Statistics

Mean Of Grouped Data Using The Direct Method

A cricket player played 7 matches and scored 88, 72, 90, 94, 56, 62, and 76 runs. What is the average score or mean score of the player in these 7 matches?

We know that,

"The mean is the sum of all observations divided by the number of observations".

We can use the following formula to find the mean of the given data.

$$\text{Mean} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

Using this formula, the mean of scores obtained by the player

$$= \frac{\text{Sum of scores}}{\text{Number of scores}} = \frac{88+72+90+94+56+62+76}{7} = \frac{538}{7} = 76.86$$

Thus, the average score of the player is 76.86. There are many situations where the number of observations is very large and then it is not possible to use this formula to find the mean of the given data.

Let us consider such a situation.

We have the marks distribution of 100 students of a class in a Mathematics test. The distribution is given in the form of groups like 0 – 10, 10 – 20 .... To find the mean in such situations, we follow another method.

Example 1: The following table shows the various income brackets of the employees of a company.

Income of employees (in Rupees)	10000 – 15000	15000 – 20000	20000 – 25000	25000 – 30000	30000 – 35000	35000 – 40000	40000 – 45000
No. of employees	32	18	13	7	14	5	11

Find the average monthly income of each employee.

Solution:

Firstly, we will calculate the class marks of each class interval and then the product of the class marks with the corresponding frequencies. By doing so, we obtain the following table.

Income of employees (in Rupees)	No. of employees: f_i	Class Mark: x_i	$x_i \times f_i$
10000 – 15000	32	12500	400000
15000 – 20000	18	17500	315000
20000 – 25000	13	22500	292500
25000 – 30000	7	27500	192500
30000 – 35000	14	32500	455000
35000 – 40000	5	37500	187500
40000 – 45000	11	42500	467500
Total (∑)	100		2310000

The mean can now be easily calculated using the formula.

Mean $(\bar{x}) = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{2310000}{100} = 23100$

Hence, the average income of the employees is Rs 23100.

Example 2: The mean of the following frequency distribution is 62.8 and the sum of all the frequencies is 50. Find the value of f_1 and f_2.

Class Interval	**Frequency**
0 – 20	5
20 – 40	f_1
40 – 60	10
60 – 80	f_2
80 – 100	7
100 – 120	8
Total	50

Solution:

Firstly, let us find the class marks of each class interval and then the product of class marks with the corresponding frequencies for each class interval as shown in the following table.

Class Interval	**Frequency: f_i**	**Class Mark: x_i**	**$x_i \times f_i$**
0 – 20	5	10	50
20 – 40	f_1	30	30 f_1
40 – 60	10	50	500
60 – 80	f_2	70	70 f_2
80 – 100	7	90	630
100 – 120	8	110	880
	$30 + f_1 + f_2 = 50$		**$2060 + 30 f_1 + 70 f_2$**

It is given that the sum of the frequencies is 50.

$\therefore 30 + f_1 + f_2 = 50$

$f_1 + f_2 = 20$

$f_1 = 20 - f_2$ ... (1)

The mean is given as 62.8.

$\therefore (\bar{x}) = \dfrac{\sum f_i x_i}{\sum f_i} = 62.8$

$\dfrac{2060 + 30 f_1 + 70 f_2}{50} = 62.8$

$2060 + 30 f_1 + 70 f_2 = 62.8 \times 50 = 3140$

$30 f_1 + 70 f_2 = 3140 - 2060 = 1080$

$30 f_1 + 70 f_2 = 1080$

Using equation (1),

$30(20 - f_2) + 70 f_2 = 1080$

$600 + 40 f_2 = 1080$

$40 f_2 = 1080 - 600 = 480$

$f_2 = \dfrac{480}{40} = 12$

On putting the value of f_2 in equation (1), we obtain

$f_1 = 20 - f_2 = 20 - 12 = 8$

Thus, the values of f_1 and f_2 are 8 and 12 respectively.

Mean Of Grouped Data Using Assumed Mean Method

We know that direct method can be used to find the mean of any data given in grouped form, but the calculation in direct method becomes very tough when the data is given in the form of large numbers, because finding the product of x_i and f_i becomes difficult and time consuming.

Therefore, we introduce **assumed mean method** to find the mean of grouped data. This method is also known as **shift of origin method**. This is an easier method to find the mean as it involves less calculation.

Example 1: The following table shows the life time (in hours) of 20 bulbs manufactured by a reputed company.

Life time (in hours)	Number of bulbs
1000 – 2000	2
2000 – 3000	3
3000 – 4000	6
4000 – 5000	5
5000 – 6000	3
6000 – 7000	1

Find the average life time of a bulb in hours.

Solution:

Firstly, we calculate the class mark of each class-interval and assume one of them as the assumed mean. Here, we take $a = 3500$. The calculations of the deviations and the product of the deviations with the corresponding frequencies have been represented in the following table.

Life time (in hours)	Class Mark: x_i	Deviation $d_i = x_i - a$	f_i	$d_i \times f_i$

1000 – 2000	1500	−2000	2	−4000
2000 – 3000	2500	−1000	3	−3000
3000 – 4000	3500	0	6	0
4000 – 5000	4500	1000	5	5000
5000 – 6000	5500	2000	3	6000
6000 – 7000	6500	3000	1	3000
			$\sum f_i = 20$	**$\sum d_i f_i = 7000$**

Using the formula,

$$\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$$

$$\bar{x} = 3500 + \frac{7000}{20}$$

= 3500 + 350

= 3850

Thus, the average lifetime of a bulb is 3850 h.

Example 2:

The following table shows the percentage of girls in the top 33 schools of Delhi.

Percentage of Girls	Number of schools
15 – 25	5
25 – 35	6
35 – 45	10
45 – 55	5
55 – 65	4
65 – 75	2

75 – 85	1

Find the mean percentage of girls by using the assumed mean method.

Solution:

Firstly, we calculate the class mark of each class-interval and assume one of them as the assumed mean. Here, we take *a* = 50. The calculations of the deviations and the product of the deviations with the corresponding frequencies have been represented in the following table.

Percentage of girls	Class Mark x_i	Deviation $d_i = x_i - a$	Frequency f_i	$d_i \times f_i$
15-25 25-35 35-45 45-55 55-65 65-75 75-85	20 30 40 50 60 70 80	−30 −20 −10 0 10 20 30	5 6 10 5 4 2 1	−150 −120 −100 0 40 40 30
			$\sum f_i = 33$	$\sum d_i f_i = -260$

Using the formula, $\overline{x} = a + \frac{\sum f_i d_i}{\sum f_i}$, we obtain

$$\overline{x} = 50 + \frac{-260}{33}$$
$$= 50 - 7.88$$
$$= 42.12$$

Thus, the mean of the given data is 42.12 i.e., the mean percentage of girls in the 33 schools of Delhi is 42.12.

Mean Of Grouped Data Using Step Deviation Method

In the assumed mean method, we assume the mean as *a* from the class-marks and we use the following formula to find the actual mean.

$$\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$$

Where, d_i is the deviation of a from x_i.

When the data is given in the form of large numbers, then sometimes it is difficult and even lengthy to find the mean by using the assumed mean method. In such cases, we follow another method called **step deviation method** or **shift of origin and scale method.** Using this method, we can further reduce the calculation.

Example 1: Calculate the mean of the following data using step deviation method.

Class-interval	**0 – 20**	**20 – 40**	**40 – 60**	**60 –80**	**80 –100**	**100 –120**	**120 – 140**
Frequency	**12**	**22**	**16**	**13**	**17**	**8**	**12**

Solution:

Let us assume a = 50

Let us calculate the deviations of class marks from the assumed mean a. The calculations are shown in the following table.

Class-interval	**Frequency** f_i	**Class Mark** x_i	$d_i = x_i - a$	$u_i = \frac{x_i - a}{h}$	$f_i \times u_i$

0 – 20	12	10	–40	–2	–24
20 – 40	22	30	–20	–1	–22
40 – 60	16	50 = *a*	0	0	0
60 – 80	13	70	20	1	13
80 – 100	17	90	40	2	34
100 – 120	8	110	60	3	24
120 – 140	12	130	80	4	48
Total	$\sum f_i = 100$				$\sum u_i f_i = 73$

Highest common factor of all the *di*'s = *h* = 20

Now, using the formula $\bar{x} = a + h \times \frac{\sum f_i u_i}{\sum f_i}$, we obtain

$$\bar{x} = 50 + 20 \times \frac{73}{100}$$

$$\bar{x} = 50 + \frac{146}{10}$$

$$\bar{x} = 50 + 14.6$$

$$\bar{x} = 64.6$$

Thus, the mean of the given data is 64.6.

Mode Of Grouped Data

Consider the following frequency table.

Observation	1	2	3	4	5	6
Frequency	28	18	25	21	41	16

From the table, we can easily observe that the observation 5 has maximum frequency which is 41.

Thus, the mode of the given data is 5.

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 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 use a formula to find the mode. Therefore, first of all, let us look at the formula to be used in such cases.

The formula used to find the mode when the data is given in groups is as follows.

Mode = $l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$

Where, l = lower limit of the modal class

h = size of class interval (all class sizes should be equal)

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

Before discussing the previous example, let us first know what a **modal class** is.

"A class corresponding to maximum frequency is known as modal class and the mode of the data lies inside the **modal class**".

Let us discuss the previous example to understand the method of finding the mode, when the data is given in a grouped form.

Now, consider the previous example again.

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

Here, the maximum frequency is 25, and this corresponds to the group 60 – 70. Therefore, the modal class is 60 – 70.

We can calculate the mode by using the following formula.

Mode = $l+\left(\frac{f_1-f_0}{2f_1-f_0-f_2}\right)\times h$

From the above table,

$l = 60$

$f_1 = 25$

$f_0 = 19$

$f_2 = 22$

$h = 70 - 60 = 10$

Using the formula,

Mode = $l+\left(\frac{f_1-f_0}{2f_1-f_0-f_2}\right)\times h$

$=60+\left(\frac{25-19}{2(25)-19-22}\right)\times 10$

Mode = $60+\left(\frac{6}{9}\right)\times 10$

$=60+\frac{20}{3}$

= 60 + 6.67

= 66.67

Therefore, the mode of the given data is 66.67.

Here, we can observe that the mode lies inside the modal class 60 − 70.

Let us solve more problems using this formula.

Example 1: The following table shows the class intervals and the frequency corresponding to them.

Class Interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
Frequency	12	15	24	16	21

Find the mode of the given data.

Solution:

In the given table, the maximum frequency is 24 corresponding to the class interval 40 – 60. Therefore, the modal class is 40 – 60.

From the table,

$l = 40$

$h = 20$

$f_0 = 15$

$f_2 = 16$

$f_1 = 24$

Mode = $l+\left(\frac{f_1-f_0}{2f_1-f_0-f_2}\right)\times h$

$= 40+\left(\frac{24-15}{2(24)-15-16}\right)\times 20$

Mode = $40+\left(\frac{9}{17}\right)\times 20$

$= 40+\frac{180}{17}$

= 40 + 10.59

= 50.59

Thus, the mode of the given data is 50.59.

Example 2: Data collected about the weights of 70 students of a class is given in the following table.

Weight (in kg)	**40 – 44**	**45 – 49**	**50 – 54**	**55 – 59**	**60 – 64**	**65 – 69**
Number of students	**6**	**11**	**23**	**17**	**9**	**4**

Find the modal weight of the class.

Solution:

It can be seen that the class intervals are of inclusive type, so we need to make them of exclusive type to calculate the mode.

The modified table will be as follows:

Original Class	**Class Boundaries**	**Frequency**
40 – 44	39.5 – 44.5	6
45 – 49	44.5 – 49.5	11
50 – 54	49.5 – 54.5	23
55 – 59	54.5 – 59.5	17
60 – 64	59.5 – 64.5	9
65 – 69	64.5 – 69.5	4

Here, the maximum frequency is 23 which corresponds to the class interval 49.5 – 54.5. Therefore, the modal class is 49.5 – 54.5.

From the table, we have

$l = 49.5$

$h = 5$

$f_0 = 11$

$f_2 = 17$

$f_1 = 23$

Mode = $l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$

$= 49.5 + \left(\frac{23-11}{2\times23-11-17}\right) \times 5$
$= 49.5 + \left(\frac{12}{18}\right) \times 5$
$= 49.5 + \frac{10}{3}$

= 49.5 + 3.33

= 52.83

Thus, the modal weight of the class is 52.83 kg.

Example 3: If the modal class is 45 – 60 and mode of the distribution is 58, then find the missing frequency in the following table.

Class Interval	0 – 15	15 – 30	30 – 45	45 – 60	60 – 75	75 – 90	90 – 105
Frequency	2	5	10	–	21	12	3

Solution:

The mode of the distribution is given as 58, which lies in the class interval 45 – 60.

Hence, 45 – 60 is the modal class.

Let the frequency of modal class 45 – 60 be f.

From the table,

$l = 45$

$h = 15$

$f_0 = 10$

$f_2 = 21$

$f_1 = f$

Using the formula, mode = $l+\left(\frac{f_1-f_0}{2\times f_1-f_0-f_2}\right)\times h$,

Mode = $l+\left(\frac{f_1-f_0}{2\times f_1-f_0-f_2}\right)\times h$

$$58=45+\left(\frac{f-10}{2f-10-21}\right)\times 15$$
$$58-45=\left(\frac{f-10}{2f-31}\right)\times 15$$
$$13=\left(\frac{f-10}{2f-31}\right)\times 15$$
$$13(2f-31)=15(f-10)$$
$$26f-403=15f-150$$
$$11f=253$$
$$f=\frac{253}{11}=23$$

Thus, the missing frequency is 23.

Median of Grouped Data Using Cumulative Frequency Table

Let us consider a group of numbers such as 17, 12, 15, 20, and 14. We can easily find the median of the given numbers. To find the median, first of all, we arrange the terms in increasing order. Then, we will have the sequence as 12, 14, 15, 17, and 20.

Here, we have 5 observations and the middle term is the 3rd term i.e., 15.

We know that median is the middle most observation of the data. Therefore, we can say 15 is the median of the given group of numbers.

However, we face problems when the data is given in grouped form. In such cases, we cannot use the above method to find the median. The method to find the median in such cases is different from the above. Let us discuss the method by taking an example.

The following table shows the wages of 60 workers of a company. Let us find the median of wages of workers using this information.

Wage (in Rs)	0 – 1000	1000 – 2000	2000 – 3000	3000 – 4000	4000 – 5000
No. of workers	10	15	20	10	5

Sometimes, the data is given as 'less than type' or 'more than type'. In these cases, firstly we have to make cumulative frequency distribution table and then we can find the median using the same formula as in the previous example.

Let us discuss each of them briefly.

1. Less than Type

Let us consider the following table which contains the marks obtained out of 100 marks in an exam by the students of Grade 10 of a school.

Marks	**Number of Students**
Less than 10	2
Less than 20	6
Less than 30	9
Less than 40	21
Less than 50	41
Less than 60	66
Less than 70	90
Less than 80	108
Less than 90	117
Less than 100	120

To find the median of such data, first of all, we need to write the table in the form of class intervals and find the frequency of each class interval. The frequencies given for less than type table are cumulative frequencies. Now, the number of students obtaining marks in the class interval 0 – 10 is same as the number of students obtaining marks less than 10. Therefore, the frequency of class interval 0 – 10 is 2. Now, the students obtaining marks less than 20 include the students of class interval 0 – 10 and 10 – 20. Therefore, frequency of class interval 10 – 20 = 6 – 2 = 4

Similarly, frequency of class interval 20 – 30 = 9 – 6 = 3

In this way, we can find the frequency of a class interval by taking the difference between the numbers of students of a group and its preceding group which is shown by the following table.

Class Interval	**Number of Students (frequency)**
0 – 10	2
10 – 20	6 – 2 = 4
20 – 30	9 – 6 = 3
30 – 40	21 – 9 = 12
40 – 50	41 – 21 = 20
50 – 60	66 – 41 = 25
60 – 70	90 – 66 = 24
70 – 80	108 – 90 = 18
80 – 90	117 – 108 = 9
90 – 100	120 – 117 = 3

Since we have obtained the frequency distribution table, we can easily convert it into a cumulative frequency distribution table and follow the same method as we have studied in the beginning of this learning section. By doing so, we will easily find out the median of the given data.

2. More than Type

Let us consider the following table as the performance (in %) of taking calls by the customer care executives in a call centre.

Performance (in %)	**Number of customer care executives**

More than 0	1500
More than 20	1280
More than 40	964
More than 60	585
More than 80	209

To find the median of such data, first of all, we are required to write the table in the form of class intervals and find the frequency of each class interval. By looking at the given table, we write the class intervals of this data as 0 – 20, 20 – 40, 40 – 60, 60 – 80, and 80 – 100. Now, the class more than 80 contains the customer care executives (C.C.E.) of class interval 80 – 100. Therefore, frequency of class interval 80 – 100 is 209. The class more than 60 contains the C.C.E. of class intervals 60 – 80 and 80 – 100. Therefore, frequency of class interval 60 – 80 = 585 – 209 = 376

Similarly, frequency of class interval 40 – 60 = 964 – 585 = 379

In this way, we can find the frequency of a class interval by taking the difference between the number of C.C.E .of a class and its following class as shown below.

Performance (in %)	**Number of customer care executives**
0 – 20	1500 – 1280 = 220
20 – 40	1280 – 964 = 316
40 – 60	964 – 585 = 379
60 – 80	585 – 209 = 376
80 – 100	209

Since we have obtained the frequency distribution table, we can easily convert it into a cumulative frequency distribution table and follow the same method as we have studied in the beginning. By doing so, we will easily find out the median of the given data.

Remember

- When a question is of more than or less than type, the cumulative frequency is always given.
- When the question is of more than type, the cumulative frequency is always in decreasing order.
- When the question is of less than type, the cumulative frequency is always in increasing order.

Let us solve some examples based on finding the median of less than type and more than type observations.

Example 1: The following table shows the literacy rate (in %) of 20 cities. Find the median literacy rate.

Literacy rate (in %)	**Less than 30%**	**Less than 40%**	**Less than 50%**	**Less than 60%**
Number of cities	**2**	**6**	**14**	**20**

Solution:

The given table is of less than type. Therefore, the number of cities given in the table represents the cumulative frequencies. The frequencies and the class intervals are shown in the following table.

Literacy rate (in %)	**Number of cities (c_f)**	**Frequency**
Less than 30 30 – 40 40 – 50 50 – 60	2 6 14 20	2 4 8 6
		n = 20

Here, number of observations, $n = 20$

$$\therefore \frac{n}{2} = 10$$

The cumulative frequency greater than and nearest to 10 is 14.

∴ 40 – 50 is the median class.

l (lower limit of median class) = 40

h (class size) = 10

n (number of observations) = 20

cf = (cumulative frequency of the class preceding the median class) = 6

f (frequency of median class) = 8

Using the formula,

Median $= l + \dfrac{\left(\frac{n}{2} - cf\right)}{f} \times h$

$$= 40 + \left(\frac{10-6}{8}\right) \times 10$$
$$= 40 + \frac{4}{8} \times 10$$
$$= 40 + 5$$
$$= 45$$

Therefore, 45% is the median literacy rate of 20 cities.

Example 2: If the median of the following distribution is 28.5, then find the value of *x* and *y*.

Class Interval	Frequency
0 – 10 **10 – 20** **20 – 30** **30 – 40** **40 – 50** **50 – 60**	**5** **x** **20** **13** **y** **7**
	Total = 60

Solution:

First of all, we have to construct a cumulative frequency table.

Class Interval	Frequency	Cumulative Frequency

0 – 10 10 – 20 20 – 30 30 – 40 40 – 50 50 – 60	5 x 20 13 y 7	5 5 + x 25 + x 38 + x 38 + x + y 45 + x + y
		45 + x + y

It is given that $n = 60$

$45 + x + y = 60$

$y = 15 - x$ ... (1)

But, median = 28.5

i.e., median lies between the class interval 20 – 30.

Hence, median class = 20 – 30

Therefore, we have

$l = 20, f = 20, cf = 5 + x, h = 10$, and $\frac{n}{2} = 30$

It is given that the median is 28.5.

$$l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h = 28.5$$

$$20 + \frac{30 - 5 - x}{20} \times 10 = 28.5$$

$$20 + \frac{25 - x}{2} = 28.5$$

$$\frac{40 + 25 - x}{2} = 28.5$$

$65 - x = 57$

$x = 65 - 57 = 8$

Putting this value in equation (1), we obtain

$y = 15 - x$

$y = 15 - 8 = 7$

Hence, $x = 8$ and $y = 7$

Example 3: The following table shows the heights of buildings of a city. Find the median height.

Height of building (in feet)	More than 0	More than 100	More than 200	More than 300	More than 400
Number of buildings	28	25	20	12	5

Solution:

The given table is of more than type. Therefore, the number of buildings given in the table represents the cumulative frequencies. The frequencies and the class intervals have been shown in the following table.

Height of building (in feet)	**Number of buildings**	**Frequency**	**Cumulative frequency (cf)**
0 – 100	28	28 – 25 = 3	3
100 – 200	25	25 – 20 = 5	8
200 – 300	20	20 – 12 = 8	16
300 – 400	12	12 – 5 = 7	23
More than 400	5	5	28
		n = 28	

Here, $n = 28$

$$\therefore \frac{n}{2} = 14$$

The cumulative frequency which is greater than and nearest to 14 is 16.

∴ 200 – 300 is median class.

We also have,

l (lower limit of median class) = 200

h (class size) = 100

n (number of observations) = 28

cf = (cumulative frequency of the class preceding the median class) = 8

f (frequency of median class) = 8

Using the formula, we have

Median $= l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$

$= 200 + \left(\frac{14-8}{8}\right) \times 100$

= 200 + 75

= 275

Hence,275 feet height is the median height.

Example 4: Lowest temperature of the day of a metro city was recorded for 30 consecutive days and the results are shown below:

Temperature (in °C)	2.1 – 3	3.1 – 4	4.1 – 5	5.1 – 6
Number of days	5	15	3	7

Find the median lowest temperature.

Solution:

It can be observed that the class intervals in the given table are of inclusive type, so we need to convert them to exclusive type so that we can get continuous class intervals. Here, the difference between upper limit of a class and lower limit of next class is 0.1. On dividing this difference by 2, we get 0.05 which can be added to every upper limit and subtracted from every lower limit to get new limits. These limits are known as class boundaries. Using these class boundaries, we can prepare cumulative frequency distribution table and thus, we can get continuous frequency distribution.

The new table can be prepared as follows:

Original Class	**Class Boundaries**	**Frequency**	**Cumulative Frequency (Less than type)**
2.1 – 3	2.05 – 3.05	5	5
3.1 – 4	3.05 – 4.05	15	20
4.1 – 5	4.05 – 5.05	3	23
5.1 – 6	5.05 – 6.05	7	30
		Total = 30	

Here, $n = 30$

$\therefore \frac{n}{2} = 15$

The cumulative frequency which is greater than and nearest to 15 is 20.

∴ 3.05 – 4.05 is median class.

We also have,

l (lower limit of median class) = 3.05

h (class size) = 1

n (number of observations) = 30

cf = (cumulative frequency of the class preceding the median class) = 5

f (frequency of median class) = 15

Using the formula, we have

Median $= l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$

$$= 3.05 + \left(\frac{\frac{30}{2} - 5}{15}\right) \times 1$$
$$= 3.05 + 0.67$$
$$= 3.72$$

Hence, 3.72 °C is the median lowest temperature.

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.

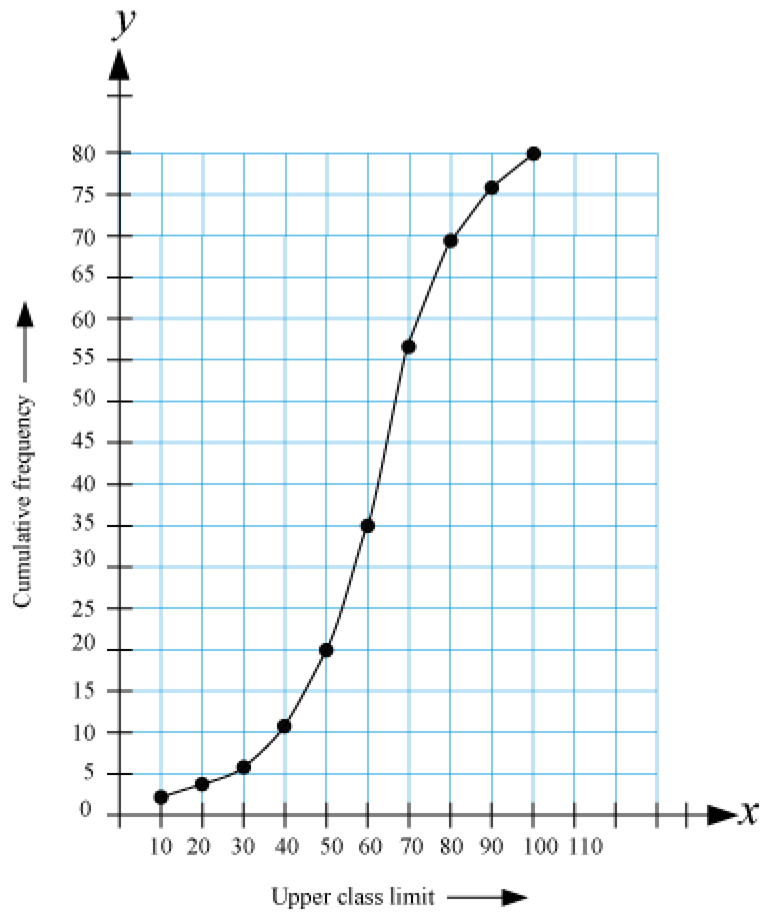
(2) Now, we mark the upper limits on the horizontal line and the cumulative

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

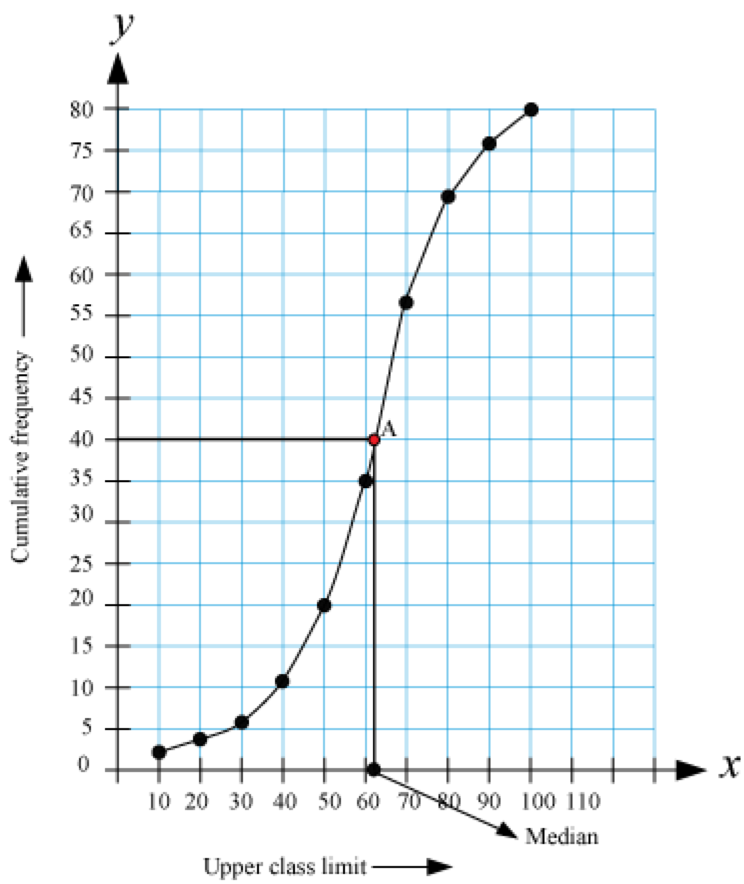
An important application of ogive in statistics is to find the median.

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

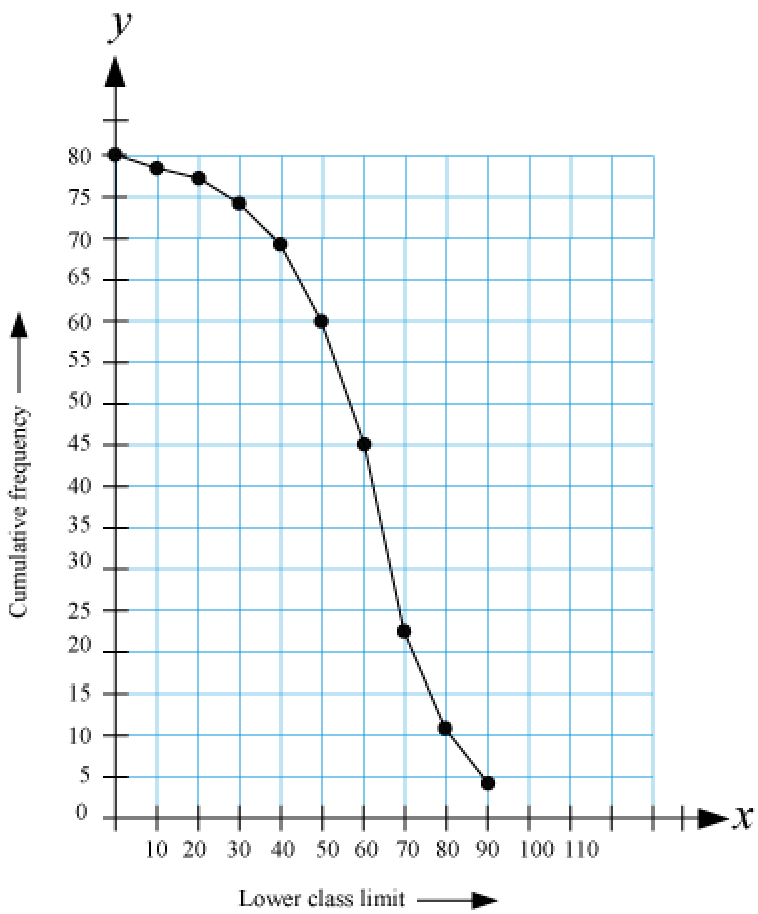
Construction of Ogive of more than type

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.

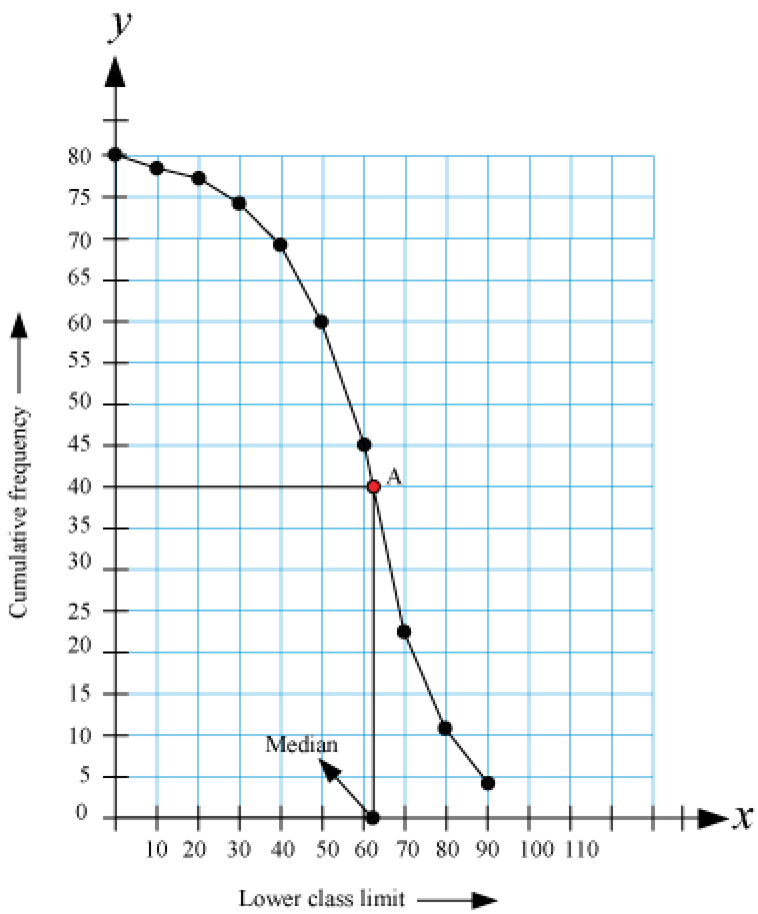


To find the median with the help of ogive

Number of observations, n = 80

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

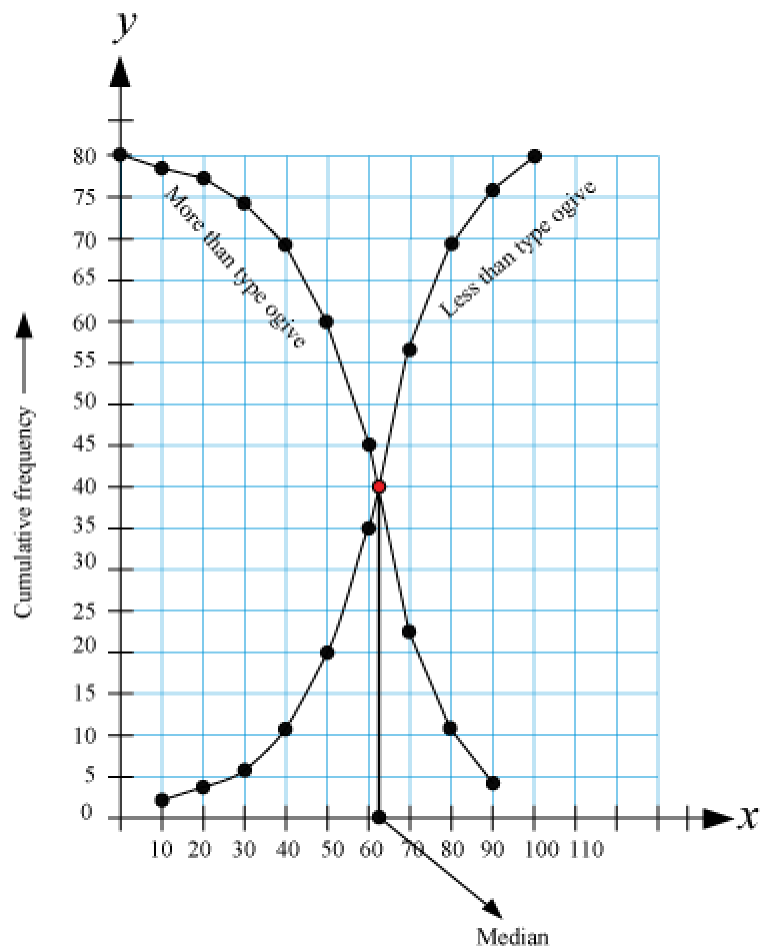
Mark the point 40 on the vertical line and then draw a horizontal line from this point. Let this line intersect the ogive at point A. Now, draw a vertical line through A. Median is the point at which this vertical line intersects the horizontal line.



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.

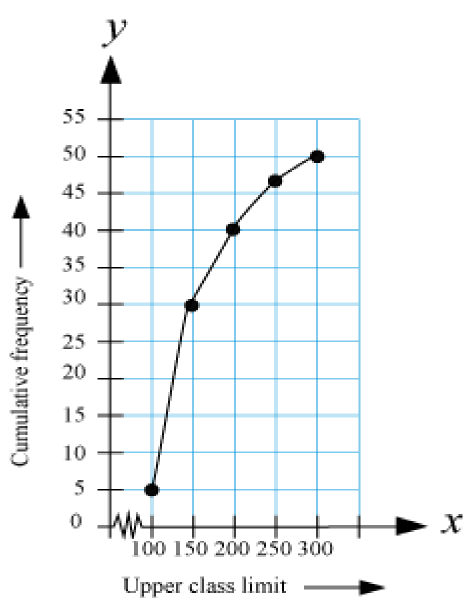
Solution:

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.

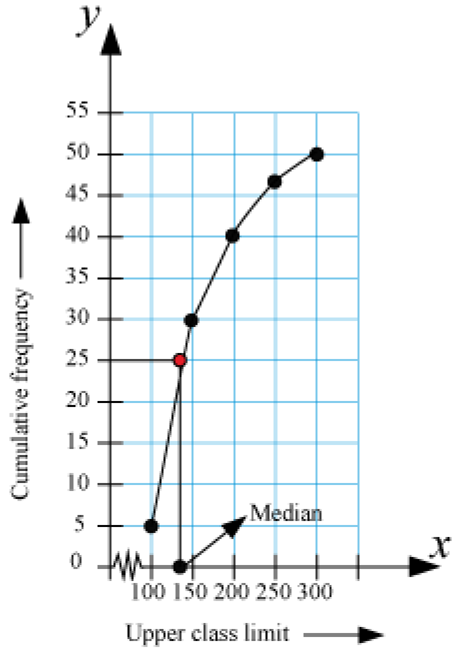
The ogive obtained has been represented in the following figure.



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

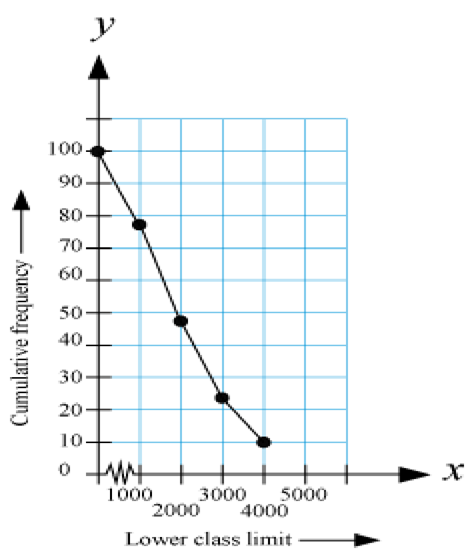
Convert it into a more than type of distribution and draw its ogive.

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.

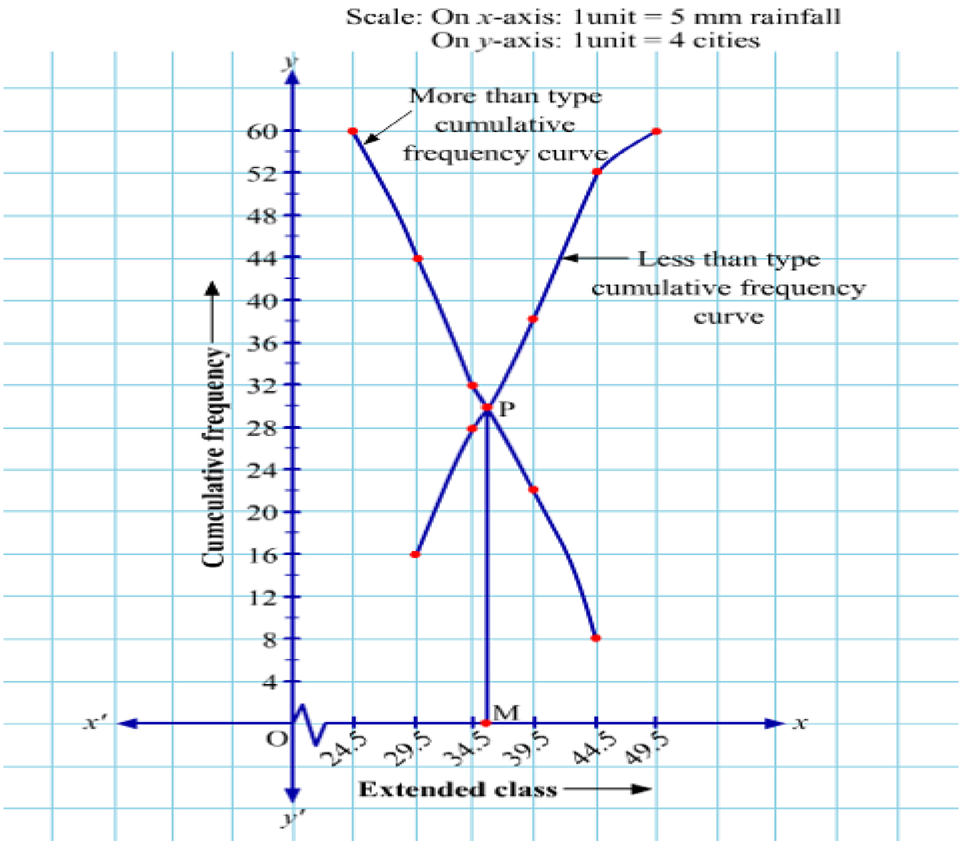
The table is given below.

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.

Relation Between Mean, Median, And Mode

We have studied three measures of central tendency - mean, median, and mode. But do you know that there is an empirical relationship between these three measures.

The relation between them is given by

Mean – Mode = 3(Mean – Median)

or

3 Median = Mode + 2 Mean

We can use this relationship to find the value of any of them, when the other two are known. It can also be used to check if the values of the three central tendencies obtained are correct or not as they will satisfy this relation. However, this relation is not always true.

First of all, let us verify it by taking an example.

Example: The following data shows the numbers of days for which the patients were hospitalized.

No. of Patients	No. of days attending the hospital
0 – 10	2
10 – 20	6
20 – 30	9
30 – 40	7
40 – 50	4
50 – 60	2

Justify the relationship between mean, median, and mode of the given data.

Solution:

To verify the relation between mean, median, and mode, we have to find out their values from the given data and then substitute them to see the authenticity of the relationship.

Firstly, let us find the mean of the given data.

Mean

To find out the mean, we construct the following table.

No. of Patients	No. of days for which the patients were hospitalized: f_i	Class Mark: x_i	$f_i \times x_i$
0 – 10	2	5	10
10 – 20	6	15	90
20 – 30	9	25	225
30 – 40	7	35	245
40 – 50	4	45	180
50 – 60	2	55	110
	$\sum f_i = 30$		$\sum f_i x_i = 860$

We know that, Mean = $\frac{\sum f_i x_i}{\sum f_i}$

$= \frac{860}{30} = \frac{86}{3}$

Now, let us find the mode of the given set of data.

Mode

In the given table, 20 − 30 is the modal class as this group has the maximum frequency.

From the table,

l = 20 (Lower limit of the modal class)

h = 10 (Class size)

f_1 = 9 (Frequency of the modal class)

f_0 = 6 (Frequency of the class preceding the modal class)

f_2 = 7 (Frequency of the class succeeding the modal class)

Now, Mode = $l+\left(\frac{f_1-f_0}{2\times f_1-f_0-f_2}\right)\times h$

Mode $=20+\left(\frac{9-6}{2\times 9-6-7}\right)\times 10$

Mode $= 20+\frac{3}{5}\times 10$ = 26

Median

To find out the median, we construct the following table.

No. of Patients	Frequency	Cumulative Frequency
0 – 10	2	2
10 – 20	6	8
20 – 30	9	17
30 – 40	7	24
40 – 50	4	28
50 – 60	2	30
		30

Here, $n = 30$

$\therefore \frac{n}{2} = 15$

The cumulative frequency greater than 15 is 17.

Hence, the median class is 20 – 30.

From the table,

l = 20 (Lower limit of median class)

cf = 8 (Cumulative frequency of class preceding the median class)

f = 9 (Frequency of median class)

h = 10 (Class size of the groups)

Now, Median $= l + \left(\frac{\frac{n}{2} - cf}{f} \right) \times h$

Median = $20 + \left(\frac{15 - 8}{9} \right) \times 10$

Median = $20 + \frac{70}{9} = \frac{250}{9}$

Now, Mode + 2 Mean = 26 + 2 × $\frac{86}{3}$

= 26 + $\frac{172}{3}$

= $\frac{250}{3}$

But, 3 Median = 3 × $\frac{250}{9}$ = $\frac{250}{3}$ = Mode + 2 Mean

Hence it is verified that,

3 Median = Mode + 2 Mean

Let us go through few more examples to learn the application of the concept.

Example 1: For a certain frequency distribution, mean and median are obtained as 123.5 and 127 respectively. What is the value of mode?

Solution:

We have the relation

3 Median = Mode + 2 Mean

$\Rightarrow$ 3(127) = Mode + 2(123.5)

$\Rightarrow$ 381 = Mode + 247

$\Rightarrow$ Mode = 381 – 247

$\Rightarrow$ Mode = 134

Example 2: For a certain frequency distribution, median and mode are obtained as 55 and 70 respectively. What is the value of mean?

Solution:

We have the relation

3 Median = Mode + 2 Mean

$\Rightarrow$ 3(55) = 70 + 2 Mean

$\Rightarrow$ 165 = 70 + 2 Mean

$\Rightarrow$ 2 Mean = 165 – 70

$\Rightarrow$ 2 Mean = 95

$\Rightarrow$ Mean = 47.5

Example 3: For a certain frequency distribution, mean is 6 more than mode. Find the relation between median and mode.

Solution:

It is given that

Mean – Mode = 6 ...(1)

We have the relation

Mean – Mode = 3(Mean – Median)

⇒ 6 = 3(Mean – Median)

⇒ Mean – Median = 2 ...(2)

On subtracting (2) from (1), we get

Mean – Mode – (Mean – Median) = 6 – 2

⇒ Mean – Mode – Mean + Median = 4

⇒ Median – Mode = 4

Thus, median is 4 more than mode.