Measures of Central Tendency

Mean of Data Sets

Application of Mean in Real Life

The runs scored by the two opening batsmen of a team in ten successive matches of a cricket series are listed in the table.

Player A	24	50	34	24	20	96	105	50	13	27
Player B	26	22	30	10	42	98	40	54	10	122

Using this data, we can compare the performances of the players for each individual game. For example, player B performed better than player A in the first match, player A then performed better than player B in the second match, etc.

This method, however, is not useful in trying to determine the overall performances of the two players and comparing them. For this we need to calculate the average or mean score of each player. The player having the better average or mean score has the better overall performance.

In this lesson, we will learn how to find the mean of a data set.

Mean of a Data Set

Did You Know?

1. Arithmetic mean (AM), mean or average are all the same.

2. Mean is used in calculating average temperature, average mark, average score, average age, etc. It is also used by the government to find the average individual expense and income.

3. Mean cannot be determined graphically.

4. Mean is supposed to be the best measure of central tendency of a given data.

5. Mean can be determined for almost every kind of data.

Properties of Mean

1. Sum of the deviations taken from the arithmetic mean is zero.
If the mean of n observations $x_1, x_2, x_3 .... x_n$ is $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + ... + (x_n - \bar{x}) = 0$.

2. If each observation is increased by p then the mean of the new observations is also increased by p.
If the mean of n observations $x_1, x_2, x_3 .... x_n$ is $\bar{x}$ then the mean of $(x_1 + p),\ (x_2 + p),\ (x_3 + p),\ ...,\ (x_n + p)$ is $(\bar{x} + p)$.

3. If each observation is decreased by p then the mean of the new observations is also decreased by p.
If the mean of n observations $x_1, x_2, x_3 .... x_n$ is $\bar{x}$ then the mean of $(x_1 - p),\ (x_2 - p),\ (x_3 - p),\ ...,\ (x_n - p)$ is $(\bar{x} - p)$.

4. If each observation is multipled by p(where $p \neq 0$) then the mean of the new observations is also multiplied by p.
If the mean of n observations $x_1, x_2, x_3 .... x_n$ is $\bar{x}$ then the mean of $px_1,\ px_2,\ px_3,\ ...,\ px_n$ is $p\bar{x}$.

5. If each observation is divided by p(where $p \neq 0$) then the mean of the new observations is also divided by p.
If the mean of n observations $x_1, x_2, x_3 .... x_n$ is $\bar{x}$ then the mean of $\frac{x_1}{p}, \frac{x_2}{p}, \frac{x_3}{p}, ..., \frac{x_n}{p}$ is $\frac{\bar{x}}{p}$.

Solved Examples

Easy

Example 1:

The amounts of money spent by Sajan during a particular week are listed in the table.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Money spent (in rupees)	270	255	195	230	285	225	115

Find the average amount of money spent by him per day.

Solution:

$$\text{Average amount of money spent by Sajan per day} = \frac{\text{Total money spent}}{\text{Total number of days}}$$
$$= \text{Rs}\,\frac{270 + 255 + 195 + 230 + 285 + 225 + 115}{7}$$
$$= \text{Rs}\,\frac{1575}{7}$$
$$= \text{Rs}\,225$$

Example 2:

The average weight of the students in a class is 42 kg. If the total weight of the students is 1554 kg, then find the total number of students in the class.

Solution:

Let the total number of students in the class be *x*.

$$\text{Average weight of the students} = \frac{\text{Total weight of the students}}{\text{Total number of students}}$$
$$\Rightarrow \text{Total number of students} = \frac{\text{Total weight of the students}}{\text{Average weight of the students}}$$
$$\Rightarrow \therefore x = \frac{1554}{42}$$
$$= 37$$

Thus, there are 37 students in the class.

Medium

Example 1:

For what value of *x* is the mean of the data 28, 32, 41, *x*, *x*, 5, 40 equal to 31?

Solution:

$$\text{Mean of the given data set} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$
$$\Rightarrow 31 = \frac{28+32+41+x+x+5+40}{7}$$
$$\Rightarrow 217 = 2x + 146$$
$$\Rightarrow 2x = 71$$
$$\Rightarrow \therefore x = 35.5$$

Thus, for *x* = 35.5, the mean of the data 28, 32, 41, *x*, *x*, 5, 40 is 31.

Example 2:

The numbers of children in five families are 0, 2, 1, 3 and 4. Find the average number of children. If two families having 6 and 5 children are included in this data set, then what is the new mean or average?

Solution:

$$\text{Mean of the given data set} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$
$$\therefore \text{Mean of the initial data set} = \frac{0+2+1+3+4}{5} = \frac{10}{5} = 2$$

Thus, the average number of children for the five families in the initial data set is 2.

Two families are added to the initial set of families.

$$\therefore \text{Mean of the new data set} = \frac{0+2+1+3+4+6+5}{7} = \frac{21}{7} = 3$$

Thus, the average number of children for the seven families in the new data set is 3.

Example 3:

The mean of fifteen numbers is 7. If 3 is added to every number, then what will be the new mean?

Solution:

Let $x_1, x_2, x_3, \ldots, x_{15}$ be the fifteen numbers having the mean as 7 and x¯x¯ be the mean.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$
$$\Rightarrow 7 = \frac{x_1 + x_2 + x_3 + \ldots + x_{15}}{15}$$
$$\Rightarrow x_1 + x_2 + x_3 + \ldots + x_{15} = 15 \times 7$$
$$\Rightarrow x_1 + x_2 + x_3 + \ldots + x_{15} = 105 \qquad \ldots(1)$$

The new numbers are $x_1 + 3,\ x_2 + 3,\ x_3 + 3,\ \ldots,\ x_{15} + 3$.

Let $\overline{X}$ be the mean of the new numbers.

$$\overline{X} = \frac{(x_1+3)+(x_2+3)+...+(x_{15}+3)}{15}$$

$$\Rightarrow \overline{X} = \frac{(x_1+x_2+...+x_{15})+3\times 15}{15}$$

$$\Rightarrow \overline{X} = \frac{105+45}{15} \qquad \text{(By equation 1)}$$

$$\Rightarrow \overline{X} = \frac{150}{15}$$

$$\Rightarrow \therefore \overline{X} = 10$$

Thus, the mean of the new numbers is 10.

Hard

Example 1:

The average salary of five workers in a company is Rs 2500. When a new worker joins the company, the average salary is increased by Rs 100. What is the salary of the new worker?

Solution:

Let the salary of the new worker be Rs *x*.

Before the joining of the new worker, we have:

$$\text{Mean salary of the five workers} = \frac{\text{Sum of the salaries of the five workers}}{5}$$

$$\Rightarrow 2500 = \frac{\text{Sum of the salaries of the five workers}}{5}$$

$$\Rightarrow \therefore \text{Sum of the salaries of the five workers} = 2500\times 5 = 12500 \qquad ...(1)$$

After the joining of the new worker, we have:

Number of workers = 5 + 1 = 6

Average salary = Rs (2500 + 100) = Rs 2600

$$\text{Mean salary of the six workers} = \frac{\text{Sum of the salaries of the six workers}}{6}$$

$$\Rightarrow 2600 = \frac{\text{Sum of the salaries of the five workers} + \text{Salary of the new worker}}{6}$$

$$\Rightarrow 2600 = \frac{12500 + x}{6} \qquad \text{(By equation 1)}$$

$$\Rightarrow 15600 = 12500 + x$$

$$\Rightarrow \therefore x = 15600 - 12500 = 3100$$

Thus, the salary of the new worker is Rs 3100.

Example 2:

Find two numbers that lie between $\frac{2}{5}$ and $\frac{1}{2}$.

Solution:

The given numbers are $\frac{2}{5}$ and $\frac{1}{2}$.

Mean of the two numbers = $\dfrac{\frac{2}{5}+\frac{1}{2}}{2} = \dfrac{\frac{4+5}{10}}{2} = \dfrac{9}{10\times 2} = \dfrac{9}{20}$

Now, we know that the mean of any two numbers lies between the numbers.

Hence, $\frac{2}{5} < \frac{9}{20} < \frac{1}{2}$

Mean of $\frac{9}{20}$ and $\frac{1}{2}$ = $\dfrac{\frac{9}{20}+\frac{1}{2}}{2} = \dfrac{\frac{9+10}{20}}{2} = \dfrac{19}{20\times 2} = \dfrac{19}{40}$

Hence, $\frac{9}{20} < \frac{19}{40} < \frac{1}{2}$

And $\frac{2}{5} < \frac{9}{20} < \frac{19}{40} < \frac{1}{2}$

So, two numbers that lie between $\frac{2}{5}$ and $\frac{1}{2}$ are $\frac{9}{20}$ and $\frac{19}{40}$.

Example 3:

If $\bar{x}$ is the mean of the *n* observations $x_1, x_2, x_3, ..., x_n$, then prove that

$$\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)}{n} = 0$$

Solution:

It is given that $\bar{x}$ is the mean of the *n* observations $x_1, x_2, x_3, ..., x_n$.

Thus,

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$
$$\Rightarrow x_1 + x_2 + x_3 + ... + x_n = n\bar{x} \qquad ...(1)$$

Now,

$$\begin{aligned}\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)}{n} &= \frac{\left(x_1 - \bar{x}\right) + \left(x_2 - \bar{x}\right) + \left(x_3 - \bar{x}\right) + ... + \left(x_n - \bar{x}\right)}{n} \\ &= \frac{\left(x_1 + x_2 + x_3 + ... + x_n\right) - \left(\bar{x} + \bar{x} + \bar{x} + ... + \bar{x}\right)}{n} \\ &= \frac{n\bar{x} - n\bar{x}}{n} \qquad \text{(By equation 1)} \\ &= \frac{0}{n} \\ &= 0\end{aligned}$$

Thus, the given result is proved.

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.

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

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 **20 − 40** **40 − 60** **60 − 80** **80 − 100** **100 − 120**	**5** f_1 **10** f_2 **7** **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 20 − 40 40 − 60 60 − 80 80 − 100 100 − 120	5 f_1 10 f_2 7 8	10 30 50 70 90 110	50 30 f_1 500 70 f_2 630 880

	$30 + f_1 + f_2 = 50$		$2060 + 30 f_1 + 70 f_2$

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.

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

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

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

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

$$30 f_1 + 70 f_2 = 1080$$

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.

Let us solve more examples using the assumed mean method.

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	20	−30	5	−150
25-35	30	−20	6	−120
35-45	40	−10	10	−100
45-55	50	0	5	0

55-65	60	10	4	40
65-75	70	20	2	40
75-85	80	30	1	30
			$\sum f_i = 33$	$\sum d_i f_i = -260$

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

$$\begin{aligned}\bar{x} &= 50 + \frac{-260}{33} \\ &= 50 - 7.88 \\ &= 42.12\end{aligned}$$

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.

Let us solve one more example to understand the step deviation method better.

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 $\overline{x} = a + h \times \frac{\sum f_i u_i}{\sum f_i}$, 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.

Medians of Data Sets Having Odd or Even Number of Terms

Median as the Measure of Central Tendency

Let us consider the group of half dozen individuals shown in the picture. There are five children standing with a very tall man.

You can see that the distribution of height in this group is unbalanced or asymmetrical because one individual is much taller than the others. When we calculate the mean height of the five children, the value so obtained will be close to the actual height of each child.

However, when we calculate the mean height of the six persons (including the really tall man), the value so obtained will give the impression that each individual in the group is quite tall. So, in this case, the mean will not be an appropriate measure of central tendency. In situations such as this, we use median as the measure of central tendency.

In this lesson, we will study about median and the method to calculate the same for any given data.

Method to Find Median

Median can be defined as follows:

Median is the value of the middlemost observation when the data is arranged in increasing or decreasing order.

The method to find median can be summarized as follows:

Step 1: Arrange the data in increasing or decreasing order.

Step 2: Let *n* be the number of observations. Here, two cases arise.

Case 1: When *n* is even, the median of the observations is given by the formula

Median = Mean of the $\left(\frac{n}{2}\right)^{\text{th}}$ **and** $\left(\frac{n}{2}+1\right)^{\text{th}}$ **observations**

Case 2: When *n* is odd, the median of the observations is given by the formula

Median = Value of the $\left(\frac{n+1}{2}\right)^{\text{th}}$ **observation**

Did You Know?

1. Median is used to measure the distribution of earnings, to calculate the poverty line, etc.

2. Median is independent of the range of the series as it is the middle value of a data. So, it is not affected by extreme values or end values.

3. The median of a data is incapable of further algebraic or mathematical treatment. For example, if we have the median of two or more groups, then we cannot find the median of the bigger group formed by combining the given groups.

4. Median is affected by the fluctuation in data as it depends only on one item, i.e., the middle term.

Know Your Scientist

Antoine Augustin Cournot

Antoine Augustin Cournot (1801−1877) was a French economist, philosopher and mathematician. The term median was introduced by him in 1843. He used this term for the value that divides a probability distribution into two equal parts or halves.

In the field of economic analysis, he developed the concept of functions and probability. He introduced the demand curve to show the relationship between price and demand for any given item. He is best remembered for his theory of strategic behaviour of competitors in a market having only two players, i.e., in a duopoly.

Solving Problems Based on Finding the Median of a Given Data

Know More

Advantages of median

1. Median is better suited for non-symmetrical distributions as it is not much affected by very low and high values. Non-symmetrical distribution means the data is distributed in such a way that the values toward one end are much higher or lower than the values toward the other end. For example, 1, 2, 3, 4, 25, 30, 50, 60 is a non-symmetrical distribution.

2. Knowing the median test score is important to people who want to know whether they belong to the 'better half of the population' or not.

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**".

Example 2:
Find the mode of the following marks obtained by 15 students.
2, 5, 1, 0, 8, 11, 8, 12, 19, 18, 13, 10, 9, 8, 1

Solution:
We arrange the given data as follows:
0, 1, 1, 2, 5, 8, 8, 8, 9, 10, 11, 12, 13, 18, 19
We observe that 8 occurs most often. So, the mode is 8.

Solved Examples

Easy

Example 1:

Find the median of these observations: 324, 250, 234, 324, 250, 196, 189, 250, 313, 227.

Solution:

On writing the observations in ascending order, we have the following sequence.

189, 196, 227, 234, 250, 250, 250, 313, 324, 324

Here, the number of observations (*n*) is 10, which is an even number.

∴ Median = Mean of the $\left(\frac{10}{2}\right)^{\text{th}}$ and $\left(\frac{10}{2}+1\right)^{\text{th}}$ observations

⇒ Median = Mean of the 5^{th} and 6^{th} observations

Here, the 5^{th} and 6^{th} observations are the same value, i.e., 250.

∴ Median $\frac{250+250}{2}$ = 250

Medium

Example 1:

The minimum temperatures (in °C) for fifteen days in a city are recorded as follows:

4.5, 4.7, 3.9, 5.2, 5.0, 4.2, 4.6, 4.2, 4.2, 4.5, 5.7, 2.3, 6.0, 3.5, 4.0

Find the median of the minimum temperatures.

Solution:

On arranging the data in ascending order, we obtain the following sequence.

2.3, 3.5, 3.9, 4.0, 4.2, 4.2, 4.2, 4.5, 4.5, 4.6, 4.7, 5.0, 5.2, 5.7, 6.0

Here, the number of observations (*n*) is 15, which is an odd number.

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

⇒ Median = $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

Here, the 8^{th} observation is 4.5.

Thus, the median of the minimum temperatures for the fifteen days was 4.5°C.

Example 2:

The marks obtained (out of 50) by fifteen students are 27, 31, 29, 35, 30, 42, 45, 41, 37, 32, 28, 36, 44, 34 and 43. Find the median. If the marks 27 and 44 are replaced by 25 and 46, then what will be the new median?

Solution:

The given marks can be arranged in ascending order as follows:

27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 41, 42, 43, 44, 45

Here, the number of observations (*n*) is 15, which is an odd number.

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

= $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

= 8^{th} observation

= 35

After replacing 27 and 44 by 25 and 46, the marks are arranged in ascending order as follows:

25, 28, 29, 30, 31, 32, 34, 35, 36, 37, 41, 42, 43, 45, 46

∴ New median = $\left(\frac{n+1}{2}\right)^{\text{th}}$ observation

= $\left(\frac{15+1}{2}\right)^{\text{th}}$ observation

= 8^{th} observation

= 35

Hard

Example 1:

The following observations are arranged in ascending order.

11, 17, 20, 25, 39, 2*y*, 3*y* + 1, 69, 95, 112, 135, 1204

If the median of the data is 53, then find the value of *y*.

Solution:

The observations in ascending order are given as follows:

11, 17, 20, 25, 39, 2*y*, 3*y* + 1, 69, 95, 112, 135, 1204

Here, the number of observations (n) is 12, which is even.

∴ Median = Mean of $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2}+1\right)^{\text{th}}$ observations

It is given that the median of the observations is 53.

∴ 53 = Mean of $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2}+1\right)^{\text{th}}$ observations.

⇒ 53 = Mean of $\left(\frac{12}{2}\right)^{\text{th}}$ and $\left(\frac{12}{2}+1\right)^{\text{th}}$ observations

⇒ 53 = Mean of $(6)^{\text{th}}$ and $(7)^{\text{th}}$ observations

The 6th and 7th observations are 2*y* and 3*y* + 1 respectively.

So,

$$53 = \frac{(2y)+(3y+1)}{2}$$
$$\Rightarrow 106 = 5y+1$$
$$\Rightarrow 5y = 105$$
$$\Rightarrow \therefore y = \frac{105}{5} = 21$$

Median of Grouped Data Given in Discrete Form Using Cumulative Frequency Table

The middle value obtained after arranging a data in ascending or descending order is known as **median** of the data.

The median of a data having an odd number of terms is simply the value of the middle observation whereas the median of a data having an even number of terms is the mean of the two middle observations of the data.

However, there are conditions where a grouped data is provided. In such cases, the above mentioned method cannot be directly applied to find the median of a data.
The method used in such cases is explained below with the help of an example.

The following table represents the salaries of 150 workers of a company.

Salary (in Rs)	8000	12000	16000	20000	24000
Number of workers	25	40	30	50	5

Let us find the median salary of the workers for the given information.

First, we have to find the cumulative frequency of the data.
The **cumulative frequency** (c.f.) for a class is the total of the frequencies above or below it.

The cumulative frequency of each class in the above table is calculated as follows:

Salary (in Rs) (x)	Number of workers (f)	Cumulative frequency (c.f.)
8000	25	25
12000	40	40 + 25 = 65
16000	30	30 + 65 = 95
20000	50	50 + 95 = 145
24000	5	5 + 145 = 150
	$\sum f = 150$	

The total frequency, N = $\sum f$ = 150

$\Rightarrow \frac{N}{2} = \frac{\sum f}{2} = \frac{150}{2} = 75$

The median is the $\left(\frac{N}{2}\right)^{th}$ observation, i.e., the 75^{th} observation.

Now, we can see that the 75^{th} observation occurs in the cumulative frequency 95 whose corresponding salary is Rs 16000.

∴ The median salary of the above data is Rs 16000.

Let us try to solve some more problems to understand the concept better.

Example 1:
The following table shows the heights of 53 students of a class. What is the median height of the class?

Height (in cm)	120	130	140	150	160
Number of students	4	8	17	20	4

Solution:

Height (in cm) (x)	Number of students (f)	Cumulative frequency (c.f.)
120	4	4
130	8	8 + 4 = 12
140	17	17 + 12 = 29
150	20	20 + 29 = 49
160	4	4 + 49 = 53
	$\sum f = 53$	

The total frequency, N = $\sum f = 53$

$$\Rightarrow \frac{N+1}{2} = \frac{\sum f+1}{2} = \frac{53+1}{2} = 27$$

The median is the $\left(\frac{N+1}{2}\right)^{th}$ observation, i.e., the 27^{th} observation.

Now, we can see that the 27^{th} observation occurs in the cumulative frequency 29 whose corresponding height is 140 cm.

∴ The median height of the class is 140 cm.

Example2:
What is the median of the given data?

x	***f***
5	3
10	7
15	4
20	11
25	2
35	15
40	12
50	4

Solution:

x	***f***	**c.f.**
5	3	3
10	7	7 + 3 = 10
15	4	4 + 10 = 14
20	11	11 + 14 = 25
25	2	2 + 25 = 27
35	15	15 + 27 = 42
40	12	12 + 42 = 54
50	4	4 + 54 = 58
	$\sum f = 58$	

The total frequency, N = $\sum f = 58$

$$\Rightarrow \frac{N}{2} = \frac{\sum f}{2} = \frac{58}{2} = 29$$

The median is the $\left(\frac{N}{2}\right)^{th}$ observation, i.e., the 29^{th} observation.

Now, we can see that the 29^{th} observation occurs in the cumulative frequency 42, whose corresponding variable ***x*** is 35.

∴ The median of the above data is 35.

Quartile Deviation

The range and quartile deviation are measures of dispersion that depend on the values of the variables at a particular position in the distribution. The range is based on extreme values in the distribution. It does not consider the deviation among the values. In order to study the variation among the values, the measure of inter-quartile range is used.

Inter-Quartile Range = Third Quartile − First Quartile = $Q_3 - Q_1$

Quartile deviation is the half of the difference between third quartile, Q_3 and first quartile, Q_1 of the series.

$$\therefore \text{Quartile deviation} = \frac{Q_3 - Q_1}{2}$$

Quartile deviation gives half of the range of middle 50% observations. Quartile deviation is also known as semi-inter-quartile range.

Calculation of quartile deviation

1. For an individual series, the first and third quartiles can be calculated using the following formula:

Q_1 = Value of $\frac{(n+1)}{4}$th ordered observation

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

2. For a discrete series, the first and third quartiles can be calculated using the following formula:

If $N = \sum f$, then

Q_1 = Value of $\frac{(N+1)}{4}$th ordered observation

Q_3 = Value of $\frac{3(N+1)}{4}$th ordered observation

3. For a continuous series, the first and third quartiles can be calculated using the following formula:

$$Q_1 = L + \left(\frac{\frac{N}{4} - \text{c.f.}}{f}\right) \times h$$

$$Q_3 = L + \left(\frac{\frac{3N}{4} - \text{c.f.}}{f}\right) \times h$$

Here, L = lower limit of the quartile class

f = frequency of the quartile class

h = class interval of quartile class

c.f. = total of all the frequencies below the quartile class

N = total frequency, ∑f∑f

Solved Examples

Example 1:

Find the quartile deviation for the following data.

15, 65, 30, 70, 50, 25, 40, 75, 45, 60

Solution:

First, arrange the observations in ascending order, as shown below:

15, 25, 30, 40, 45, 50, 60, 65, 70, 75

Here, number of observations, n = 10

Q_1 = Value of $\frac{(10+1)}{4}$th observation

= Value of 2.75^{th} observation

= Value of 2^{nd} observation + 0.75(Value of 3^{rd} observation − Value of 2^{nd} observation)

= 25 + 0.75(30 − 25)

= 25 + 0.75 × 5

= 28.75

Q_3 = Value of $\frac{3(10+1)}{4}$ th observation

= Value of 8.25^{th} observation

= Value of 8^{th} observation + 0.25(Value of 9^{th} observation − Value of 8^{th} observation)

= 65 + 0.25(70 − 65)

= 65 + 0.25 × 5

= 66.25

$\therefore$ Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{66.25 - 28.75}{2} = \frac{37.5}{2} = 18.75$

Example 2:

Find the quartile deviation for the following data.

Age of workers	25	35	45	55	65

Number of workers	8	12	20	16	4

Solution:

Here, the data is in ascending order. First, prepare the table with cumulative frequencies(less than type).

Age of workers(x_i)	Number of workers(f_i)	Cumulative frequency
25	8	8
35	12	20 $\leftarrow Q_1$
45	20	40
55	16	56 $\leftarrow Q_3$
65	4	60
Total	$N = 60$	

Here, $N = 60$

$$\frac{N+1}{4} = \frac{60+1}{4} = \frac{61}{4} = 15.25$$

$$\therefore Q_1 = 35$$

$$\frac{3(N+1)}{4} = \frac{3(60+1)}{4} = \frac{183}{4} = 45.75$$

$$\therefore Q_3 = 55$$

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2} = \frac{55-35}{2} = \frac{20}{2} = 10$$

Example 3:

Compute the quartile deviation for the following data.

C.I.	10-19	20-29	30-39	40-49	50-59
Frequency	3	4	6	5	2

Solution:

In the given data, the classes are discontinuous. First, convert the classes into continuous classes and prepare the table with cumulative frequencies(less than type).

C.I.	Continuous classes	Frequency	Cumulative frequency
10-19	9.5-19.5	3	3
20-29	19.5-29.5	4	7 ← Q_1
30-39	29.5-39.5	6	13
40-49	39.5-49.5	5	18 ← Q_3
50-59	49.5-59.5	2	20
Total		N = 20	

Here, N = 20

$$\therefore \frac{N}{4} = \frac{20}{4} = 5$$

So, Q_1 lies in the class 19.5-29.5.

$\therefore$ L = 19.5, f = 4, h = 10 and c.f. = 3

$$\Rightarrow Q_1 = L + \left(\frac{\frac{N}{4} - \text{c.f.}}{f}\right) \times h = 19.5 + \left(\frac{5-3}{4}\right) \times 10 = 19.5 + 5 = 24.5$$

Also, $\frac{3N}{4} = \frac{3 \times 20}{4} = 15$

So, Q_3 lies in the class 39.5-49.5.

$\therefore$ L = 39.5, f = 5, h = 10 and c.f. = 13

$$\Rightarrow Q_3 = L + \left(\frac{\frac{3N}{4} - \text{c.f.}}{f}\right) \times h = 39.5 + \left(\frac{15-13}{5}\right) \times 10 = 39.5 + 4 = 43.5$$

$\therefore$ Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{43.5 - 24.5}{2} = \frac{19}{2} = 9.5$

Mode Of A Data Set

Bhangra's is a very popular shop that sells watches of foreign brands in Delhi's posh Connaught Place market. The owner of the shop decided to stock the brand whose watches were selling the most. He decided to look at the previous month's sales, which is listed as

Brand	Number of watches sold
Alpha	17

Townzen	23
Tag Heuim	7
Twatch	13

Based on this information, the owner decided to stop keeping Tag Heuim watches because very few people buy them. He also decided to keep more varieties of Townzen watches because most people were buying this brand.

In this data set, the highest occurring event (23) corresponds to Townzen watches and is known as the **mode** of this data set. Just like mean and range, mode is another measure of central tendency of a group of data. It can be defined as

The value of a set of data that occurs most often is called the mode of the data.

Here, the data set did not contain too many terms and could thus be easily arranged in ascending order. However, in case of very large data, it is not always easy to arrange it in ascending or descending order. Therefore, in such cases, it is better to arrange the data set in the form of a table with tally marks.

A collection of data can have more than one mode. The data sets having one mode or two modes or more than 2 modes are said to have **uni-mode** or **bi-mode** or **multi-mode**.

Let us now look at some more examples to understand this concept better.

Example 1:

Find the mode of the following numbers.

2, 6, 7, 5, 4, 2, 6, 7, 9, 7, 8, 3, 2, 11

Solution:

The given set of numbers can be arranged in ascending order as

2, 2, 2, 3, 4, 5, 6, 6, 7, 7, 7, 8, 9, 11

Here, 7 and 2 occur most often (3 times). Therefore, both 7 and 2 are the modes of the given set of numbers.

Example 2:

Find the mode of the following data set.

1000, 200, 700, 500, 600, 160, 270, 300, 360, 950

Solution:

The increasing order of the given numbers is

160, 200, 270, 300, 360, 500, 600, 700, 950, 1000

Here, every number is occurring only once.

Thus, the given data has no mode.

Note: The above example shows that the mode of a data may or may not be unique. Also, there are some data sets which do not have any mode.

Example 3:

Determine whether the data 35, 30, 32, 35, 40, 30, 25, 30, 22, 30 has uni-mode, bi-mode or multi-mode.

Solution:
From the data, we observe that 30 repeats maximum times (4 times). Thus, the mode of this data is 30. Since there is only one mode of this data, so the given data has uni-mode.

Example 4:
Find the type of mode of the data 25, 23, 23, 25, 27, 26, 23, 24, 23, 25, 28, 25.

Solution:
From the data, we observe that 23 and 25 repeats maximum times (4 times). Thus, the mode of this data is 23 and 25. Since there are two modes of this data, so this data has bi-mode.

Example 5:
Which type of mode is represented by the data as shown below?

Data	Frequency
0	5
8	6
16	9
24	9
32	9
40	8

Solution:

From the frequency table, we observe that 16, 24 and 32 repeats maximum times (9 times). Thus, the mode of this data is 16, 24, and 32. Since there are three modes of this data, so this data has multi-mode.