Measures of Central Tendency

Measures of Central Tendency-Meaning and Objectives

Objective

In this lesson, we will understand the meaning and objectives of central tendency.

Introduction

In the previous few lessons, we studied about the techniques of collecting, organising and presenting the data. The next step is to analyse the data. A measure of central tendency is one of the important tools of statistical analysis. Central tendency refers to a central value or a representative value of a statistical series.

The central value represents the entire data in the sense that the values of observations in the data lie close to the central value. Thus, central value can be said to represent the entire data set. A central value makes the raw data easy to understand and analyse. For example, suppose one is asked about the wages of workers in a factory.

Now, if as an answer to this question the wages of all the workers in the factory are presented, then it would be very difficult and tedious to draw any meaningful conclusion. On the other hand, if a single value of wage is ascertained that represents the wages of all the workers in the factory, then a clearer conclusion can be drawn. Thus, a central value provides a brief description of the data set and makes it more understandable.

Measures of central tendency- Meaning and Objectives

Central tendency refers to a central value that represents the entire statistical series. The central value is sometimes simply referred as average. The following are the objectives of a measure of central tendency.

1. Provides concise information: A central value presents a brief and concise information about the entire data and makes it more understandable. It provides a systematic and quick view of the data set. For example, average marks of all the students in a class is said to represent the marks of all the students and provides a view about the performance of the class.

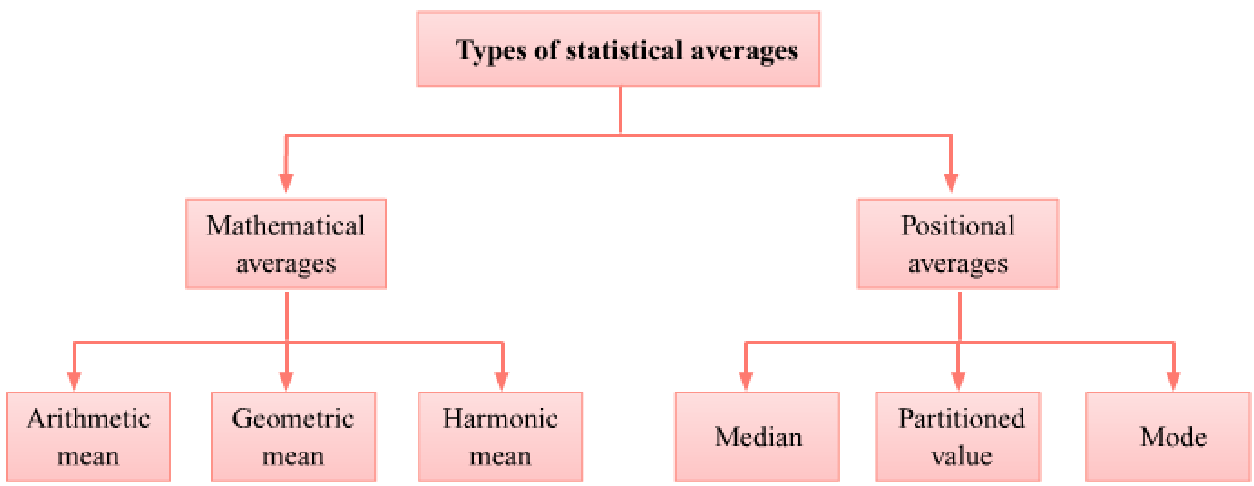
2. Enables comparative studies: The measures of central tendency facilitates easy comparison between two or more data set. For example, suppose the marks of students of three classes are to be compared. The comparison would become much easier if the marks of each class are represented by a single value. On the other hand, a comparison of marks with the raw data of would be quite a difficult task.

3. Facilitates formulation of policies: The average value of the various economic variables such as income, price, etc. helps in the formation of suitable economic policies. For example, if the per capita income (average income) is found to be low, then it becomes quite clear that such policies are to be formed that aim at increasing the per capita income.

4. Helps in establishing relationship: With the help of measures of central tendency, we can establish the relationship between two or more variables. This is because with average values it becomes easier to estimate their exact relationship. For example, the average marks of students and the average hours of their T.V. viewing can be compared. Accordingly, it can be judged whether T.V. viewing affects the performance of the students.

Types of Statistical Averages

The following table presents the different types of statistical averages.



In the subsequent lessons, we will study about arithmetic mean, median and mode.

Arithmetic Mean- Meaning, Properties and Computation Methods

Objective

In this lesson, we will understand the calculation of arithmetic mean in different kinds of series namely, individual series, discrete series and continuous series using the following three methods.

- Direct Method
- Short-Cut Method or the Assumed Mean Method

- Step-Deviation Method

In addition, we will also understand the following concepts.

- Properties of Mean
- Merits and Demerits of Mean
- Weighted Arithmetic Mean
- Calculation of Combined Arithmetic Mean
- Calculation of Corrected Arithmetic Mean

Arithmetic Mean

Arithmetic mean is the most widely used and simplest of the measures of central tendency. It is simply an average of all the items in a series. It is denoted by $\overline{X}$. Algebraically, it is obtained by dividing the sum of the values of all the items in the series by the total number of items.

$$\text{Arithmetic Mean} = \frac{\text{Total Value of the Observations}}{\text{No. of Observations}}$$

Arithmetic mean can be classified into the following two categories.

i. Simple Arithmetic Mean or Average

ii. Weighted Arithmetic Mean or Weighted Average

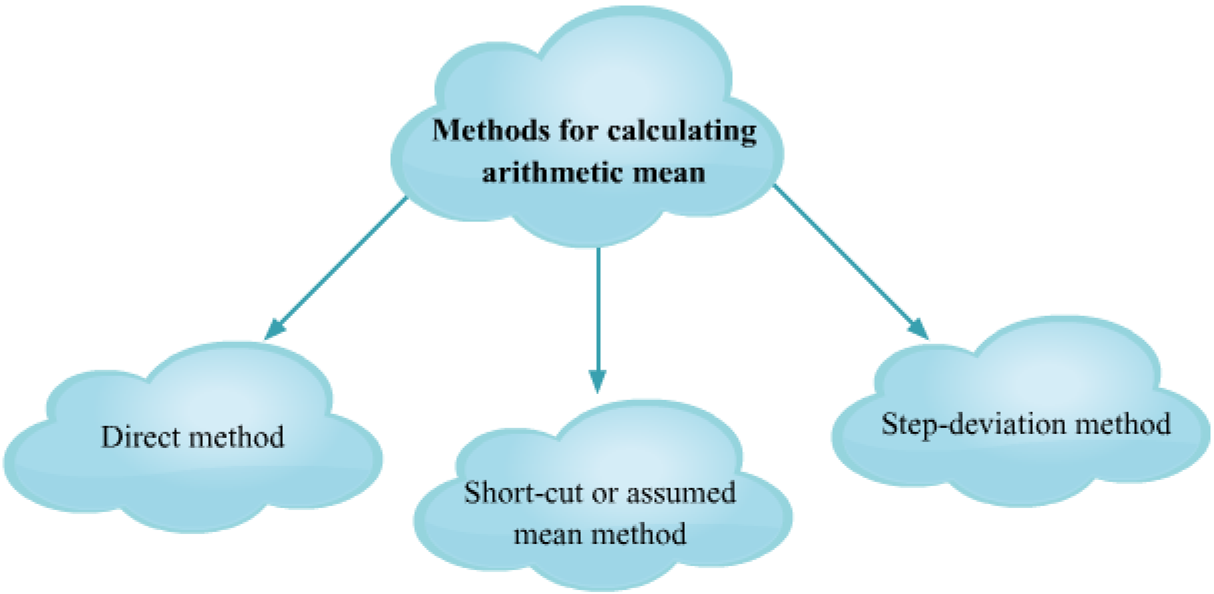
Calculation of Simple Arithmetic Mean

Arithmetic Mean can be calculated using either of the following three methods:

i. Direct Method

ii. Short-Cut or Assumed Mean Method

iii. Step-Deviation Method



Let us understand the estimation of arithmetic mean (A.M.) in the three kinds of series namely, individual series, discrete series and continuous series.

Calculation of Simple A.M. in Individual Series

(i) *Direct method*

Under direct method, A.M. is calculated by simply dividing the the sum of values of all the items in the series by the total number of items in the series.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots\ldots + X_n}{N}$$ or, $$\overline{X} = \frac{\Sigma X}{N}$$

where,
$\overline{X}$ represents Arithmetic Mean
ΣX represents sum of all observations
N represents number of observations

<u>Example</u>: Following are the marks of ten students in a class.

25, 31, 14, 16, 28, 45, 38, 47, 41, 29

Calculate the mean marks for the class.

Solution:

$$\text{Mean} = \overline{X} = \frac{\Sigma X}{N} = \frac{25+31+14+16+28+45+38+47+41+29}{10} = 31.4$$

Thus, the mean marks is 31.4

(ii) *Short-Cut or Assumed mean method*

The direct method of calculation of mean becomes tedious if, the number of items in the series is large and the value of each item is high. In such cases, to simplify the calculation process, short-cut or the assumed mean method is used.

The following steps are involved in the calculation of mean using the assumed mean method.

***Step* 1**: From the given values of items, decide an assumed mean.

***Step* 2**: Take deviation of each item from the assumed mean as (*X*-*A*) and denote by '*d*'.

***Step* 3**: Obtain the sum of the deviations by- *Σd.*

***Step* 4**: Mean is calculated using the following formula.

$$\overline{X} = A + \frac{\Sigma d}{N}$$

where,
A represents assumed mean

Σd represents sum of deviation of values from the assumed mean

N represents number of items in the series
d = (*X* – *A*)

Example: Continuing with the illustration given above, let us calculate the value of mean using the assumed mean method.

Solution
Here, we take 25 as the assumed mean. So, we take deviations of each item in the series from 25.

Student	**Marks (*X*)**	***X* – *A* (*d*)**
A	25 = *A*	0
B	31	6
C	14	– 11

D	16	– 9
E	28	3
F	45	20
G	38	13
H	47	22
I	41	16
J	29	4
N = 10		*Σd* = 64

$$\overline{X} = A + \frac{\Sigma d}{N} = 25 + \frac{64}{10}$$
$$= 25 + 6.4 = 31.4$$

Thus, mean marks is 31.4

(iii) *Step-Deviation method*

The step-deviation method further simplifies the calculation procedure. Under this method, the value of the deviations is further lowered by dividing each deviation by a common factor. The following steps are involved in the calculation of mean using the step-deviation method.

***Step* 1**: From the given value of items, decide an assumed mean.

***Step* 2**: Take deviation of each item from the assumed mean as (*X*-*A*) and denote it by *d.*

Step* 3**: Divide each deviation obtained in ***step 2 by a common factor and denote it by *d'*.

***Step* 4**: Obtain the sum of step deviations as *Σd'.*

***Step* 5**: Mean is calculated using the following formula:

$$\overline{X} = A + \frac{\sum d'}{N} \times i$$

where,
A represents assumed mean

Σd' represents sum of step deviation of values from the assumed average

N represents number of items in the series

$$d' = \frac{X - A}{\text{Common Factor}}$$

i represents common factor

Example: Consider again the illustration given above. Let us now calculate the mean using the step-deviation method.

Solution

Let us take the common factor as 5 and divide each deviation by 5.

Students	**Marks (X)**	**$X – A$ (d)**	$\frac{(X-A)}{i}$ **= (d')**
A	25 = A	0	0
B	31	6	1.2
C	14	– 11	– 2.2
D	16	– 9	– 1.8
E	28	3	0.6
F	45	20	4
G	38	13	2.6
H	47	22	4.4
I	41	16	3.2
J	29	4	0.8
N = 10			$\Sigma d'$ = 12.8

$$\overline{X} = A + \frac{\sum d'}{N} \times i$$

$$= 25 + \frac{12.8}{10} \times 5 = 31.4$$

Note 1: The only difference between the calculation procedure of the step-deviation method and the short-cut method is that under the step-deviation method, the deviations are divided by a common factor. Accordingly, the formula for mean also changes slightly and we take sum of step deviations instead of simple deviations. The ratio of the sum of step deviations and number of items $\left(\frac{\Sigma d'}{N}\right)$ is then multiplied by the common factor.

Note 2: Irrespective of the method used, the value of mean obtained is the same.

Calculation of Simple A.M. in Discrete Series

Like individual series, in a discrete series also, the arithmetic mean can be calculated using either of the three methods namely, direct method, short-cut method or step-deviation method.

(i) *Direct method*

The following steps are involved in the calculation of mean in discrete series using the direct method.

***Step* 1**: Obtain the product of the frequencies and the observations.

Step* 2**: Obtain the sum of the product obtained in ***step 1.

***Step* 3**: The mean is calculated using the following formula.

$$\overline{X} = \frac{\Sigma fX}{\Sigma f}$$

where,

ΣfX represents summation of the product of frequency and *X* values.

Σf represents sum total of frequencies.

<u>Example</u>: The following table presents the expenditure of 10 household. Calculate the mean expenditure.

Expenditure (*X*)	***f***	***fX***
1,000	2	2,000
1,200	2	2,400
1,500	3	4,500
1,800	1	1,800
2,000	2	4,000
	$N = \Sigma f = 10$	$\Sigma fX = 14,700$

Solution

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{14,700}{10} = 1,470$$

Thus, the average expenditure incurred by the household is Rs 1,470.

(ii) *Short-Cut Method*

The following steps are involved in the calculation of mean using the short-cut method.

***Step* 1**: Like the method in individual series, we first decide an assumed mean.

***Step* 2**: From the assumed mean, take deviation of each item as (*X*-*A*) and denote by *d*.

***Step* 3**: Multiply each deviation by the corresponding frequencies and obtain the sum as *Σfd.*

***Step* 4**: Mean is calculated using the following formula.

$$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$$

where,
A represents assumed mean

d = *X* – *A*
Σfd represents summation of the product of frequency and deviations

Σf or *N* represents sum total of frequency

<u>Example</u>: Calculate the mean expenditure of the households in the above illustration using the short-cut method.

Solution

Let us take 1,000 as the assumed mean and obtain deviation of each observation from 1,000.

Expenditure (*X*)	***f***	***X* – *A* = *d***	***fd***
1,000 = *A*	2	0	0
1,200	2	200	400
1,500	3	500	1,500
1,800	1	800	800

2,000	2	1,000	2,000
	N = 10		*Σfd* = 4,700

$$\overline{X}=A=\frac{\Sigma fd}{N}$$

$$=1,000+\frac{4,700}{10}=1,470$$

Thus, the mean expenditure is Rs 1,470.

(iii) *Step-Deviation Method*

The following steps are involved in the calculation of mean using the step-deviation method.

***Step* 1**: From the given values of the items, assume a mean.

***Step* 2**: Take deviation of each item from the assumed mean and denote it by *d.*

Step* 3**: Divide each deviation obtained in ***step 2 by a common factor to obtain the step-deviations denoted by (*d*').

***Step* 4**: Multiply each step deviation by the corresponding frequencies and obtain the sum denoted by *Σfd'.*

***Step* 5**: Mean is calculated using the following formula.

$$\overline{X}=A+\frac{\Sigma fd'}{\Sigma f}\times i$$

where,

$$d'=\frac{X-A}{i}$$

i represents common factor

<u>Example</u>: Consider again the illustration of the expenditure of households. Let us calculate the mean expenditure using the step-deviation method.

Households (*X*)	***f***	***X* – *A* (*d*)**	$\frac{(X-A)}{i}$ **(*d'*)**	***fd'***
1,000 =*A*	2	0	0	0
1,200	2	200	2	4

1,500	3	500	5	15
1,800	1	800	8	8
2,000	2	1,000	10	20
	$\Sigma f = N = 10$			$\Sigma fd' = 47$

Solution

$$\overline{X} = A + \frac{\sum fd'}{N} \times i$$
$$= 1000 + \frac{47}{10} \times 100 = 1,470$$

Like the direct method and the short-cut method, the mean expenditure calculated using the step-deviation method is Rs 1,470.

Calculation of Simple A.M. in Continuous Series

The calculation procedure for the calculation of mean under the continuous series is the same as under discrete series. The only difference being that in the continuous series, we take the mid-value of each class interval in place of the individual values.

Thus, as a first step in the calculation of mean in the continuous series, we calculate the mid-value of each class interval. Then, the same calculation procedure is followed as under discrete series treating the mid-values as the individual observations.

(i) *Direct method*

In the direct method, the following formula is used for the calculation of mean.

$$\overline{X} = \frac{\Sigma fm}{\Sigma f}$$

where,
Σfm represents sum of the product of frequency and mid-values
Σf represents sum total of frequency

m represents mid values

$$m = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

<u>Example</u>: The following data presents the temperature of 15 cities. Calculate the mean temperature of the cities.

Temperature	Number of Cities
0°-5°	1
5°-10°	3
10°-15°	5
15°-20°	4
20°-25°	6

Solution

Temperature	Number of Cities	Mid-Points (*m*)	*fm*
0°-5°	1	2.5	2.5
5°-10°	2	7.5	15
10°-15°	5	12.5	62.5
15°-20°	4	17.5	70
20°-25°	3	22.5	67.5
	N = 15		217.5

$$\overline{X} = \frac{\Sigma fm}{N} = \frac{217.5}{15} = 14.5°\text{C}$$

Thus, the average temperature is 14.5° C.

(ii) *Short-Cut Method*

The following is the formula for the calculation of mean using the short-cut method.

$$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$$

where,
$d = (m - A)$
A represents assumed mean

Σfd represents sum of the product of frequency and deviations
Σf or N represents sum total of frequency

<u>Example</u>: Let us calculate the, mean temperature of the cities as given in the above illustration using the short-cut method.

Solution

Here, the assumed mean is decided from the mid-points. Let us take 2.5 as the assumed mean. So, we calculate the deviations of the mid-values from 2.5

Temperature (*X*)	**Number of Cities (*f*)**	**Mid-Point (*m*)**	***(m*– A) (*d*)**	***fd***
0°-5°	1	2.5 = *A*	0	0
5°-10°	2	7.5	5	10
10°-15°	5	12.5	10	50
15°-20°	4	17.5	15	60
20°-25°	3	22.5	20	60
	N = 15			∑*fd* = 180

$$\overline{X} = A + \frac{\Sigma fd}{N}$$
$$= 2.5 + \frac{180}{15} = 14.5°$$

(iii) *Step-Deviation Method*

The following is the formula for the calculation of mean using the step-deviation method.

$$\overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times i$$

where,

$\Sigma f d'$ = Sum of the product of frequency and condensed deviations

$$d' = \frac{m-A}{i}$$

<u>Example</u>: Consider again the illustration given above. Let us calculate the mean temperature using the step-deviation method.

Temperature	**No. of cities (*f*)**	**Mid-point (*m*)**	**(*m* – A) (*d*)**	$\frac{(X-A)}{i}$ **(*d'*)**	***fd'***
0°-5°	1	2.5= *A*	0	0	0

5°-10°	2	7.5	5	1	2
10°-15°	5	12.5	10	2	10
15°-20°	4	17.5	15	3	12
20°-25°	3	22.5	20	4	12
	N = 15				*Σfd'* = 36

Solution

$$\overline{X} = A + \frac{\sum fd'}{N} \times i$$
$$= 2.5 + \frac{36}{15} \times 5 = 14.5$$

Properties of Arithmetic Mean

Arithmetic mean exhibits certain important properties. The following are some of the properties of A.M.

(i) The sum of deviations taken of all the values of a distribution from their arithmetic mean is always zero.

i.e. $\Sigma\left(X_i - \overline{X}\right) = 0$

where,

X_i refers to all the values of a distribution

$\overline{X}$ refers to the arithmetic mean.

***<u>Example</u>*- Consider the following data.

X	2	4	6	8	10

$$\text{Mean}, \overline{X} = \frac{2+4+6+8+10}{5} = 6$$

Now taking deviation of each value of *X* from the mean

X	$X-\overline{X}$
2 4 6 8 10	– 4 – 2 0 2 4
	$\Sigma\left(X-\overline{X}\right)=\text{Zero}$

Note: In case of discrete and continues series, $\Sigma f\left(X_i-\overline{X}\right)$ is zero.
Example: Consider the following series.

X	f	fX	$(X-\overline{X})$	$f(X-\overline{X})$
2	3	6	–4	–12
5	1	5	–1	–1
7	3	21	1	3
9	2	18	3	6
10	1	10	4	4
		ΣfX= 60		$\sum f(X-\overline{X})=\text{Zero}$

$$\overline{X}=\frac{\sum fX}{N}=\frac{60}{10}=6$$

(ii) The total of the squared deviations from the arithmetic mean is minimal. In other words, if the deviations are taken from the arithmetic mean and squared then, the sum obtained is less than what would have been obtained had the deviations been taken from any other value.

***<u>Example</u>*- Consider the following data.

X	5	10	15	20	25

$$\text{Mean}, \overline{X}=\frac{5+10+15+20+25}{5}=15$$

Let us now take the deviation of the values from mean as well as from another value say, 10 and compare the result.

Deviation taken from mean	**Deviation taken from 10**

X	$(X-\overline{X})$	$(X-\overline{X})^2$	X	**(X – 10)**	**(X – 10)²**
5	– 10	100	5	– 5	25
10	– 5	25	10	0	0
15	0	0	15	5	25
20	5	25	20	10	100
25	10	100	25	15	225
		$(\Sigma X-\overline{X})^2=250$			$(\Sigma X - 10)^2 = 375$

Clearly, 250 < 375

or, $\Sigma(X-\overline{X})^2 < \Sigma(X-10)^2$

Hence, the sum of the squared deviation from mean is less than the sum of squared deviations from any other value.

(iii) If constant value is added (or subtracted) to all the values of the series, the arithmetic mean will also increase (or, decrease) by the same constant factor.

Example- Consider the following data.

X	2	4	6	8	10	12

$$\text{Mean}, \overline{X} = \frac{2+4+6+8+10+12}{6} = 7$$

Let us now add 2 to each value and calculate mean.

X	(X + 2)
2	4
4	6
6	8
8	10
10	12
12	14

	$\Sigma(X + 2) = 54$

$$\text{Mean}, \overline{X} = \frac{54}{6} = 9$$

Change in mean = (9 – 7) = 2

Thus, mean increases by 2. In other words, on adding 2 to value of each item the mean also increases by 2.

Similarly, let us subtract 2 from each value and calculate mean.

X	**(*X* + 2)**
2	
4	2
6	4
8	6
10	8
12	10
	$\Sigma\,(X + 2) = 30$

$$\text{Mean}, \overline{X} = \frac{30}{6} = 5$$

Thus on subtracting 2 from each item, the mean also decreases by 2 (i.e.7 – 5 = 2).

(iv) Similarly, if all values are multiplied by some constant number, the arithmetic average will also be multiplied by the same factor.

Example: Continuing with the illustration given above, let us multiply each observation by 2 and calculate the mean.

X	***X* × 2** **(*x*)**
2	4
4	8
6	12
8	16
10	20
12	24
	$\Sigma x = 84$

$$\text{Mean} = \frac{\Sigma x}{6} = \frac{84}{6} = 14$$

Thus, we see that on multiplying value of each item by 2, the resultant mean value also gets multiplied by 2.

Merits of Arithmetic Mean

i. Of all the measures of central tendency, it is the easiest and simplest to calculate.

ii. It is based on each and every observation in the series.

iii. It gives equal importance to all the items of a series.

iv. It is not positional like median and mode. In fact, it is a calculated value.

v. It is free from any personal bias.

vi. It can be used for mathematical calculations, such as, multiplication, division, addition, etc.

vii. It acts as a good base for comparison between two or more series.

viii. It is a stable measure of central tendency. In other words, it is minimally affected by the change in the series.

Demerits of Arithmetic Mean

i. Arithmetic mean is most affected by the presence of extreme items. In other words, it is easily distorted by the extreme values.

ii. It cannot be calculated for an open-ended series.

iii. It cannot be ascertained graphically.

iv. It can sometimes be misleading and give absurd results.

v. Sometimes, the computed value of the mean may not be from the given series.

Weighted Arithmetic Mean

We know that simple A.M. gives equal weightage to all the osbervations in the data set. However, sometimes it is required to accord different weights to different observations based on their relative importance.

For example, in the analysis of the mean prices of different consumer items, assigning higher weights to the more important items such as food, clothing etc. and lower weights to the lesser important ones such as cosmetics, stationary etc.

Would provide more meaningful results. Similarly, in the analysis of the household expenditure, higher weights can be accorded to those expenditures that are comparatively more important such as housing, food and fuel. In such cases, for the estimation of average we use the weighted arithmetic mean.

The formula for calculating weighted arithmetic mean is

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W}$$

Where, $\overline{X}_w$ is the weighted arithmetic mean

X refers to the variables

and, *W* represents the weights assigned.

Example- Calculate weighted arithmetic mean for the following data.

X	5	10	15	20	25
W	3	1	4	5	2

Solution

X	***W***	***WX***
5	3	15
10	1	10
15	4	60
20	5	100
25	2	50
	$\Sigma W = 15$	$\Sigma WX = 235$

$$\text{Weighted Arithemetic Mean, } \overline{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{235}{15} = 15.66$$

Thus, weighted arithmetic mean is 15.66.

Example- The following table presents the marks of two students in different subjects. Identify the one who gets the highest marks.

Subject	Weight	Marks of student 1	Marks of student 2
Physics	4	50	45
Chemistry	1	45	50
Math	2	38	40
English	3	35	35

Solution

To identify which of the two gets highest marks, we calculate the weighted mean marks for each student.

Weighted mean marks for student 1

Subject	Weight (W)	Marks (X)	WX
Physics	4	50	200
Chemistry	1	45	45
Math	2	38	76
English	3	35	105
	$\Sigma W = 10$		$\Sigma WX = 426$

$$\text{Weighted Arithemetic Mean}, \overline{X}_w = \frac{\sum WX}{\sum W} = \frac{426}{10} = 42.6$$

Weighted mean marks for student 2

Subject	Weight (W)	Marks (X)	WX
Physics	4	45	180
Chemistry	1	50	50
Math	2	40	80
English	3	35	105
	$\Sigma W = 10$		$\Sigma WX = 415$

$$\text{Weighted Arithemetic Mean}, \overline{X}\text{w} = \frac{\sum WX}{\sum W} = \frac{415}{10} = 41.5$$

Since, weighted mean marks of student 1 is greater than the weighted mean marks of student 2, student 1 obtains greater marks than students 2.

Miscellaneous Illustrations

Calculation of Combined A.M.

The combined arithmetic mean for two or more series can be calculated using the following formula.

$$\overline{X}_{1,2} = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2}{N_1 + N_2}$$

where,

$\overline{X}_{1,2}$ represents combined arithmetic mean of parts 1,2 of the series

$\overline{X}_1$ represents arithmetic mean of part 1 of the series

$\overline{X}_2$ represents arithmetic mean of part 2 of the series

N_1 represents number of items of series 1

N_2 represents number of items of series 2

***Example*-** The following information represents the works of students in two classes A and B. Calculate the combined mean marks of students of two classes.

	Mean marks	**Number of students**
Class A	30	45
Class B	35	40

Solution

We know,

$$\overline{X}_{12} = \frac{\overline{X}_1 N_1 + \overline{X}_2 N_2}{N_1 + N_2}$$

Here,

$\overline{X}_1 = 30$

$\overline{X}_2 = 35$

$N_1 = 45$

$N_2 = 40$

Substituting the given values in the formula,

$$\text{Combined Arithemetic Mean}, \overline{X}_{12} = \frac{30\times45+35\times40}{45+40}$$
$$=32.35$$

Thus, the combined mean marks of the two classes are 32.35

***Example*-** The mean weight of 60 children in locality A is 20. There are 40 children in locality B. Calculate the mean weight of children in locality B if the combined mean is 22.

Solution

Here the given information can be summarised as follows.

Locality	Mean	Number of Children
A	$\left(\overline{X}_1\right)=20$	$N_1 = 60$
B	$\left(\overline{X}_1\right)=?$	$N_2 = 40$

We know,

$$\overline{X}_{12} = \frac{N_1\overline{X}_1+N_2\overline{X}_2}{N_1+N_2}$$
$$\text{or, } \frac{60\times20+40\times\overline{X}_2}{60+40}$$
$$\text{or, } 2,200=1,200+40\overline{X}_2$$
$$\text{or, } \overline{X}_2=25$$

Thus mean weight of students in locality B is 25.

Example- There are 100 students in a class of a school. The students are divided into sections A and B. The mean marks of students in section A is 14 and mean marks of students in section 13 is 12. Find the number of students in section A if, combined mean marks is 13.

Solution

The given information can be summarised as follows.

Section	Mean Marks	Number of Children

A	$\left(\overline{X}_1\right)=14$	N_1 =?
B	$\left(\overline{X}_1\right)=12$	N_2 =?

$N_1 + N_2 = 100$

We know,

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$
$$\text{or, } 13 = \frac{N_1 \times 14 + N_2 \times 12}{100}$$
$$\text{or, } 1300 = 14N_1 + 12N_2$$
$$\text{or, } 14N_1 + 12N_2 = 1300$$
$$\text{or, } 7N_1 + 6N_2 = 650 \qquad \text{......(1)}$$

Also, $N_1 + N_2 = 100$ …..(2)
Solving equation (1) and (2) simultaneously

$7N_1 + 6N_2 = 650$
$(N_1 + N_2 = 100) \times 7$

$$\begin{array}{ccccc} 7N_1 & + & 6N_2 & = & 650 \\ 7N_1 & + & 7N_2 & = & 700 \\ (-) & (-) & (-) & & \\ \hline & & -N_2 & = & -50 \\ & & N_2 & = & 50 \end{array}$$

Substituting value of N_2 in equation (2)
$N_1 + 50 = 100$
or, $N_1 = 50$

Thus, there are 50 students in section A.

Calculation of Corrected A.M.

The correct A.M. is calculated using the following formula.

$$\overline{X} = \frac{\Sigma X(\text{Wrong Total}) + \text{Correct Value} - \text{Incorrect Value}}{N}$$

***Example*-** The arithmetic mean of 10 observations in the series was 200. Later, it was found that one of the items in the series was misread as 30 instead of 25. Calculate the correct arithmetic mean.

$$(\overline{X})_{\text{Wrong}} = \frac{\Sigma X}{N}$$
$$200 = \frac{\Sigma X}{10}$$

$(\Sigma X)_{\text{Wrong}} = 2{,}000$

$$(\overline{X})_{\text{Correct}} = \frac{\Sigma X_{(\text{Wrong})} + \text{Correct value} - \text{Incorrect Value}}{N}$$
$$= \frac{2,000 + 25 - 30}{10}$$
$$= 19.95$$

Hence, the correct arithmetic mean is 19.95.

***Example*-** The mean of 20 items in a series was 300. Later, it was found that two items were misread as 50 and 40 instead of 60 and 55 respectively. Calculate the correct arithmetic mean of the series.

Solution

$$(\overline{X})_{\text{Wrong}} = \frac{\Sigma X}{N}$$
$$\text{or, } 300 = \frac{\Sigma X}{20}$$

$(X)_{\text{Wrong}} = 6{,}000$

$$(\overline{X})_{\text{Correct}} = \frac{(\Sigma X)_{\text{Wrong}} + \text{Correct Value} - \text{Incorrect Value}}{N}$$
$$= \frac{6,000 + 60 + 55 - 50 - 40}{20}$$
$$= 301.25$$

Thus, the correct arithmetic mean is 301.25.

Requisites of a Good Average

A good average must posses certain characteristics. The following are some of the important requisites of a good average.

(i) ***Simple to calculate and easy to understand***: An average should be easy to calculate and should not involve any tedious and lengthy calculations. In addition, it should be easy to understand by the masses.

(ii) ***Based on all the observations***: It should be based on all the observations in the data. If it is based on only a few items in the series then, it cannot be said to appropriately represent the entire data set.

(iii) ***Capable of algebraic treatment***: An average should be capable of further algebraic treatment so as to facilitate further analysis of the data.

(iv) ***Least affected by extreme values***: It should not be unduly affected by extreme values in the series. In other words, too large values and too small values in the series should not affect the estimation of average.

(v) ***Least affected by fluctuations in sample***: An average should not be affected much by the fluctuations in sample. In other words, if two or more samples are taken from a universe, then there should not be much difference in the average.

Median- Meaning, Properties and Computation Methods

Introduction

In this lesson, we will study the calculation of median under different series namely individual series, discrete series and continuous series. In addition to this, we will study the graphical location of median.

Median- Meaning

In the previous lesson, we studied about the arithmetic mean which is a mathematical average. In this lesson, we will study about one of the positional average, median. Median refers to the value that divides a series into two equal parts. It is a centrally located value in the sense that half of the observations in the series lie above the median and the rest half lie below the median.

As against the A.M., the median is not based on all items in the series. Irrespective of the values of the observations in the series, median is just the centrally located or the middle value. According to the concept of median, the middle value is able to represent the entire data set.

Calculation of Median in Individual Series

The individual series can have either odd number of items or even number of items. The following are the steps involved in the calculation of median in individual series.

***Step* 1**: Arrange different items of the series either in ascending or descending order.

***Step* 2**: Count the number of items in the series and denote it by *N*.

***Step* 3**: If number of items in the series is odd, then the following formula is used to calculate the median.

$$M = \text{Size of } \left(\frac{N+1}{2}\right)^{\text{th}} \text{item}$$

OR

If number of items in the series is even then, the following formula is used to calculate the median.

$$M = \frac{\text{Size of } \left(\frac{N}{2}\right)^{\text{th}} \text{item} + \text{Size of } \left(\frac{N}{2}+1\right)^{\text{th}} \text{item}}{2}$$

Example: The following table presents the runs scored by 11 players in a cricket match. Calculate the median score.

Cricketer	Runs Scored
Sameer	101
Devender	53
Ajay	74
Amit	28
Rohit	0
Mohit	13
Rajender	34
Rakesh	25
Mukesh	15
Surender	0
Rohan	21

Solution

Arranging the data in descending order

Cricketer	Runs Scored
Samer	101
Ajay	74
Devender	53

Rajender	34
Amit	38
Rakesh	25
Rohan	21
Mukesh	15
Mohit	13
Rohit	0
Rohan	0

$$\text{Median} = \text{Size of} \left(\frac{N+1}{2}\right)^{th} \text{item} = 6^{th} \text{ item} = 25$$

Thus, Rakesh's score is the median which is 25.

Example- The following data presents the marks of 8 students in a class. Calculate the median.

S. No.	1	2	3	4	5	6	7	8
Marks	25	20	35	22	28	30	27	29

Solution

Arranging the given data in ascending order

S. No.	1	2	3	4	5	6	7	8
Marks	20	22	25	27	28	29	30	35

Here, the number of observation is even

$$\text{Median} = \frac{\text{Size of } \left(\frac{N}{2}\right)^{th} \text{ item} + \text{Size of } \left(\frac{N}{2}+1\right)^{th} \text{ item}}{2}$$

$$= \frac{\text{Size of } \left(\frac{8}{2}\right)^{th} \text{ item} + \text{Size of } \left(\frac{8}{2}+1\right)^{th} \text{ item}}{2}$$

$$= \frac{\text{Size of } 4^{th} \text{ item} + \text{Size of } 5^{th} \text{ item}}{2}$$

$$= \frac{27+28}{2} = 27.5$$

Thus, median is 27.5

Calculation of Median in Discrete Series

The following steps are involved in the calculation of median in discrete series.

***Step* 1**: Arrange the data in either ascending or descending order.
***Step* 2**: Convert simple frequencies into cumulative frequencies.
***Step* 3**: Add the simple frequency $\sum f$ as N.
***Step* 4**: Determine the value of (N + 12)N + 12^{th} item
***Step* 5**: The value of X corresponding to (N + 12)N + 12^{th} item is the median value.

<u>Example</u>: For the following data, calculate the value of median.

Items (*X*)	**Frequency (*f*)**
10	3
15	7
20	5
25	4
30	6

35	2
40	3
	$\sum f = 30$

Solution

The given data is already arranged in the ascending order.

Item (*X*)	**Frequency (*f*)**	**Cumulative Frequency (*f*)**
10	3	3
15	7	10
20	5	15
25	4	19
30	6	25
35	2	27
40	3	30
	$\sum f = 30$	

$$M = \text{Size of } \left(\frac{N+1}{2}\right)^{\text{th}} \text{item}$$
$$= \text{Size of } \left(\frac{30+1}{2}\right)^{\text{th}} \text{item} = 15.5^{\text{th}} \text{item}$$

This corresponds to item 25. Thus, the median is 25.

Calculation of Median in Continuous Series

In a continuous series, the following steps are involved in the calculation of median.

***Step* 1**: Convert simple frequencies to cumulative frequencies.
***Step* 2**: Calculate the sum of simple frequencies i.e. $\sum f$ as *N*.

***Step* 3**: Determine the value of $\left(\frac{N}{2}\right)^{\text{th}}$ item.

***Step* 4**: Median class corresponds to the cumulative frequency that includes the $\left(\frac{N}{2}\right)^{th}$ item

***Step* 5**: Compute median by the following formula,

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Here,

l_1 represents lower limit of median class
c.f. represents cumulative frequency of the class preceding the median class
f represents frequency of median class

i represents the size of median class interval

Example: For the following data, calculate the median.

Items (*X*)	**Frequency (*f*)**
0-10	12
10-20	15
20-30	13
30-40	11
40-50	9
	$\sum f = 60$

Solution

Items (*X*)	**Frequency (*f*)**	**Cumulative Frequency (*c.f*)**
0-10	12	12
10-20	15	27
20-30	13	40
30-40	11	51
40-50	9	60
	$\sum f = 60$	

$$M = \frac{N}{2} = (\frac{60}{2}) = 30$$

This corresponds to the class interval (20 - 30). This is the median class.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{30-27}{13} \times 10$$

$$= 22.30$$

Miscellaneous Illustrations

Example 1: Calculate the median in the data given below by arranging the series in descending order.

Item	Frequency
0-10	5
10-20	10
20-30	13
30-40	17
40-50	15
50-60	13
60-70	7

Solution

The descending order of the series is as follows.

Item (X)	Frequency (f)	Cumulative Frequency (c.f.)
60-70	7	7
50-60	13	20
40-50	15	35
30-40	17	52
20-30	13	65
10-20	10	75
0-10	5	80
	$\sum f = 80$	

$M = \text{size of} \left(\frac{N}{2}\right)^{th} \text{item} = \left(\frac{80}{2}\right) = 40$

This corresponds to the class interval (30-40). This is the median class.

Now, as the series is arranged in descending order, the formual will be different from the one used for the ascending order series. The new formula is as follows.

$$M = l_2 - \frac{\frac{N}{2} - c.f.}{f} \times i$$

where,

l_2 represents the upper limit of the median class.

f represents the frequency of the median class.

c.f. represents the cumulative frequency of the class preceeding the median class.

So,

$$M = 40 - \frac{40-35}{17} \times 10$$
$$= 37.06$$

Hence, the median is 37.06

Note: When the data is arranged in descending order, we take the upper limit of the median class rather than the lower limit.

Example 2: The frequency distribution of the wages received by workers are given. Calculate the median.

Wages (Rs in hundred)	**Number of Workers**
Less than 10	7
Less than 20	23
Less than 30	41
Less than 40	60
Less than 50	71
Less than 60	80
Less than 70	86
Less than 80	90

Solution

Here, the wages are given in the less than form. For calculating the median, we first convert them into class intervals as follows.

Wages (Rs in hundred)	**Frequency (*f*)**	**Frequency (c.*f.*)**

0-10	7	7
10-20	16	23
20-30	18	41
30-40	19	60
40-50	11	71
50-60	9	80
60-70	6	86
70-80	4	90
	$\sum f = 90$	

$$M = \text{size of} \left(\frac{N}{2}\right)^{th} \text{item} = \left(\frac{90}{2}\right) = 45$$

This corresponds to the class interval (30 - 40). This is the median class.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{45 - 41}{19} \times 10$$

$$= 32.10$$

Example 3: Calculate median for the following frequency distribution.

Marks **(*X*)**	**Number of students** **(*f*)**
More than 0	40
More than 10	35
More than 20	28
More than 30	20
More than 40	14
More than 50	6

Solution

Just like in less than series, here the more than series is first converted into class interval.

Marks (X)	(Number of students) Frequency (f)	Cumulative Frequency ($c.f.$)
0-10 10-20	5 7	5 12
20-30	8	20
30-40 40-50 50-60	6 8 6	26 34 40
	N = 40	

$$M = \text{size of} \left(\frac{N}{2}\right)^{th} \text{item} = \left(\frac{40}{2}\right) = 20$$

This corresponds to the class interval (20 - 30). This is the median class.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{20-12}{8} \times 10$$

$$= 30$$

Example 4: The following table presents the marks of 60 students in a class. Calculate the median.

X	Frequency (f)
1-5 6-10 11-15 16-20 21-25 26-30	6 9 12 13 11 9
	$\sum f$ = 60

Solution

Note that this is an inclusive class interval series. In order to calculate median, we will first convert the inclusive class interval series into the exclusive class interval series as follows.

X	**Frequency** ***(f)***	**Cumulative Frequency** ***(c.f)***
0.5-5.5	6	6
5.5-10.5	9	15
10.5-15.5	12	27
15.5-20.5	13	40
20.5-25.5	11	51
25.5-30.5	9	60

$$M = \text{size of} \left(\frac{N}{2}\right)^{th} \text{item} = \left(\frac{60}{2}\right) = 30$$

This corresponds to the class interval (15.5-20.5). This is the median class.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 15.5 + \frac{30-27}{13} \times 5$$

$$= 16.65$$

Example 5: For the given data, calculate the value of median.

Marks **(*X*)**	**Frequency** **(*f*)**
0-5	7
5-20	11
20-40	14
40-70	12
70-90	6

Solution

Note that here the class interval are unequal. However, this does not pose any problem in the estimation of the median. The method for calculation of median remains the same as in the case of equal class interval.

Marks **(*X*)**	**Frequency** **(*f*)**	**Cumulative frequency** **(*c.f*)**

0-5	7	7
5-20	11	18
20-40	14	32
40-70	12	44
70-90	6	50

$$M = \text{size of } \left(\frac{N}{2}\right)^{th} \text{item} = \left(\frac{50}{2}\right) = 25$$

This corresponds to the class interval (20-40). This is the median class.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{25-18}{14} \times 20$$

$$= 30$$

Calculation of Missing Frequency

Example: For the following data, find the value of the missing frequency, if the median is 16.

Class Interval	0-5	5-10	10-15	15-20	20-25
Frequency (*f*)	10	7	5	? = f_1	13

Solution

Class Interval	**Frequency (*f*)**	**Cumulative Frequency (*c.f.*)**
0-5	10	10
5-10	7	17
10-15	*5*	22
15-20	f_1	22+ f_1
20-25	13	35 + f_1

Now, as median is 16, so the median class is (15-20).
Now,

$$\text{Median} = l + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$\text{or, } 16 = 15 + \frac{\left(\frac{35 + f_1}{2}\right) - 22}{f_1} \times 5$$

$$\text{or, } 1 = \frac{\left(\frac{35 + f_1}{2}\right) - 22}{f_1} \times 5$$

$$\text{or, } f_1 = 15$$

Hence, the value of the missing frequency is 15.

Example 6: For the following data, find the missing frequencies, if the median is 32 and total of frequencies is 100.

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency (*f*)	10	15	? = f_1	25	? = f_2	25

Solution

Class Interval	**Frequency (*f*)**	**Cumulative Frequency (*c.f.*)**
0-10	10	10
10-20	15	25
20-30	f_1	$25 + f_1$
30-40	25	$25 + f_1 + 25 = 50 + f_1$
40-50	f_2	$50 + f_1 + f_2$
50-60	25	$75 + f_1 + f_2$
	$N = 100$	

We know,
$N = 100$

Taking the sum of remaining frequencies we get.
$75 + f_1 + f_2 = 100$
or, $f_1 + f_2 = 25$ **…..(1)**

Now, as median is 32, so the median class is (30 – 40).

Now,

$$\text{Median} = l + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$\text{or, } 32 = 30 + \frac{50 - (25 + f_1)}{25} \times 10$$

$$\text{or, } f_1 = 20$$

Substituting the value of f_1 in equation (1)

$f_1 + f_2 = 25$
$20 + f_2 = 25$
$f_2 = 5$

Thus, the missing frequencies are 20 and 5 respectively.

Graphical Location of Median

Graphically, the median is located with the help of ogives. The following steps are involved in the graphical location of median.

***Step* 1**: For the given data calculate the less than and more than cumulative frequencies.

***Step* 2**: With the help of less than frequencies, draw the 'less than ogive'.

***Step* 3**: With the help of more than frequencies, draw the 'more than ogive'.

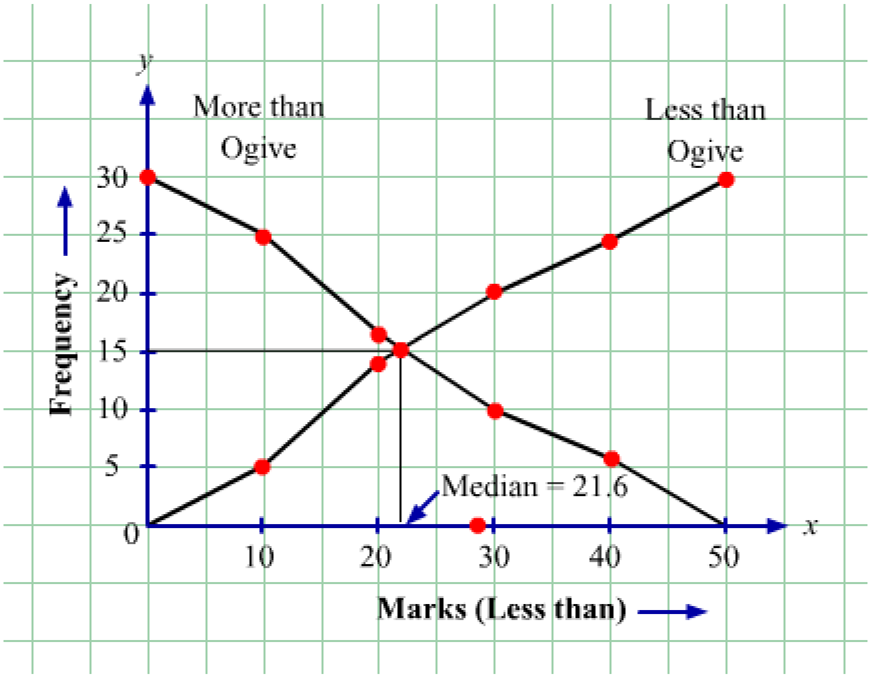
***Step* 4**: Now, from the intersecting point of the two ogives, draw a perpendicular on *x*-axis.

***Step* 5**: The point on the *x*-axis corresponding to the perpendicular is the median value.

Example: For the following data, determine the median with the help of ogive curves.

Marks (*X*)	**Number of students (*f*)**	**Cumulative Frequency *c.f* (less than)**	**Cumulative Frequency *c.f* (more than)**
0-10	5	5	30
10-20	9	14	25
20-30	6	20	16
30-40	4	24	10

40-50	6	30	6



Besides determining the median graphically using the two ogives, we can also ascertain median using only less than ogive. This procedure is explained in the below video.

Merits of Median

The following are some of the merits of median.

- Median is very easy to calculate.
- It is possible to determine the median graphically as well.
- Unlike mean, it can also be calculated for open-ended classes as well as classes with unequal intervals.
- The extreme values in the series do not affect the median.

Demerits of Median

The following are some of the demerits of median.

- Calculation of median requires the arrangement of data in either the ascending or the descending order.
- Further algebraic treatment of median is not possible.

- Median is not based on all the items in the series.
- Fluctuations in the sample can affect the median.

Mode- Meaning, Properties and Computation Methods

Objectives

After going through this lesson, you shall be able to understand the following concepts.

- Meaning and Computation of Mode
- Graphical Presentation of Mode
- Merits and Demerits of Mode

Meaning of Mode

In the previous lesson, we studied about the median. In this lesson, we will study about another positional average, mode. Mode refers to the value in the statistical series that has the highest frequency. In other words, the value that occurs (or repeats itself) most frequently in a series is known as mode. Thus, the value that occurs most frequently is the central value and is said to be a representative of the entire data set. It is denoted by *Z*. For example, suppose a survey is done to analyse the pattern of demand for rice. It is found that maximum number of people demand 2 kg of rice. In this case, it can be said that mode is 2 kg rice.

Calculation of Mode in Individual Series

In individual series, mode can be calculated using either of the following two methods.

- Inspection Method
- Changing the Individual Series into Discrete Series

(i) *Inspection method*: This is the simplest method of locating the mode. In this method, the given data is examined and the value that is found to occur the most is the mode or the modal value.

Example: The following marks are scored by 10 students in a class. Find the modal marks.

Student (Roll No.)	Marks Scored
1	15
2	17

3	14
4	15
5	15
6	15
7	18
8	14
9	12
10	13

Solution

A close examination of the data reveals that 15 occurs the highest number of times in the series i.e. 4 times. Thus, 15 is the modal marks.

(ii) ***Changing the individual series into discrete series***: When there are a large number of observations in a series then, the location of mode by the inspection method can prove to be difficult and cumbersome. In such cases the given individual series is first converted into discrete series or frequency series. The value that is found to have the highest frequency is the modal value.

<u>Example</u>: Considering the example given above, let us find mode for the series, by converting it into discrete series.

Solution: The given series is converted into discrete series as follows.

Conversion of Individual Series into Discrete Series

<table>
<tr><th>Marks Scored</th><th>Tally Marks</th><th>Frequency</th><th></th></tr>
<tr><td>12</td><td>|</td><td>1</td><td></td></tr>
<tr><td>13</td><td>|</td><td>1</td><td></td></tr>
<tr><td>14</td><td>||</td><td>2</td><td></td></tr>
<tr><td>15</td><td>||||</td><td>4</td><td>→ Modal Value</td></tr>
<tr><td>17</td><td>|</td><td>1</td><td></td></tr>
<tr><td>18</td><td>|</td><td>1</td><td></td></tr>
</table>

Calculation of Mode in Discrete Series

The following are the two methods of locating mode in a discrete series.

- Inspection Method
- Grouping Method

(i) ***Inspection method***: Similar to the inspection method in individual series, in discrete series the value having the highest frequency is the mode.

Example: The following table presents the salary of 40 employees in a factory. Compute the modal salary.

Salary (in Rs)	**Number of Employees**
12,000	5
10,000	20
15,000	8
18,000	3
16,000	4
Total	40

Solution: In this series, 10,000 has the highest frequency (20). Thus, 10,000 is the modal salary.

(ii) ***Grouping method***: In a discrete series, at times more than one value may have the highest frequency. Another possibility is that the difference between the maximum frequency and its preceding as well as succeeding frequency is very small. In such cases, it becomes difficult to find mode by mere inspection. Here, the mode is located using the grouping method. Under this method, a grouping table and an analysis table is prepared.

Steps to Prepare Grouping Table

A grouping table comprises of 6 columns. These are:

***Column* 1**: Here, the value having the highest frequency is marked.

***Column* 2**: In this column, the given frequencies are grouped in 2's.

***Column* 3**: Here, leaving the first frequency, the remaining frequencies are grouped in 2's.

***Column* 4**: Now, the frequencies are grouped in 3's.

***Column* 5**: In this column, leaving the first frequency, the remaining frequencies are grouped in 3's.

***Column* 6**: In this column, leaving the first two frequencies, the remaining frequencies are grouped in 3's.

After preparing the grouping table, the analysis table is prepared by finding out the highest frequency in each column.

Example: For the following data, compute the value of mode.

Price	Frequency
5	4
7	9
10	12
6	18
8	25
4	12
11	18
16	15
12	7
3	2

Solution

The grouping table for the given series is as follows.

Grouping Table

Price	Frequency (I)	II	III	IV	V	VI
5	4	13		4 + 9 + 12 = 25		
7	9		21		9 + 12 + 18 = 39	
10	12	30				12 + 18 + 25 = 55
6	18		43	18 + 25 + 12 = 55		
8	25	37			25 + 12 + 18 = 55	
4	12		30			12 + 18 + 15 = 45
11	18	33		18 + 15 + 7 = 40		
16	15		22		15 + 7 + 2 = 24	
12	7	9				
3	2					

On the basis of this grouping table, an analysis table is prepared. For each column of the grouping table, we analyse which item/group of items correspond to the highest frequency.

In other words, in the analysis table, we record those items that correspond to the highest frequency in different columns. On the basis of the above grouping table, the analysis table is prepared in the following manner:

Analysis Table

Columns	Price									
	5	**7**	**10**	**6**	**8**	**4**	**11**	**16**	**12**	**3**
I					✓					
II					✓	✓				
III				✓	✓					
IV				✓	✓	✓				
V					✓	✓	✓			
VI			✓	✓	✓					
Total			1	3	6	3	1			

From the analysis table, it is clear that the value 8 is repeated the maximum number of times. Hence, 8 is the modal value.

Calculation of Mode in Continuous Series

Similar to the calculation under discrete series, under continuous series as well, the mode can be located by following either the inspection method or the grouping method.

(i) ***Inspection method***: In this method, the class having the maximum frequency, known as- the modal class, is to be identified and the modal value has to be calculated using the following formula:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here,
Z = value of Mode
l_1 = lower limit of modal class
f_0 = Frequency of the preceding modal class
f_1 = Frequency of the modal class
f_2 = Frequency of the subsequent modal class or post modal class

i = Class interval of the modal class

Example: For the following data, compute the value of mode.

Class Interval	Frequency	
0-10	25	
10-20	18	
20-30	35	Modal Class
30-40	12	
40-50	19	

Solution

By inspection, we can say that the modal class is 20 – 30 as it has the highest frequency of 35.

$$\begin{aligned} z &= l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i \\ &= 20 + \frac{35-18}{2\times35-18-12} \times 10 = 24.25 \end{aligned}$$

(ii) ***Grouping method***: In this method, the modal class is computed by preparing a grouping table like the one prepared in discrete series. After finding the modal class, the mode is calculated by using the formula given above.

Example: Calculate mode for the following data.

Class Interval (*C.I.*)	**Frequency (*f*)**
0-10	2
10-20	4
20-30	5
30-40	10
40-50	8
50-60	4

Solution

<table>
<tr><th colspan="7">Grouping Table</th></tr>
<tr><th rowspan="2">(C.I.)</th><th>(f)</th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><th>I</th><th>II</th><th>III</th><th>IV</th><th>V</th><th>VI</th></tr>
<tr><td>0-10</td><td>2</td><td rowspan="2">6</td><td></td><td rowspan="3">11</td><td></td><td></td></tr>
<tr><td>10-20</td><td>4</td><td rowspan="2">9</td><td rowspan="3">$\boxed{19}$</td><td></td></tr>
<tr><td>20-30</td><td>5</td><td rowspan="2">$\boxed{15}$</td><td rowspan="3">$\boxed{23}$</td></tr>
<tr><td>30-40</td><td>$\boxed{10}$</td><td rowspan="2">$\boxed{18}$</td><td rowspan="3">$\boxed{22}$</td></tr>
<tr><td>40-50</td><td>8</td><td rowspan="2">12</td><td></td></tr>
<tr><td>50-60</td><td>4</td><td></td><td></td><td></td></tr>
</table>

Analysis Table						
Column	**C.I.**					
	0 – 10	**10 – 20**	**20 – 30**	**30 – 40**	**40 – 50**	**50 – 60**
I				√		
II			√	√		
III				√	√	
IV				√	√	√
V		√	√	√		
VI				√	√	√
Total		1	2	6	3	1

Thus, (30 – 40) is the modal class.

Now, mode can be determined by applying the following formula.

$$z = l_1 + \frac{+f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 30 + \frac{10-5}{2 \times 10 - 5 - 8} \times 10$$

$$= 37.14$$

Thus, mode is 37.14.

Example: For the following frequency distribution, compute the value of mode.

Mid-Value	**Frequency**
2	3
6	7
10	15
14	22
18	27
22	14
26	9
30	2

Solution

Here, since the mid-value are given, we will have to first convert them into class intervals and then prepare the grouping table.

<table>
<tr><th colspan="7">Grouping Table</th></tr>
<tr><th rowspan="2">(C.I.)</th><th>(f)</th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><th>(I)</th><th>II</th><th>III</th><th>IV</th><th>V</th><th>VI</th></tr>
<tr><td>0 – 4</td><td>3</td><td rowspan="2">10</td><td></td><td rowspan="3">25</td><td></td><td></td></tr>
<tr><td>4 – 8</td><td>7</td><td rowspan="2">22</td><td rowspan="3">44</td><td></td></tr>
<tr><td>8 – 12</td><td>15</td><td rowspan="2">37</td><td rowspan="3">$\boxed{64}$</td></tr>
<tr><td>12 – 16</td><td>22</td><td rowspan="2">$\boxed{49}$</td><td rowspan="3">$\boxed{63}$</td></tr>
<tr><td>16 – 20</td><td>$\boxed{27}$</td><td rowspan="2">$\boxed{41}$</td><td rowspan="3">$\boxed{50}$</td></tr>
<tr><td>20 – 24</td><td>14</td><td rowspan="2">23</td><td rowspan="3">25</td></tr>
<tr><td>24 – 28</td><td>9</td><td rowspan="2">11</td><td></td></tr>
<tr><td>28 – 32</td><td>2</td><td></td><td></td><td></td></tr>
</table>

Analysis Table							
Column	**C.I.**						
	0 – 4	**4 – 8**	**8 – 12**	**12 – 16**	**16 – 20**	**20 – 24**	**24 – 28**
I					√		
II					√	√	
III				√	√		
IV				√	√	√	
V					√	√	√
VI			√	√	√		
Total			1	3	6	3	1

Hence, the modal class is 16-20.

Now,

$$z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$
$$= 16 + \frac{27-22}{2\times 27 - 22 - 14} \times 4$$
$$= 17.11$$

Hence, the mode is 17.11.

Example: Find the value of mode for the following data.

Class	Frequency
Below 10	3
Below 20	7
Below 30	15
Below 40	20
Below 50	22
Below 60	28
Below 70	32

Solution

For the computation of mode, we first convert the given data into class interval.

Grouping Table						
(C.I.)	**(*f*)**					
	(I)	**II**	**III**	**IV**	**V**	**VI**

<table>
<tr><td>0 – 10</td><td>3</td><td rowspan="2">7</td><td></td><td rowspan="3">$\boxed{15}$</td><td></td><td></td></tr>
<tr><td>10 – 20</td><td>4</td><td rowspan="2">$\boxed{12}$</td><td></td><td></td></tr>
<tr><td>20 – 30</td><td>$\boxed{8}$</td><td rowspan="2">$\boxed{13}$</td><td>$\boxed{17}$</td><td>$\boxed{15}$</td></tr>
<tr><td>30 – 40</td><td>5</td><td rowspan="2">7</td><td rowspan="3">13</td><td></td><td></td></tr>
<tr><td>40 – 50</td><td>2</td><td rowspan="2">8</td><td rowspan="3">12</td><td></td></tr>
<tr><td>50 – 60</td><td>6</td><td rowspan="2">10</td><td></td></tr>
<tr><td>60 – 70</td><td>4</td><td></td><td></td><td></td></tr>
</table>

Analysis Table

Column	**C.I.**						
	0 – 10	**10 – 20**	**20 – 30**	**30 – 40**	**40 – 50**	**50 – 60**	**60 – 70**
I			√				
II			√	√			
III		√	√				
IV	√	√	√				
V		√	√	√			
VI			√	√	√		
Total	1	3	6	3	1		

Clearly, (20-30) is the modal class.

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$
$$= 20 + \frac{8-4}{2\times 8 - 4 - 5} \times 10$$
$$= 25.7$$

Hence, mode is 25.7.

Graphical Presentation of Mode

Besides, algebraic computation, mode can also be determined graphically. The following steps are involved in the graphical determination of mode.

Step 1: Prepare a histogram for the given data.

Step 2: In the histogram, the highest rectangle represents the modal class.

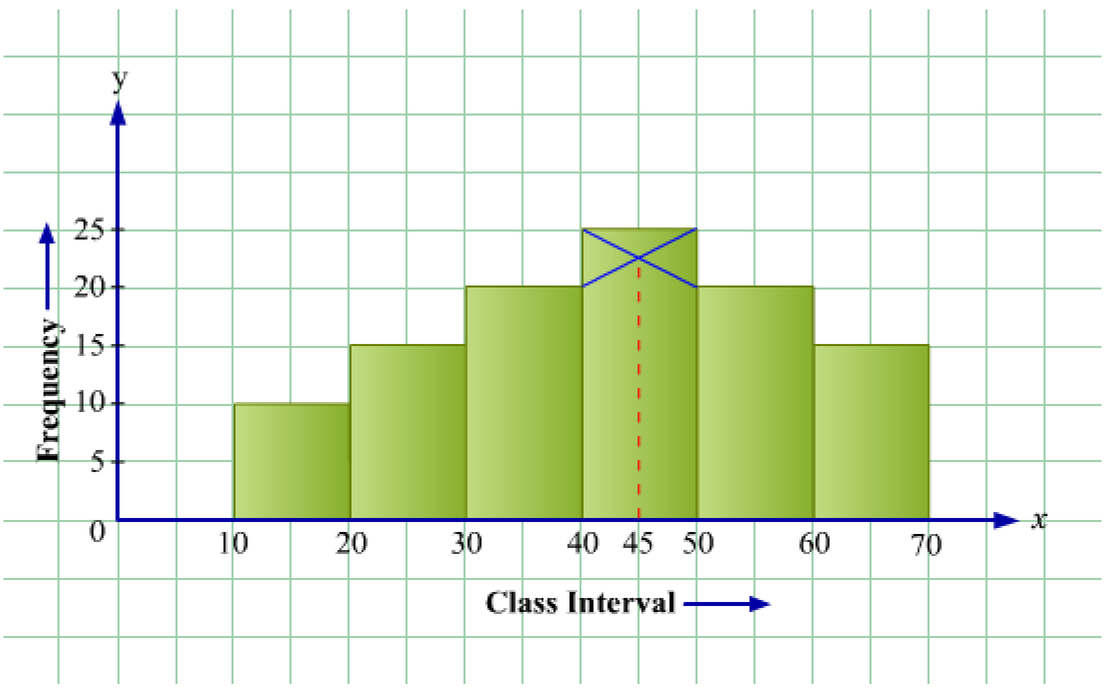
Step 3: In the modal class rectangle, draw two lines diagonally to the upper corner to the rectangles.

Step 4: Now, from the point of intersection of the lines drawn in ***step 3***, draw a perpendicular on the *x*-axis. The point where the perpendicular touches the *x*-axis gives the modal value.

Example: For the given data, determine mode graphically.

Class Interval	Frequency
0-10	5
10-20	10
20-30	15
30-40	20
40-50	25
50-60	20
60-70	15

Solution



Thus, mode is 45.

Note: Mode can also be determined with the help of frequency polygon. In this case, a perpendicular is drawn on the *x*-axis, from the highest point of the frequency polygon to locate the modal value.

Merits of Mode

- It is unaffected by the presence of extreme values.
- It indicates the value of the series in the best way.
- It is comparatively easier to calculate mode, as the information regarding all values and all frequencies is not required. Only the most frequent (maximum) value is to be known.
- In addition to the mathematical computation of mode, it is also possible to determine the mode graphically.
- Mode can also be determined in the case of open ended classes.

Demerits of Mode

- Mode is not based on all the values of a series.
- It is incapable of any further algebraic treatment.
- Sometimes, it is difficult to ascertain the definite value of mode.
- It fails to represent the small values of a series. Therefore, it may not be the best indicator of the series.

Quartiles, Deciles, Percentile, Empirical Relationship between A.M, Median and Mode

Objective

After going through this lesson, you shall be able to understand the following concepts:

- Quartiles- Meaning and Computation
- Deciles- Meaning and Computation
- Percentiles- Meaning and Computation
- Empirical Relationship Between A.M., Median and Mode

Introduction

In the previous lessons, we studied the meaning and computation of two positional averages, median and mode. In this lesson, we will study about some more positional averages namely, quartiles, deciles and percentiles.

Quartiles

Quartile refers to the value that divides the series into four equal parts. A study of quartile gives a more clear picture of the composition of the series. For a series there are three quartiles, Q_1, Q_2, and Q_3.

- The first quartile (Q_1), also known as the lower quartile has 25% of the items in the series below it and 75% of the items above it.

- The second quartile (Q_2) has 50% of the items both above and below it. The second quartile is the same as median.

- The third quartile (Q_3), also known as the upper quartile has 75% of the items in the series below it and 25% of the items above it.

Let us understand the computation of the quartiles in the three statistical series, namely, individual series, discrete series and continuous series.

Calculation of Quartiles in Individual Series

For the computation of quartiles, the given series is first arranged in either ascending or descending order. The quartiles can then be located using the following formula.

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ item}$$

$$Q_3 = \text{Size of } \left(\frac{3(N+1)}{4}\right)^{\text{th}} \text{ item}$$

<u>Example</u>: For the given data, compute the first and the third quartile.

Marks (*X*)	12	14	15	9	6	11	4	8	16	15	13

Solution

Marks (*X*)	**Marks arranged in ascending order**
12	4
14	6

15	8
9	9
6	11
11	12
4	13
8	14
16	15
15	15
13	16
	$N = 11$

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{item} = \left(\frac{11+1}{4}\right)^{th} \text{item} = 3^{rd} \text{item}$$

$$= 8$$

$$Q_3 = \text{Size of } \left(\frac{3(N+1)}{4}\right)^{th} \text{item} = \left(\frac{3(11+1)}{4}\right)^{th} \text{item} = 9^{th} \text{item}$$

$$= 15$$

Hence, the first and third quartile are 8 and 15 respectively.

Calculation of Quartiles in Discrete Series

In discrete series, the formula for estimation of quartile remains the same as in individual series.

The following are the steps involved in the calculation of quartiles in the discrete series.

***Step* 1**: Arrange the data in either ascending or descending order.

***Step* 2**: Calculate the cumulative frequencies for the given data.

***Step* 3**: For the first quartile locate the size of $\left(\frac{N+1}{4}\right)^{th}$ item in the cumulative frequency column and corresponding *x* value is Q_1.Similary, for the third quartile locate the size of Size of $\left(\frac{3(N+1)}{4}\right)^{th}$ item in the cumulative frequency. The corresponding *x* value is Q_3.

Example: The following table presents the wages of 80 workers in a factory. Calculate

a) Maximum income of the lowest 25% workers

b) Minimum income of the highest 25% workers

Wages (Rs in '000)	Number of Workers
5	4
6	8
7	9
8	14
9	16
10	18
11	8
12	3

Solution

Wages (in Rs '000)	Number of Workers (*f*)	Cumulative Frequency (*c.f.*)
5	4	4
6	8	12
7	9	21
8	14	35
9	16	51
10	18	69
11	8	77
12	3	80

(a) Maximum income of the lowest 25% workers implies the computation of the first quartile.

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$$
$$= \text{Size of } \left(\frac{80+1}{4}\right)^{th} \text{ item} = 20.25^{th} \text{item}$$

This corresponds to 21in the cumulative frequency.

Hence, the maximum income of the lowest 25% workers is Rs 7,000.

(b) Minimum income of the highest 25% workers implies the computation of the third quartile.

$$Q_3 = \text{Size of } \frac{3(N+1)}{4}^{\text{th}} \text{ item} = 3\left(\frac{80+1}{4}\right)^{\text{th}} \text{ item}$$
$$= 60.75^{\text{th}} \text{ item}$$

This corresponds to 69 in the cumulative frequency. Hence, the minimum income of the highest 25% workers is Rs 10,000.

Calculation of Quartiles in Continuous Series

The following are the steps involved in the calculation of quartiles in the continuous series.

***Step* 1**: Arrange the data in either ascending or descending order

***Step* 2**: Calculate the cumulative frequencies for the given data.

***Step* 3**: For the first quartile locate the size of $\left(\frac{N}{4}\right)^{\text{th}}$ item in the cumulative frequency column. Similarly, for the third quartile locate the size of Size of $\left(\frac{3N}{4}\right)^{\text{th}}$ item in the cumulative frequency.

***Step* 4**: Calculate the first quartile is calculated using the following formula

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

Here,

l_1 represents lower limit of the quartile class.
N represents sum of frequencies
c.f. represents cumulative frequency of the class preceding the Q_1 class

i represents class interval

On the other hand, the third quartile is calculated using the following formula.

$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f.}{f} \times i$$

Here,

l_1 represents lower limit of the quartile class.

N represents sum of frequencies

c.f. represents cumulative frequency of the class preceding the Q_1 class

i represents class interval

Example: The following frequency distribution presents the income of 70 families. Calculate values of the first quartile and third quartile for the data.

Income (Rs in '000)	**Number of Persons**
10-20	12
20-30	16
30-40	18
40-50	14
50-60	10

Solution

Income (X) (Rs in '000)	**Number of Persons (f)**	***c.f***	
10 – 20	12	12	
20 – 30	16	28	$\rightarrow Q_1$
30 – 40	18	46	
40 – 50	14	60	$\rightarrow Q_3$
50 – 60	10	70	
	$\sum f = 70$		

Solution

$$Q_1 = \text{Size of } \left(\frac{N}{4}\right)^{th} \text{ item} = \left(\frac{70}{4}\right)^{th} \text{ item} = 17.5^{th} \text{item}$$

This lies in the class interval (20 − 30)

Now,

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$= 20 + \frac{17.5 - 12}{16} \times 10 = 23.43$$

Thus, the value of first quartile is 23.43

$$Q_3 = \text{Size of } 3\left(\frac{N}{4}\right)^{th} \text{ item} = 3\left(\frac{70}{4}\right) = 52.5^{th} \text{item}$$

This lies in the class interval (40 − 50)

$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f.}{f} \times i$$

$$= 40 + \frac{52.5 - 46}{14} \times 10 = 44.64$$

Deciles

Decile refer to the values that divides the series into ten equal parts.

Calculation of Deciles in Individual Series

In individual series the data is first arranged in either ascending or descending order. The decile is then calculated using the following formula.

$$D = \text{Size of } x\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

Here, *x* is the decile to be calculated. For example, if 4^{th} decile is to be calculated then, the formula becomes

$$D = \text{Size of } 4\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

Similarly, if 8^{th} decile is to be calculated then, the formula becomes

$$D = \text{Size of } 8\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

Example: For the following data compute the value of fifth decile.

(***X***)	4	5	6	7	8	9	10

Solution

$$D_5 = \text{Size of } 5\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

$$\text{or, } D_5 = \text{Size of } 5\left(\frac{7+1}{10}\right)^{th} \text{ item}$$

$$\text{or, } D_5 = \text{Size of } 4^{th} \text{item} = 7$$

Thus, value of 5^{th} decile is 7.

Calculation of Deciles in Discrete Series

In discrete series also, the formula for calculating decile remains the same as in individual series.

Example: For the following data calculate the value of seventh decile.

Marks	**Number of Students**
10	7
11	6
12	8
13	10
14	12
15	9

Solution

Marks (*X*)	**Number of Students (*f*)**	**Cumulative Frequency (*c.f.*)**
10	7	7
11	6	13
12	8	21
13	10	31
14	12	43
15	9	52

Thus, in order to calculate, 7^{th} decile

$$D_7 = \text{Size of } 7\left(\frac{N+1}{10}\right)^{th} \text{ item}$$

$$= \text{Size of } 7\left(\frac{52+1}{10}\right)^{th} \text{ item} = 37.1^{th} \text{ item}$$

This corresponds to 43^{rd} cumulative frequency. Thus, the value of seventh decile (D_7) is 14

Calculation of Deciles in Continuous Series

The following are the steps involved in the calculation of quartiles in continuous series.

***Step* 1**: Arrange the data in ascending order or discrete series.

***Step* 2**: Calculate the cumulative frequencies for the given data.

***Step* 3**: Locate the decile by $\text{Size of } x\left(\frac{N}{10}\right)^{th} \text{ item}$ in the cumulative frequency column, where x is the decile to be calculated.

***Step* 4**: Calculate the decile using the following formula.

$$D = l_1 + \frac{x\left(\frac{N}{10}\right) - c.f.}{f} \times i$$

Here,

l_1 represents the lower limit of the decile class

c.f. represents cumulative frequency of the class preceding decile class

f represents frequency of decile class

i represents the class interval

<u>Example</u>: Calculate the value of sixth decile for the following data.

Age Group	Number of Persons

15-20	17
20-25	13
25-30	18
30-35	12
35-40	13
40-45	7
45-50	10

Solution

Age Group	Number of of Persons (*f*)	*c.f.*
15-20	17	17
20-25	13	30
25-30	18	48
30-35	12	60
35-40	13	73
40-45	7	80
45-50	10	90

Solution

$$6^{th}\text{decile} = \text{Size of } 6\left(\frac{90}{10}\right)^{th} \text{ item} = 54^{th} \text{ item}.$$

This lies in the class interval $(30-35)$.

$$D_6 = l_1 + \frac{6\left(\frac{N}{10}\right) - c.f.}{f} \times i$$

$$= 30 + \frac{(54-48)}{12} \times 5 = 32.5$$

Thus, the value of sixth decile is 32.5

Percentiles

Percentile refers to the value that divides the series into 100 equal parts.

Calculation of Percentiles in Individual Series

In Individual series, the value of percentile is estimated by using the following formula

$$P = \text{Size of } x\left(\frac{N+1}{100}\right)^{th} \text{ item}$$

where, *x* is the percentile to be calculated.

Example: For the following data calculate the value of 60^{th} percentile.

10	15	8	16	14	10	9	12	18

Solution

$$60^{th}\text{percentile} = 60\left(\frac{9+1}{100}\right) = 6^{th}\text{item}$$

Thus, the value of sixth percentile is 10.

Note: If the size of the $(\_)^{th}$ item comes to be in fraction, say 7.8^{th}, then, in order to calculate the percentile value, we need to use the following formula.

$$P_n = \text{Size of } n^{th} \text{ item} + \text{fraction } \{\text{Size of } (n+1)^{th} \text{item} - \text{ Size of } n^{th} \text{item}\}$$

For example,

$$P_{10} = \text{Size of } 10\left(\frac{28+1}{100}\right)^{th} \text{ item}$$

$$\text{or, } P_{10} = \text{Size of } 10\left(\frac{29}{100}\right)^{th} \text{ item} = \text{Size of } (2.9)^{th} \text{ item}$$

$$\text{or, } P_{10} = \text{Size of } (2)^{nd} \text{ item} + 0.9 \ \{\text{Size of } 3^{rd} \text{item} - \text{ Size of } 2^{nd} \text{item}\}$$

Example: For the given data, calculate the value of 20^{th} percentile.

Age (in Years)	**Arranged in ascending order**
15	14
22	15
27	16
16	19
14	22
19	27

Solution

$$P_{20} = \text{Size of } 20\left(\frac{6+1}{100}\right)^{th} \text{ item}$$

$$\text{or, } P_{20} = \text{Size of } 20\left(\frac{7}{100}\right)^{th} \text{ item} = \text{Size of } (1.4)^{th} \text{ item}$$

$$\text{or, } P_{20} = \text{Size of } (1)^{th} \text{ item} + 0.4 \ \{\text{Size of } 2^{nd} \text{item} - \text{ Size of } 1^{st} \text{item}\}$$

$$\text{or, } P_{20} = 14 + 0.4 \ \{15 - 14\} = 14.4$$

$$\Rightarrow P_{20} = 14.4$$

Calculation of Percentile in Discrete Series

Just like in individual series, percentile in discrete series is also calculated by using the same formula

$$P = \text{Size of } x\left(\frac{N+1}{100}\right)^{\text{th}} \text{ item}$$

Example: The following table relates to the age of the 40 persons. Calculate the size of 37^{th} percentile.

Age (in years)	No. of persons (*f*)	*c.f*
14	7	7
15	8	15
16	5	20
17	6	26
18	4	30
19	3	33
20	7	40

Solution

$$P_{37} = \text{Size of } 37\left(\frac{40+1}{100}\right)^{\text{th}} \text{ item}$$
$$= 15.17$$

This corresponds to the cumulative frequency equal to 20. Thus, the value of 37^{th} percentile is 16

Calculation of Percentiles in Continuous Series

The following are the steps involved in the calculation of percentile in continuous series.

Step 1: Calculate the cumulative frequencies for the given data.

Step 2: Locate the percentile by $\text{Size of } x\left(\frac{N}{100}\right)^{\text{th}} \text{ item}$ in the cumulative frequency column. (where *x* is the percentile to be calculated).

Step 3: Calculate the percentile by using the following formula:

$$P_x = l_1 + \frac{x\left(\frac{N}{100}\right) - c.f.}{f} \times i$$

Here, l_1 represents the lower limit of the percentile class

c.f. represents cumulative frequency of the class preceding percentile class

f represents the frequency of percentile class

i represents the class interval

Example: The following table relates to marks of 40 students in a class. Calculate the value of the 17th percentile.

Marks	**Number of Students**
0-5	2
5-10	4
10-15	10
15-20	14
20-25	7
25-30	3

Solution

Marks	**Number of Students (*f*)**	**Cumulative Frequency (*c.f*)**
0-5	2	2
5-10	4	6
10-15	10	16
15-20	14	30
20-25	7	37
25-30	3	40

Solution

$$P_{17} = \text{Size of } 17\left(\frac{N}{100}\right)^{\text{th}} \text{ item}$$
$$= 17 \times \left(\frac{40}{100}\right)^{\text{th}} \text{ item} = 6.80^{\text{th}} \text{ item}$$

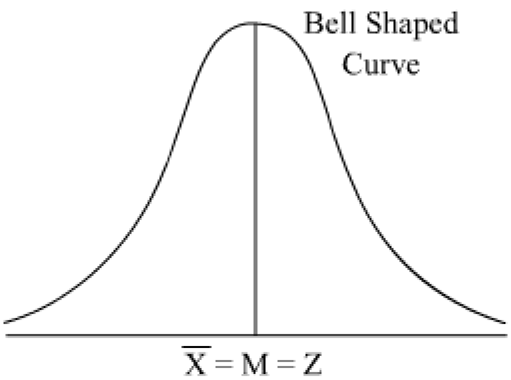
This lies in the class-interval of (10-15).

$$P_{17} = l_1 + \frac{17\left(\frac{N}{100}\right) - c.f.}{f} \times i$$
$$= 10 + \frac{(6.80 - 6)}{10} \times 5 = 10.4$$

Thus, the value of 17th percentile is 10.4

Empirical Relationship between A.M, Median and Mode

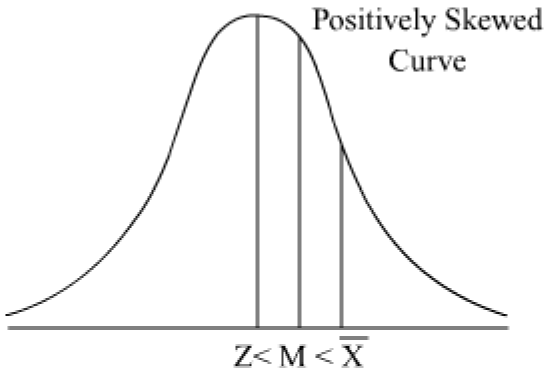
As studied in the previous lessons, there are two types of distribution-symmetrical distribution and asymmetrical distribution. In a symmetrical distribution the frequency curve is bell shaped. Thus, in such distributions mean, median and mode are identical and equal.



However, in assymetrical distribution, the value of mean, median and mode is different.

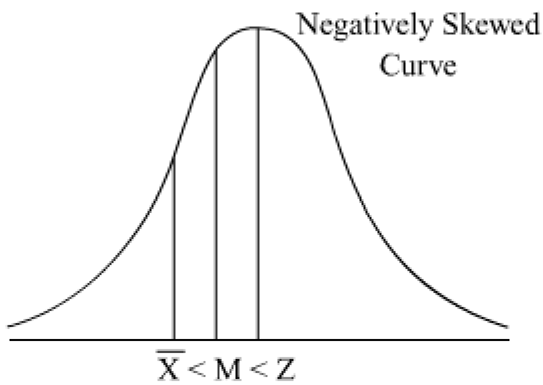
In a ***positively skewed distribution***, the value of the mean $(\overline{X})$ and median (*M*) is greater than the value of mode (*Z*). Algebraically, this is expressed by the following inequality.

$$\left(\overline{X} > M > Z\right)$$



In contrast to this, in a ***negatively skewed distribution***, the value of mean $\left(\overline{X}\right)$ and median (*M*) is less than the value of mode (*Z*). Algebraically, this is expressed by the following inequality.

$\left(\overline{X} < M < Z\right)$



In assymetrical distributions there exists an empirical relationship between mean, median and mode. The relationship as given by Karl Pearson is as follows.

Mode = 3Median – 2Mean