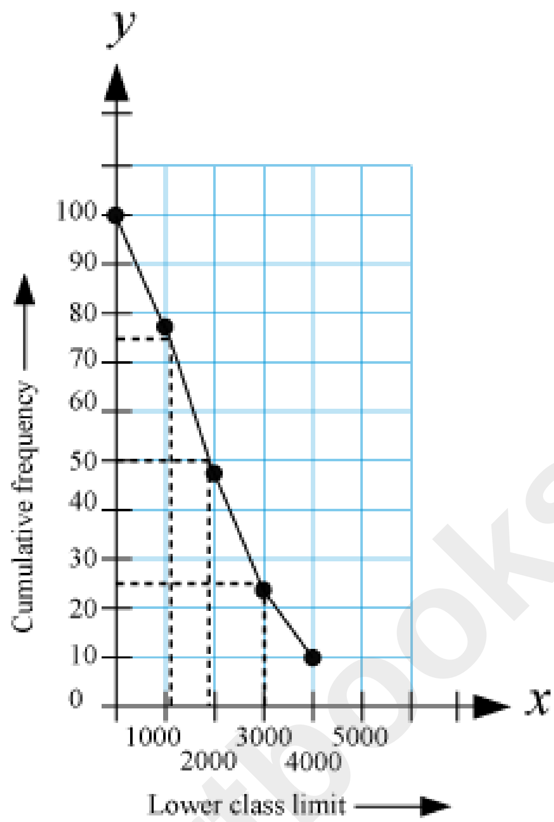
$$\text{Upper quartile} = \frac{3n}{4}^{\text{th}} \text{ term} = 75^{\text{th}} \text{ term}$$

$$\text{Lower quartile} = \frac{n}{4}^{\text{th}} \text{ term} = 25^{\text{th}} \text{ term}$$

Now, first we mark the point 50 on the vertical line and then draw a horizontal line through this point. Let this horizontal line intersect the ogive at a point. Now, draw a vertical line through this point.

Median is the point where this vertical line intersects the x-axis. In this case, the value comes out to be 1900. Hence, 1900 is the median of the given distribution table.

Similarly, we find the lower and upper quartiles.



Thus, from the graph, we find that:

Upper quartile = 3000 (approximately)

Median = 1900 (approximately)

Lower quartile = 1100 (approximately)

Example 3:

The following data shows the runs scored by a batsman in 8 different matches.

50, 98, 105, 94, 76, 60, 86, 55

For this data, find (i) lower quartile (ii) upper quartile (iii) inter-quartile range

Solution:

On arranging the given data in ascending order, we obtain

50, 55, 60, 76, 86, 94, 98, 105

Here, number of observations, $n = 8$ (even)

(i)

According to the definition, lower quartile is the middle observation of the data before median i.e., 50, 55, 60, 76.

∴ Lower quartile $= \frac{55+60}{2} = \frac{115}{2} = 57.5$

(ii)

According to the definition, upper quartile is the middle observation of the data after median i.e., 86, 94, 98, 105.

∴ Upper quartile $= \frac{94+98}{2} = \frac{192}{2} = 96$

(iii)

Inter-quartile range = $Q_3 - Q_1$

= 96 − 57.5

= 38.5