Statistics may rightly be called as a science of averages.

– Sir A. L. Bowley

3

Measures of Central Tendency

Contents :

3.1 Meaning

The large volume of statistical data can be organized by classification and tabulation. This shows certain characteristics of the given data. The diagrams and graphs drawn for the data show its trends and patterns. It helps in visual interpretation and comparison of the data. We need more concise and numerical representation of the data for further statistical analysis. Let us understand this with an illustration.

Suppose a person is planning his monthly budget. The expense on every commodity is a variable which changes according to the quantity consumed and the market price. Suppose he wants to decide the amount to be spared for milk. He has the values of milk expenses of past 10 months. He wants a representative value from these data to provide for the expenses on milk in his budget.

A representative value of more than one set of data can be used for their comparison and further for taking future decisions. We will illustrate this by the following situation.

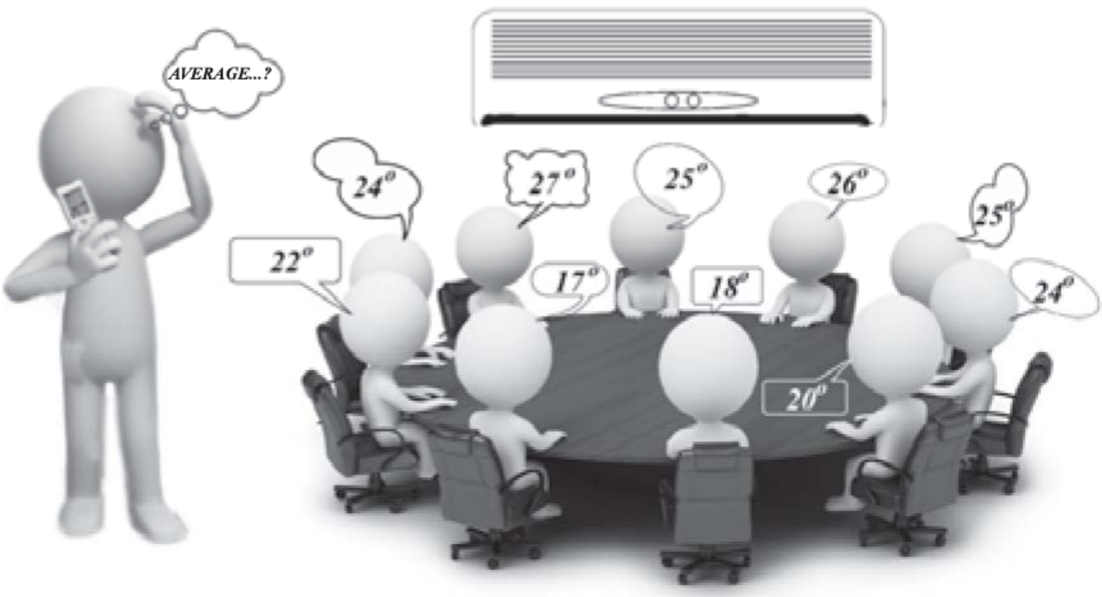
Suppose a company wants to compare the sales of two products produced by it. The sales vary from day to day. The company has data about sales of last 50 days. The pattern displayed by the sales of two products can be compared from the frequency distributions obtained from these data. But the company needs some exact measures to describe the data of sales of their two products for further conclusion and comparison.

In most of the graphs drawn for various frequency distributions, one can observe a common pattern that the values of the variable are concentrated around a certain central value. This characteristic of data is called as **central tendency** and the central value around which the values of the variable are concentrated is called as **measure of central tendency** or **average** value. The average value thus can be taken as a representative for the whole set of data. It is used for further analysis, interpretation and comparison.

Thus an average value

- Presents the data in a concise form
- Shows special characteristics of the data
- Helps in comparison between two or more sets of data

Different measures of central tendency can be obtained for the collected data. The choice of average depends upon the type of data, the purpose of average value and its further application.



3.2 Characteristics of Good Measure of Central Tendency

An average with following characteristics can be called as an ideal average :

(1) It should be well defined and rigid.

(2) It should be easy to understand and calculate.

(3) It should be based on all the observations of the data.

(4) It should be suitable for further algebraic operations.

(5) It should be a stable measure. It means that values of averages found for different samples of same size from the same population should be almost same.

(6) It should not be unduly affected by a few very large or very small observations.

We will discuss the following measures of central tendency which are widely used in data analysis.

(1) Mean (2) Median and other positional averages (3) Mode.

3.3 Arithmetic Mean or Mean

This is one of the most commonly used averages.

3.3.1 Meaning

Arithmetic mean is defined as the value obtained by sum of all observations divided by the total number of observations.

The arithmetic mean of variable x is denoted by $\bar{x}$.

Calculation of mean :

For raw or ungrouped data :

Suppose $x_1, x_2, ..., x_n$ are the n observations in the data, then Arithmetic mean is $\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$

$$= \frac{\Sigma x_i}{n}$$

Where $\Sigma x_i = x_1 + x_2 + ... + x_n$ = Sum of observations $x_1, x_2, ..., x_n$

and n = number of observations

Note : For the sake of simplicity while solving examples, we will not write the suffix i. Thus we will take x instead of x_i , d instead of d_i and f instead of f_i .

Illustration 1 : The following data show the number of scooters repaired daily at a garage. Find the mean number of scooters repaired per day.

7, 13, 4, 8, 6, 9, 10, 4

Here $n = 8$

$$\text{Mean } \bar{x} = \frac{\Sigma x}{n} = \frac{x_1 + x_2 + ... + x_8}{8}$$

$$= \frac{7 + 13 + 4 + 8 + 6 + 9 + 10 + 4}{8}$$

$$= \frac{61}{8}$$

$$= 7.625$$

$$\approx 7.63$$

Thus, the mean number of scooters repaired per day at the garage is 7.63 scooters.

Short cut method :

If the values of observations are very large, the calculation can be simplified by using assumed mean A. A is some constant value, preferably around the centre of all the observations. Assumed mean A is subtracted from observations $x_1, x_2, ..., x_n$ and the deviations are denoted by $d_1, d_2, ..., d_n$.

$$d_1 = x_1 - A,\ d_2 = x_2 - A,\ ...,\ d_n = x_n - A$$

The mean $\bar{x}$ is obtained as follows :

$$\bar{x} = A + \frac{\Sigma d_i}{n}$$

Where A = Assumed mean

$\Sigma d_i = d_1 + d_2 + ... + d_n$

and n = number of observations.

Note : The choice of value of A does not change the value of mean.

Illustration 2 : The rainfall (in mm.) at 10 different places of a district was recorded as :

126, 110, 91, 115, 112, 80, 101, 93, 97, 113

Find mean rainfall.

As the observations have large values, we will calculate mean by short cut method. We will select assumed mean A = 100. The observations x and deviations $d = x - \text{A}$ are shown in the following table :

Rainfall (mm) x	126	110	91	115	112	80	101	93	97	113	**Total**
$d = x - A,\ A$=100	26	10	–9	15	12	–20	1	–7	–3	13	**38**

Here, $n = 10$

Mean $\bar{x} = A + \frac{\Sigma d}{n}$

$= 100 + \frac{38}{10}$

$= 100 + 3.8$

$= 103.8$

Thus, the mean rainfall is 103.8 mm.

Illustration 3: The mean weight of a group of 20 persons was found to be 55 kg. Later, it was discovered that one of them reported her weight as 45 kg which was actually 54 kg. Find the correct mean of their weights.

Here, $\bar{x} = 55$ and $n = 20$

Mean $\bar{x} = \frac{\Sigma x}{n} = 55$

$\therefore \quad \frac{\Sigma x}{20} = 55$

$\therefore \quad \Sigma x = 55 \times 20 = 1100$

Thus the sum of observations is 1100 which includes a wrong value 45 instead of the correct value 54.

To find the correct sum of observations, we subtract the wrong observation and add the correct observation.

$\therefore$ correct $\Sigma x = 1100 - 45 + 54$

$= 1109$

$\therefore$ correct mean $= \dfrac{\text{correct}\sum x}{n}$

$= \dfrac{1109}{20}$

$= 55.45$

Thus, the correct mean weight is 55.45 kg.

For grouped data :

For discrete frequency distribution :

Suppose $x_1, x_2, ..., x_k$ are the observations in the given data with frequencies $f_1, f_2, ..., f_k$ respectively.

Here, n = total number of observations

$= f_1 + f_2 + ... + f_k = \Sigma f_i$

The frequency of x_1 is f_1 means observation x_1 is repeated f_1 times. The sum of all x_1 observations will be $f_1 \times x_1$ or $f_1\, x_1$. Similarly, sum of all x_2 observations will be $f_2\, x_2$ and so on.

$$\text{Mean } \bar{x} = \frac{\text{sum of all observations}}{\text{total number of observations}}$$

$$= \frac{f_1x_1 + f_2x_2 + ... + f_kx_k}{n}$$

$$= \frac{\Sigma f_i x_i}{n}$$

Where $\Sigma f_i\, x_i = f_1\, x_1 + f_2\, x_2 + ... + f_k\, x_k$

llustration 4 : The following table shows the number of children per family in a certain area. Find the mean number of children per family.

No. of children	0	1	2	3	4	5
No. of families	4	8	23	8	6	3

Here we have $k = 6$ values for the variable x.

The calculations to find mean are shown in the following table :

No. of children x	**No. of families** f	fx
0	4	0
1	8	8
2	23	46
3	8	24
4	6	24
5	3	15
Total	$n = 52$	**117**

Mean $\bar{x} = \frac{\Sigma fx}{n}$

$= \frac{117}{52}$

$= 2.25$

Thus, mean for number of children per family is 2.25 children.

Short cut method :

As done earlier for the raw data, an assumed mean A can be suitably chosen and the deviations of values $x_1, x_2, ..., x_k$ can be taken from A. Further, if all these deviations have a common factor c, we can further simplify the calculations by dividing all the deviations by c.

Thus we will have the values $d_1 = \frac{x_1 - A}{c}$, $d_2 = \frac{x_2 - A}{c}$, ..., $d_k = \frac{x_k - A}{c}$

Now, the formula for mean is written as follows :

Mean $\bar{x} = A + \frac{\Sigma f_i d_i}{n} \times c$

Where $\Sigma f_i d_i = f_1 d_1 + f_2 d_2 + ... + f_k d_k$

and n = total number of observations

$= f_1 + f_2 + ... + f_k = \Sigma f_i$

Note : The choice of values of A and c does not change the value of mean.

Illustration 5 : The time (in minutes) taken by a bus to travel between two towns on different days is shown in the following table.

Time (min.)	110	113	120	122	126
No. of days	7	17	11	10	5

Find the mean travel time.

Here the values of observations are large. We will take assumed mean $A = 120$.

The deviations are 110 – 120 = – 10, 113 – 120 = –7, 120 – 120 = 0, 122 – 120 = 2, 126 – 120 = 6.

There is no common factor in them other than 1. Hence we will take $c = 1$

Thus we will have $d = \frac{x - A}{c} = \frac{x - 120}{1} = x - 120$

The calculations for mean will be as follows :

Time (minute) x	**No. of days** f	$d = x - A$ $A = 120$	fd
110	7	– 10	– 70
113	17	– 7	– 119
120	11	0	0
122	10	2	20
126	5	6	30
Total	$n = 50$		**– 139**

Mean $\overline{x} = A + \frac{\Sigma fd}{n} \times c$

$= 120 + \frac{(-139)}{50} \times 1$

$= 120 - 2.78$

$= 117.22$

Thus, the mean travel time for the bus is 117.22 minutes

Illustration 6: The price of an item changes from shop to shop. The following data are available. Find the mean price.

Price (₹)	206	212	218	220	224	230
No. of shops	5	8	9	14	3	1

As the observations are large, we will calculate the mean using short cut method in which, we will select A = 220. The deviations of all observations from A will be –14, –8, –2, 0, 4, 10. These deviations have the highest common factor $c = 2$.

Hence we will take $d = \frac{x-A}{c} = \frac{x-220}{2}$

Calculations for mean :

Price (₹) x	**No. of shops** f	$d = \frac{x-A}{c}$ $A = 220,\ c = 2$	fd
206	5	–7	–35
212	8	–4	–32
218	9	–1	–9
220	14	0	0
224	3	2	6
230	1	5	5
Total	$n = 40$		**–65**

Mean $\overline{x} = A + \frac{\Sigma fd}{n} \times c$

$= 220 + \frac{(-65)}{40} \times 2$

$= 220 + \frac{(-130)}{40}$

$= 220 - 3.25$

$= 216.75$

Thus the mean price of this item is ₹ 216.75.

For continuous frequency distribution :

When we transform the data into a continuous frequency distribution, each frequency shows the number of observations in that class. Thus we do not know each observation of that class. The mid-value of that class is taken as a representative for all the values in that class.

For example, let us consider a class 0-5 with frequency 7. Exact values of these 7 observations are not known. Hence the mid-value 2.5 is assumed for all 7 observations of this class which actually can be any number from 0 to 5.

Considering the mid-value of each class as x, the mean can be calculated by the same method as the one used for the discrete frequency distribution described earlier.

Thus the mean will be calculated as follows :

$$\text{Mean } \bar{x} = A + \frac{\Sigma f_i d_i}{n} \times c$$

Where A = Assumed mean

c = common factor among $x_i - A$

$d_i = \frac{x_i - A}{c}$

f_i = frequency of the class with mid-value x_i

$\Sigma f_i d_i = f_1 d_1 + f_2 d_2 + \ldots\ldots + f_k d_k$

n = total number of observations

$= f_1 + f_2 + \ldots\ldots + f_k = \Sigma f_i$

Illustration 7: The following data represent monthly income (in ₹) of workers in a factory. Find their mean income.

Income (₹)	2000 - 3000	3000 - 4000	4000 - 5000	5000 - 6000	6000 - 7000	7000 - 8000	8000 - 9000
No. of workers	2	3	7	15	25	16	12

We first find the mid-value of each class.

$$\text{mid-value} = \frac{\text{upper limit of the class} + \text{lower limit of the class}}{2}$$

These mid-values will be 2500, 3500, 4500, 5500, 6500, 7500, 8500. We will take A = 5500.

The deviations $x - A$ will be –3000, –2000, –1000, 0,1000, 2000, 3000 respectively.

As these deviations have the highest common factor c = 1000, we will take $d = \frac{x - A}{c} = \frac{x - 5500}{1000}$.

Calculations for mean are shown in the following table :

Income (₹)	**No. of workers** f	**Mid-value** x	$d = \frac{x-A}{c}$ $A = 5500,\ c = 1000$	fd
2000 - 3000	2	2500	–3	–6
3000 - 4000	3	3500	–2	–6
4000 - 5000	7	4500	–1	–7
5000 - 6000	15	5500	0	0
6000 - 7000	25	6500	1	25
7000 - 8000	16	7500	2	32
8000 - 9000	12	8500	3	36
Total	$n = 80$			**74**

Mean $\bar{x} = A + \frac{\Sigma fd}{n} \times c$

$= 5500 + \frac{74}{80} \times 1000$

$= 5500 + \frac{74000}{80}$

$= 5500 + 925$

$= 6425$

Thus, the mean income of these workers is ₹ 6425.

Illustration 8 : The data about weight (in grams) of mangoes from a tree are as follows. Moreover, the minimum weight for these mangoes is 410 grams. Find the mean weight of mangoes.

Weight of mango (gram)	Below 420	Below 430	Below 440	Below 450	Below 460	Below 470
No. of mangoes	14	34	76	130	165	180

This table shows cumulative frequencies of 'less than' type. We can find the frequencies from these cumulative frequencies by subtracting the frequencies of successive classes. The lower limit of first class is given as 410 grams.

Thus the frequency distribution will be obtained as follows :

Weight of mango (gram)	410 - 420	420 - 430	430 - 440	440 - 450	450 - 460	460 - 470
No. of mangoes	14	20	42	54	35	15

The mid-values of the classes are 415, 425, ..., 465.

If we choose A = 435, the values obtained as deviations –20, –10, ..., 30 will have the highest common factor c = 10. We will take $d = \frac{x-A}{c} = \frac{x-435}{10}$.

Calculation for mean is as follows :

Weight of mango (gram)	**No. of mangoes** f	**Mid-value** x	$d = \frac{x-A}{c}$ $A = 435,\ c = 10$	fd
410 - 420	14	415	–2	–28
420 - 430	20	425	–1	–20
430 - 440	42	435	0	0
440 - 450	54	445	1	54
450 - 460	35	455	2	70
460 - 470	15	465	3	45
Total	$n = 180$			**121**

Mean $\bar{x} = A + \frac{\Sigma fd}{n} \times c$

$= 435 + \frac{121}{180} \times 10$

$= 435 + \frac{1210}{180}$

$= 435 + 6.7222$

$= 441.7222$

≈ 441.72

Thus, the mean weight of mangoes is 441.72 gm.

Activity

Take A = 415 and suitable value of c in the above example and calculate mean.

Now take A = 440 and find the deviations. What is the common factor c ? Again calculate mean with these values of A and c.

Observe that all the answers for mean are same.

Illustration 9 : The distribution of annual sales tax of different companies in a zone is given below. Find the mean sales tax of these companies :

Sales tax (thousand ₹)	0 - 10	10 - 20	20 - 30	30 - 50	50 - 70
No. of companies	3	14	32	40	21

The class lengths of this distribution are not same. The mid values of classes are 5, 15, 25, 40, 60. Taking A = 25 the deviations are – 20, – 10, 0, 15, 35 which will have the highest common factor c = 5. Thus we will take $d = \frac{x-A}{c} = \frac{x-25}{5}$.

The calculation for mean in the following table :

Sale Tax (thousand ₹)	No. of companies f	Mid-value x	$d = \frac{x-A}{c}$ $A = 25, c = 5$	fd
0 - 10	3	5	–4	–12
10 - 20	14	15	–2	–28
20 - 30	32	25	0	0
30 - 50	40	40	3	120
50 - 70	21	60	7	147
	$n = 110$			**227**

Mean $\bar{x} = A + \frac{\Sigma fd}{n} \times c$

$= 25 + \frac{227}{110} \times 5$

$= 25 + \frac{1135}{110}$

$= 25 + 10.3182$

$= 35.3182$

≈ 35.32

Thus, mean sales tax paid by these companies is ₹ 35.32 thousand.

Advantages and disadvantages of mean :

Advantages :

Mean is the most popular measure of central tendency because of its following advantages.

(1) It is rigidly defined. It has a fixed mathematical formula.

(2) It is easy to understand and calculate.

(3) It is based on all observations.

(4) It is suitable for further algebraic operations.

(5) It is comparatively more stable measure. It means that means of all samples of same size from the same population have comparatively less variation.

(6) All the observations are given equal importance in the calculation of mean.

Disadvantages :

The following disadvantages should be also taken into consideration before using mean as a measure of central tendency.

(1) It is unduly affected by too large or too small values.

(2) It cannot be calculated for data having classes with open ends.

(3) Its exact value cannot be found graphically or by inspection.

(4) If few observations are missing, exact value of mean cannot be found.

(5) Mean is not a good representative value for the data which are not evenly spread around their average.

(6) Using mean as an average is not appropriate if observations have varying importance.

Some important results about mean :

(1) The sum of deviations of all observations from mean is always zero. Symbolically, the deviations can be shown as $x_i - \overline{x}$ and hence $\Sigma (x_i - \overline{x}) = 0$

For example, consider 4 values 1, 7, 5, 3.

Their mean $\overline{x} = \frac{\Sigma x}{n} = \frac{1+7+5+3}{4} = \frac{16}{4} = 4$

The deviations from mean are shown in the following table :

x	1	7	5	3	Total
$(x - \overline{x})$	–3	3	1	–1	$\Sigma (x - \overline{x}) = 0$

The sum of deviations from any other value cannot be zero.

Activity

Find the deviations of above observations from 5. What is their sum? Is it zero ?

Now, take any value of your choice other than mean and verify that the sum of deviations from this value is not zero.

(2) If each observation from $x_1, x_2, \ldots, x_n$ is multiplied by a non zero constant b and another constant a is added to it, we get a new set of observations. We will denote these values by $y_1, y_2, \ldots, y_n$, where $y_1 = bx_1 + a$, $y_2 = bx_2 + a$, ..., $y_n = bx_n + a$.

The mean of $y_1, y_2, \ldots, y_n$ will be $\overline{y} = \frac{\Sigma y}{n} = \frac{y_1 + y_2 + \ldots + y_n}{n}$

If we know mean $\overline{x}$, we can use the formula $\overline{y} = b\overline{x} + a$ to find $\overline{y}$ which is the mean of y.

Activity

Note down the ages of 10 neighbours around your house and find their mean $\overline{x}$. What will be their ages after two years ? Find the mean $\overline{y}$ of new set of numbers y you calculated. Here, age after two years for each person is $y = x + 2$. Now observe that $\overline{y} = \overline{x} + 2$

EXERCISE 3.1

1. The weekly growths (in cm.) in saplings grown in a nursery are:
1.0, 3.2, 1.4, 1.9, 2.4, 1.6, 1.4, 2.1, 1.3, 1.5.
Find the mean growth.

2. The mean age of 4 participants in the team of a relay race was calculated to be 24 years. Later it was found that one of the palticipant's age was actually 27 years, which was wrongly recorded as 25 years. lf there is rule wherein it is not possible to participate in this race if the mean age is more than 25 years, can they participate in this race even after the correction of age?

3. The following table gives the diameters (in mm.) of different screws selected from a large consignment. Find the mean diameter.

Diameter of screw (mm.)	30	35	40	45	50	55
No. of screws	4	10	15	8	5	3

4. The marks in a test for a group of students are as follows. Find the mean marks of these students.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of students	3	5	12	16	11	5	4

5. The following information is available on the talk time (in min.) noted for 70 calls of a certain mobile phone user. Find the mean talk time.

Talk time (min.)	Less than 4	Less than 8	Less than 12	Less than 16	Less than 20
No. of calls	20	42	57	65	70

6. The information of profits (in lakh ₹) of 50 firms is given below. Find mean profit.

Profit (Lakh ₹)	0–7	7–14	14–21	21–28	28–35
No. of firms	4	9	18	12	7

7. The distribution of demand of an item on different days is as follows. Find the mean demand.

Demand (units)	5–14	15–24	25–34	35–49	50–64	65–79
No. of days	4	17	19	22	18	10

*

3.3.2 Combined Mean and Weighted Mean

Combined Mean :

If we know the means of two or more groups of observations, we can find mean of the combined group. Such a value is called as combined mean. It is denoted by $\bar{x}_c$.

Suppose $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are means of k groups having $n_1, n_2, ...n_k$ observations respectively.

The formula for combined mean is as follows :

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + ... + n_k\bar{x}_k}{n_1 + n_2 + ... + n_k}$$

Illustration 10 : A factory owner knows that the mean monthly production from January to March is 350 items, from April to August it is 254 items and from September to December it is 315 items. Find the mean monthly production for that year.

Here $n_1 = 3$ months, $n_2 = 5$ months, $n_3 = 4$ months,

$\bar{x}_1 = 350$, $\bar{x}_2 = 254$, $\bar{x}_3 = 315$

Combined mean $\bar{x}_c = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$

$$= \frac{3(350) + 5(254) + 4(315)}{3 + 5 + 4}$$

$$= \frac{1050 + 1270 + 1260}{12}$$

$$= \frac{3580}{12}$$

$$= 298.3333$$

$$\approx 298.33$$

Thus, mean monthly production of the factory over the year is 298.33 items.

Illustration 11: The ratio of women and men employees in an office is 1:2. If the means of ages of women and men are 34 years and 37 years respectively, find the mean age of all the employees of the office.

Suppose the no. of women = *a.* Then the no. of men = $2a$ as the ratio is 1:2.

Taking $n_1 = a$ and $n_2 = 2a$, $\bar{x}_1 = 34$ and $\bar{x}_2 = 37$

Combined mean $\bar{x}_c = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$= \frac{a(34) + 2a(37)}{a + 2a}$$

$$= \frac{34a + 74a}{3a}$$

$$= \frac{108a}{3a}$$

$$= 36$$

Thus, mean age of all the employees in the office is 36 years.

Illustration 12: A team has to score with a run rate of 8.25 in 20 overs to win the match. The run rate at the end of 12 overs is 7.25 What should be the least run rate to be maintained in the remaining overs to win the match ?

We know that run rate $= \dfrac{\text{Total number of runs}}{\text{Total number of overs}}$

= mean of runs

Thus we will take run rate as mean of runs per over.

Here $n_1 = 12$, $n_2 = 8$

$\bar{x}_c$ = mean number of runs from all 20 overs

= 8.25 runs

$\bar{x}_1$ = mean runs in the first 12 overs = 7.25 runs.

Combined mean $\bar{x}_c = \dfrac{n_1\bar{x}_1 - n_2\bar{x}_2}{n_1 - n_2}$

$\therefore \quad 8.25 = \dfrac{12\,(7.25) + 8\bar{x}_2}{12 + 8}$

$\therefore \quad 8.25 = \dfrac{87 + 8\bar{x}_2}{20}$

$\therefore \quad 8.25 \times 20 = 87 + 8\bar{x}_2$

$\therefore \quad 165 = 87 + 8\bar{x}_2$

$\therefore \quad 8\bar{x}_2 = 165 - 87 = 78$

$\therefore \quad \bar{x}_2 = \dfrac{78}{8}$

$= 9.75$

Thus, minimum run rate for this team in the last 8 overs should be 9.75 runs to win the match.

Weighted Mean :

We said that using arithmetic mean is not appropriate if the importance of all observations is not same. A special mean called as weighted mean can be found in such cases. Weighted mean is denoted by $\bar{x}_w$. Each observation is assigned a numerical value called weight according to its importance. The most important value is given maximum weight.

Suppose $w_1, w_2, \ldots, w_n$ are the weights assigned to observations $x_1, x_2, \ldots, x_n$ respectively.

The formula for weighted mean is given as follows :

$$\text{Weighted mean} \quad \bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n}$$

$$= \frac{\Sigma w_i x_i}{\Sigma w_i}$$

Here $\Sigma w_i x_i = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$

And $\Sigma w_i = w_1 + w_2 + \ldots + w_n$

$= \text{sum of weights}$

Illustration 13: A student gets 35 marks in theory paper, 15 marks in practical examination and 5 marks in oral examination of a subject. The school gives weights 4, 2 and 1 respectively to these types of examinations. Find the weighted mean of marks for this student.

Here $x_1 = 35, \quad x_2 = 15, \quad x_3 = 5$ and

$w_1 = 4, \quad w_2 = 2, \quad w_3 = 1$

$$\text{Weighted mean} \quad \bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3}{w_1 + w_2 + w_3}$$

$$= \frac{4(35) + 2(15) + 1(5)}{4 + 2 + 1}$$

$$= \frac{140 + 30 + 5}{7}$$

$$= \frac{175}{7}$$

$$= 25$$

Thus, the weighted mean of marks of this student is 25.

EXERCISE 3.2

1. The mean daily wage paid to 75 skilled workers of a factory was ₹ 280 whereas the mean daily wage paid to 125 unskilled workers was ₹ 150. Find the mean wage of all the workers.

2. Find the weighted mean of the percentage change in prices from the following data :

Food item	**Rice**	**Wheat**	**Tea**	**Sugar**	**Pulses**
Weight	7	10	5	8	2
Percentage change in price	134	125	115	97	120

3. 2 officers, 10 clerks and 3 peons contributed for a staff picnic. The contribution collected per person is shown in the following table.

Officer	Clerk	Peon
₹ 1000	₹ 500	₹ 200

Find the mean contribution per person using weighted mean.

4. The mean marks of a student in 7 theory papers is 62. What should be the mean marks in 3 practical examinations so that his mean marks for the entire examination is 68 ?

(The marks of each theory paper and practical examination are same.)

*

3.3.3 Geometric Mean :

Suppose we are studying a variable that changes over time. If we want to find average rate of change in the variable, the application of arithmetic mean is not appropriate.

Let us understand this with an example.

Consider an item with a price of ₹ 200. If the price increases after one month by 50% and by 25% in the next month, the prices in the successive months will be $200 \times \frac{150}{100} = 300$ and $300 \times \frac{125}{100} = 375$ ₹ respectively.

If we find the average of percentage prices for the two months using arithmetic mean, it will be $\frac{150+125}{2} = 137.5$

If We take this average for finding price of that item 2 months later, it will be $200 \times \frac{137.5}{100} \times \frac{137.5}{100} = 378.18$ ₹ which is not ₹ 375 as calculated above.

Another average called as Geometric Mean is more appropriate here.

A geometric mean is defined as the *n*th root of the product of *n* positive observations and it is denoted by G.

Thus, for n observations $x_1, x_2, ..., x_n$, $G = \sqrt[n]{x_1 \times x_2 \times ... \times x_n}$

Let us find geometric mean in the above example. The two numbers 150 and 125 which show the percentage prices will have $G = \sqrt{150 \times 125} = 136.93$

Now applying this average, the price after two months will be $200 \times \frac{136.93}{100} \times \frac{136.93}{100} = 375$ ₹ which is same as the value calculated earlier.

Note: If the given variable increases by p% then increased percentage value is written as (100 + p) whereas, if the value has decreases by p% then decreased percentage value is written as (100 – p). For example, if the price decreases 20% in a certain month, the percentage price for the next month is taken as (100 – 20) = 80.

Note : For the given observations $x_1, x_2, ..., x_n$, arithmetic mean is always equal to or more than their geometric mean. That is $\bar{x} \geq G$

$\bar{x}$ = G only if all the observations have same value.

Illustration 14 : The population of an area increased by 15 %, 18 %, 13 %, 20 % respectively for four years. Find the average rate increase in the population.

Since the values of increase in population are given in percentages, we will use the geometric mean for the average value.

Considering the percentage increase in the population, we will get the observations

$x_1 = 100 + 15 = 115 \qquad x_2 = 100 + 18 = 118,$

$x_3 = 100 + 13 = 113 \qquad x_4 = 100 + 20 = 120$

$$G = \sqrt[4]{x_1 \times x_2 \times x_3 \times x_4}$$

$$= \sqrt[4]{115 \times 118 \times 113 \times 120}$$

$$= \sqrt[4]{184009200}$$

$$= \sqrt{13564.9991}$$

$$= 116.4689$$

$$\approx 116.47$$

Thus, the average rate of growth in the population over these four years is 16.47 %.

Note : To find the 4th root in this example, the square root of 184009200 is taken and its square root is taken again. Similarly, to find the 8th root of a number the process of taking square root should be repeated three times.

Illustration 15 : The geometric mean of two numbers is 2. If one number is 4 times the other number, find the numbers.

Suppose the smaller number is $x_1 = a$.

Then the larger number which is 4 times this number will be $x_2 = 4a$

and $G = 2$.

$$G = \sqrt{x_1 \times x_2}$$

$$\therefore \quad 2 = \sqrt{a \times 4a}$$

$$\therefore \quad 2 = \sqrt{4a^2}$$

$$\therefore \quad 2 = 2a$$

$$\therefore \quad a = 1$$

Thus, we get the first number $x_1 = a = 1$ and the second number $x_2 = 4a = 4$.

Illustration 16: Find the arithmetic mean and geometric mean of two numbers 9 and 16 and verify that $\bar{x} > G$.

Here $x_1 = 9$, $x_2 = 16$ and $n = 2$

Arithemetic mean $\bar{x} = \frac{\Sigma x}{n} = \frac{9+16}{2} = \frac{25}{2} = 12.5$

Geometric mean $G = \sqrt{x_1 \times x_2} = \sqrt{9 \times 16} = \sqrt{144} = 12$

Since $\bar{x} = 12.5$ and $G = 12$ we can see that $\bar{x} > G$

Advantages of geometric mean :

(1) It is rigidly defined.

(2) It is based on all observations.

(3) It is suitable for algebraic calculations.

(4) It is comparatively less affected by too large or too small values.

Disadvantages of geometric mean :

(1) It can be found only if all the observations have positive values.

(2) It is difficult to calculate.

(3) Its calculation becomes difficult if the number of observations is large.

Activity

Find the G.M. and A.M. of 4 values 1, 7, 5, 100.

Which average is more appropriate ? Why ?

EXERCISE 3.3

(1) The following data show the number of books read by 8 students of a class during last month.

2, 1, 5, 9, 1, 3, 2, 4

Find the average number of books read using geometric mean.

(2) The value of a machine depreciates at the rate of 10 % , 7 %, 5 % and 2 % in its first four years respectively. Find the average rate of depreciation using an appropriate method.

(3) A taxi travelled 15 km on Monday and 254 km on Tuesday. Find the average distance travelled over these two days using geometric mean.

*

3.4 Measures of Positional Average

Median, Quartiles, Deciles, Percentiles :

We studied that mean is an appropriate average if we have data which are evenly distributed around the average and the data which do not have too large or too small values. It is said that mean does not become a good representative of data if these conditions are not satisfied. Another average is more suitable in such situations which is called as median. It is a positional average. In addition to median, quartiles, deciles and percentiles are also other positional averages.

3.4.1 Meaning

Median, quartiles, deciles and percentiles are called positional averages because their values are found using the value of an observation at a specific position among the values of variable in the ordered data.

Median :

Median is defined as the value of middlemost observation when the data are arranged in either ascending or descending order. It is denoted by *M*. In other words, 50% values of observations in the data are above the median and 50% observations have value less than the median.

Calculation of median :

For raw data :

As we have to find the value at the centre, the observations have to be arranged in ascending or descending order.

For n observations $x_1, x_2, \ldots, x_n$, median is found as follows :

Median M = value of the $\left(\frac{n-1}{2}\right)$ th observation

For example, if we have 15 observations, the value of the $\left(\frac{15-1}{2}\right)$th that is the 8th observation will be exactly the central value, which is called as median.

Suppose the given data consists of 20 observations. Then, as $\frac{n-1}{2} = \frac{20+1}{2} = 10.5$, we say that the 10th and the 11th observations are both in the centre. In this case, median will be taken as the mean of these two central values.

Illustration 17 : The numbers of items produced by a factory in different weeks are 80, 85, 90, 92, 68 80 72, 63, 55. Find the median production.

Arrangement of observations in ascending order is as follows :

55, 63, 68, 72, 80, 80, 85, 90, 92

Here, $n = 9$

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

$= $ value of the $\left(\frac{9+1}{2}\right)$th observation

$= $ value of the 5th observation

$= 80$

Thus, the median production of this factory is 80 items.

Illustration 18 : The daily profits of a hawker (in ₹) for last 10 days are given below. Find the median profit.

261.5, 257, 258.5, 260, 265, 249, 255.5, 262.5, 264, 267

Arranging in ascending order, the observations will be as follows :

249, 255.5, 257, 258.5, 260, 261.5, 262.5, 264, 265, 267

Median $M = $ value of the $\left(\frac{n-1}{2}\right)$th observation

$= $ value of the $\left(\frac{10-1}{2}\right)$ th observation

$= $ value of the 5.5th observation

$= \frac{\text{value of the 5th observation + value of the 6th observation}}{2}$

$= \frac{260+261.5}{2}$

$= 260.75$

Thus, median daily profits of the hawker is ₹ 260.75.

Illustration 19 : There are 11 employees in an office. The monthly salaries (in ₹) of the lowest paid 7 employees among them are 4500, 2100, 3400, 3600, 2500, 4200, 1500. What is the median monthly salary of all employees ?

Some of the observations are missing here. We do not know the salaries of highest paid 4 employees.

Suppose these values are a, b, c, d in their increasing order. These values are greater than the given observations as these are salaries of highest paid employees.

We will arrange these data in ascending order. 1500, 2100, 2500, 3400, 3600, 4200, 4500, a, b, c, d.

Here, $n = 11$

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

$= $ Value of the $\left(\frac{11+1}{2}\right)$th observation

$= $ Value of the 6th observation

$= 4200$

Thus, the median salary of these employees is ₹ 4200.

For grouped data :

For discrete frequency distribution :

Suppose $x_1, x_2, ..., x_k$ are the values of a variable with frequencies $f_1, f_2, ..., f_k$ respectively.

A frequency distribution generally shows the observations arranged in ascending order. We shall use cumulative frequencies to find the median for a frequency distribution where observations are arranged in an ascending order.

Here, the median is found as follows :

Median $M =$ value of the $\left(\frac{n+1}{2}\right)$th observation

Where $n = f_1 + f_2 + ... + f_k = \Sigma f_i$ = total number of observations

Illustration 20 : The following table shows the record of absent students of class during a month. Find the median of number of absent days per student.

No. of absent days of student	**0**	**1**	**2**	**3**	**4**	**5**
No.of students	8	12	18	9	5	1

We will find cumulative frequency distribution as shown below :

No of absent days x	0	1	2	3	4	5	**Total**
No of students f	8	12	18	9	5	1	**53**
cumulative frequency cf	8	20	38	47	52	53	–

Here $n = \Sigma f = 53$

Median $M =$ value of the $\left(\frac{n+1}{2}\right)$th observation

$=$ value of the $\left(\frac{53+1}{2}\right)$th observation

$=$ value of the 27th observation

It can be known from the cumulative frequencies that the 21st to the 38th observations have value 2.

Hence, the 27th observation has value 2. $\therefore$ median $M = 2$ days.

Thus, the median number of absent days is 2 days.

Illustration 21 : The time required for typing a report by different typists is given in the following data. Find the median typing time using it.

Time for typing (min.)	**10**	**11**	**12**	**13**	**14**
No. of typists	5	7	8	15	5

The cumulative frequency distribution is as follows :

Time for typing x	10	11	12	13	14
No. of typists f	5	7	8	15	5
Cumulative frequency cf	5	12	20	35	40

Here, $n = \Sigma f = 40$

Median M = value of the $\left(\frac{n+1}{2}\right)$th observation

= value of the $\left(\frac{40+1}{2}\right)$th observation

= value of the 20.5th observation

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

> The observations of this discrete frequency distribution are as follows in their increasing order : 10, 10, 10, 10, 10, 11, 11,11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, **12, 13**, 13, 13, ... An inspection of above observations shows that the 20.5th observation is a mean of 12 and 13.

It can be known from the cumulative frequencies that the 13th to the 20th observations have value 12 and the 21th to the 35th observations have value 13.

Thus, the 20th and the 21st observations are 12 and 13 respectively.

$\therefore \quad M = \frac{12+13}{2}$

$= 12.5$

Thus, the median time required for typing is 12.5 min.

For continuous frequency distribution :

A continuous frequency gives the values of the variable in the form of class intervals and they are generally arranged in ascending order. In such cases, we will use the cumulative frequencies to find the median. These cumulative frequencies will show us the class containing median. For this, we take Median class = class containing the $\left(\frac{n}{2}\right)$th observation

Where $n = f_1 + f_2 + \ldots + f_k = \Sigma f_i$ = total number of observations

The following formula is used to find the median :

$$\text{Median } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

Where L = lower boundary point of the median class

cf = cumulative frequency of the class prior to median class

f = frequency of the median class

c = length of median class

Illustration 22 : The number of cheques deposited in a bank per day has the following frequency distribution :

No. of cheques	**0 - 39**	**40 - 79**	**80 - 119**	**120 - 159**	**160 - 199**
No. of days	2	14	23	7	4

Find the median number of cheques deposited.

We will find the cumulative frequency distribution as shown below:

No. of cheques	**0 - 39**	**40 - 79**	**80 - 119**	**120 - 159**	**160 - 199**
No. of days	2	14	23	7	4
Cumulative frequency	2	16	39	46	50

Here $n = \Sigma f = 50$

Median class = Class of the $\left(\frac{n}{2}\right)$ th observation

= Class of the $\left(\frac{50}{2}\right)$th observation

= Class of the 25th observation

It can be known from the cumulative frequencies that all the observations from the 17th observation to the 39th observation lie in 80 - 119. Hence it will be the median class.

Since this is exclusive type of classification, we will obtain the class boundary points from the class limits. Hence the median class will be taken as 79.5 - 119.5.

Taking, $L = 79.5$, $cf = 16$, $f = 23$, $c = 40$,

$$\text{Median } M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$

$$= 79.5 + \frac{25-16}{23} \times 40$$

$$= 79.5 + \frac{9}{23} \times 40$$

$$= 79.5 + \frac{360}{23}$$

$$= 79.5 + 15.6522$$

$$= 95.1522$$

$$= 95.15$$

Thus, the median for number of cheques deposited in the bank per day is 95.15 cheques.

Illustration 23: The data about monthly expenditure on petrol for 75 families is given in the following table. Find median expenditure on petrol for these families.

Expenditure (₹) on petrol	Upto 200	Upto 400	Upto 600	Upto 800	Upto1000	Upto 1200
No. of families	2	8	17	32	57	75

We are given the cumulative frequencies. We will find the frequency distribution.

Expenditure (₹)	Upto 200	200 - 400	400 - 600	600 - 800	800 - 1000	1000 - 1200
No. of families	2	6	9	15	25	18
Cumulative Frequency cf	2	8	17	32	57	75

Here $n = \Sigma f = 75$

Median class = class of the $\left(\frac{n}{2}\right)$ th observation

= class of the $\left(\frac{75}{2}\right)$th observation

= class of the 37.5th observation

It can be known from the cumulative frequencies that the 37th and the 38th observations both lie in the class 800 - 1000. Thus the median class will be 800-1000.

Taking, $L = 800,\ cf = 32,\ f = 25,\ c = 200,$

Median $M = L + \dfrac{\frac{n}{2} - cf}{f} \times c$

$= 800 + \dfrac{37.5 - 32}{25} \times 200$

$= 800 + \dfrac{5.5}{25} \times 200$

$= 800 + \dfrac{1100}{25}$

$= 800 + 44$

$= 844$

Thus, the median monthly expenditure on petrol of these families is ₹ 844.

Illustration 24: The level of air pollution (in ppm) in a city on different days is as follows. Find the median level of pollution.

Level of pollution (*ppm*)	**250 and above**	**270 and above**	**290 and above**	**310 and above**	**320 and above**	**330 and above**	**340 and above**
No. of days	150	133	108	76	41	20	7

A cumulative frequency distribution of 'more than' type is given here. We shall obtain the frequency distribution as well as the cumulative distribution of 'less than' from it.

Level of pollution	**250 - 270**	**270 - 290**	**290 - 310**	**310 - 320**	**320 - 330**	**330 - 340**	**340 and above**
No. of days	17	25	32	35	21	13	7
Cumulative frequency *cf*	17	42	74	109	130	143	150

$n = \Sigma f = 150$. Here, the classes have unequal lengths.

Median class = class of the $\left(\frac{n}{2}\right)$ th observation

= class of the $\left(\frac{150}{2}\right)$th observation

= class of the 75th observation

It can be known from the cumulative frequencies that the 75th observation lies in the class 310 - 320.

Hence median class will be 310 - 320.

Taking $L = 310$, $cf = 74$, $f = 35$, $c = 10$;

Median $$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$

$$= 310 + \frac{75-74}{35} \times 10$$

$$= 310 + \frac{1}{35} \times 10$$

$$= 310 + \frac{10}{35}$$

$$= 310 + 0.2857$$

$$= 310.2857$$

$$\approx 310.29$$

Thus, the median for level of pollution 310.29 *ppm*

Advantages and disadvantages of median :

Advantages :

(1) It is easy to calculate and to understand.

(2) It can be found by inspection.

(3) It can be located by graph.

(4) It is the only available average when the frequency distribution has open ended classes.

(5) It is less affected by too large or too small values.

(6) It can be calculated even if certain data are missing.

Disadvantages :

(1) It is not rigidly defined.

(2) It is not based on all values.

(3) It is not suitable for further algebraic operations.

(4) It is less stable measure of central tendency as compared to mean.

Other positional averages :

We saw that the median divides the data in two equal parts. Sometimes we require values that divide data in more parts. We shall now study a few such positional averages.

Quartiles :

When the observations of the given data are arranged in ascending order, three values which divide the data in four equal parts are called quartiles. These three quartiles are denoted by Q_1, Q_2 and Q_3 respectively.

First 25% values of the data will be less than or equal to Q_1, next 25% values of the data will be between Q_1 and Q; and the further 25% values will be between Q; and Q_3. Hence we have 50% data values lying between Q_1 and Q_3 whereas 25% observations have value above Q_3.

We can say that the *j*th quartile Q_j will divide the data such that 25*j*% observations will be below Q_j. ($j = 1, 2, 3$)

Thus, quartile Q_2 will have (25 × 2)% or 50% observations below Q_2. Hence Q_2 = Median = M

For median :

50 %	50 %

Lowest Value — M — Highest Value

For quartiles :

	25 %		25 %		25 %		25 %	
Lowest Value		Q_1		Q_2		Q_3		Highest Value

Deciles :

Suppose the observations are arranged in the ascending order. Nine values which will divide the data in ten equal parts are called as deciles which are denoted by D_1 , D_2, ..., D_9 respectively. 10% observations will have value less than D_1, 20% observations will have value less than D_2 and so on. Thus, 10 j % observations will have value less then the jth decile D_j (j = 1, 2, ...9).

We can see that $D_5 = M = Q_2$.

Percentiles :

Suppose the observations are arranged in the ascending order. The 99 values which divide the data in 100 equal parts are called percentiles which are denoted by P_1 , P_2 ..., P_{99} respectively. Here, 100 j % observations will have value less than the jth percentile P_j (j = 1, 2, ...,99).

We can see that $D_1 = P_{10}$, $D_2 = P_{20}$, ..., $D_9 = P_{90}$ Similarly, $Q_1 = P_{25}$ and $Q_3 = P_{75}$

Moreover, $M = Q_2 = D_5 = P_{50}$

Since median, quartiles, deciles and percentiles are all positional averages, the calculation will have similar method. The following table shows the formula for finding the jth quartile Q_j , the jth decile D_j and the jth percentile P_j.

Type of Data	***j*th Quartile** ***j* = 1, 2, 3**	***j*th Decile** ***j* = 1, 2,...9**	***j*th Percentile** ***j* = 1, 2,..., 99**
Raw data and Discrete frequency distribution	Q_j = value of the $j\left(\frac{n+1}{4}\right)$th observation	D_j = value of the $j\left(\frac{n+1}{10}\right)$th observation	P_j = value of the $j\left(\frac{n+1}{100}\right)$th observation
Continuous frequency distribution	Class of Q_j = class of the $j\left(\frac{n}{4}\right)$th observation $Q_j = L + \frac{j\left(\frac{n}{4}\right) - cf}{f} \times c$ Where L = lower boundary point of class of Q_j cf = cumulative frequency of the class prior to class of Q_j f = frequency of the class of Q_j c = length of class of Q_j	Class of D_j = class of the $j\left(\frac{n}{10}\right)$th observation $D_j = L + \frac{j\left(\frac{n}{10}\right) - cf}{f} \times c$ Where L = lower boundary point of class of D_j cf = cumulative frequency of the class prior to class of D_j f = frequency of the class of D_j c = length of class of D_j	Class of P_j = class of the $j\left(\frac{n}{100}\right)$th observation $P_j = L + \frac{j\left(\frac{n}{100}\right) - cf}{f} \times c$ Where L = lower boundary point of class of P_j cf = cumulative frequency of the class prior to class of P_j f = frequency of the class of P_j c = length of class of P_j

Illustration 25: Find Q_1, D_7, P_{40} for the following data showing runs scored by a batsman in his 20 innings.

32, 28, 47, 63, 71, 9, 60, 10, 96, 14, 31, 148, 53, 67, 29, 10, 62, 40, 80, 54

Arrangement of observations in ascending order is as following :

9, 10, 10, 14, 28, 29, 31, 32, 40, 47,

53, 54, 60, 62, 63, 67, 71, 80, 96, 148

Here, $n = 20$

> While finding the positional average, if the value of its position is a fraction, then linear approximation is used to find the value of the positional average. For example, to find the 5.25th value, we add 25% of the distance between the 5th observation and the next that is the 6th observation to the 5th observation.

Quartile Q_1 = value of the $\left(\frac{n-1}{4}\right)$ th observation

= value of the $\left(\frac{20+1}{4}\right)$th observation

= value of the 5.25th observation

= value of the 5th observation + 0.25(value of the 6th observation – value of the 7th observation)

= 28 + 0.25(29 – 28)

= 28 + 0.25

= 28.25

Decile D_7 = value of the $7\left(\frac{n+1}{10}\right)$ th observation

= value of the $7\left(\frac{20+1}{10}\right)$ th observation

= value of the 14.7th observation

= value of the 14th observation + 0.7(value of the 15th observation – value of the 14th observation)

= 62 + 0.7(63 – 62)

= 62 + 0.7

= 62.7

Percentile P_{40} = value of the $40\left(\frac{n+1}{100}\right)$th observation

= value of the $40\left(\frac{20+1}{100}\right)$th observation

= value of the 8.4th observation

= value of the 8th observation + 0.4(value of the 9th observation – value of the 8th observation)

= 32 + 0.4(40 – 32)

= 32 + 3.2

= 35.2

Thus the values of Q_1 , D_7 , P_{40} are 28.25 runs, 62.7 runs, and 35.2 runs respectively.

Illustration 26 : The following data refer to the milk quantity in 90 bags filled by a machine in a dairy.

Milk content (ml.)	**485 - 490**	**490 - 495**	**495 - 500**	**500 - 505**	**505 - 510**
No. of bags	5	21	33	23	8

Find Q_3, D_2, P_{55} and interpret them.

Let us find the table of cumulative frequencies for this continuous frequency distribution.

Milk content (ml.)	**485 - 490**	**490 - 495**	**495 - 500**	**500 - 505**	**505 - 510**
No. of bags	5	21	33	23	8
Cumulative frequency *cf*	5	26	59	82	90

Here, $n = 90$

For quartile Q_3 :

$$\text{Class of } Q_3 = \text{Class of the } 3\left(\frac{n}{4}\right)\text{th observation}$$

$$= \text{Class of the } 3\left(\frac{90}{4}\right)\text{th observation}$$

$$= \text{Class of the 67.5th observation}$$

It can be known from the cumulative frequencies that the 67th and the 68th observations are in the class 500 - 505.

Taking L = 500, $cf = 59$, $f = 23$, $c = 5$,

$$\text{Quartile } Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

$$= 500 + \frac{67.5 - 59}{23} \times 5$$

$$= 500 + \frac{8.5}{23} \times 5$$

$$= 500 + \frac{42.5}{23}$$

$$= 500 + 1.8478$$

$$= 501.8478$$

$$\approx 501.85$$

Thus, maximum quantity of milk is 501.85 ml in 75 % bags having least milk content.

For decile D_2 :

$$\text{Class of } D_2 = \text{Class of the } 2\left(\frac{n}{10}\right)\text{th observation}$$

$$= \text{Class of the } 2\left(\frac{90}{10}\right)\text{th observation}$$

$$= \text{Class of the 18th observation}$$

It can be found from the cumulative frequencies that the 18th observation is in the class 490-495.

Taking $L = 490,\ cf = 5,\ f = 21,\ c = 5,$

$$D_2 = L + \frac{2\left(\frac{n}{10}\right) - cf}{f} \times c$$

$$= 490 + \frac{18-5}{21} \times 5$$

$$= 490 + \frac{13}{21} \times 5$$

$$= 490 + \frac{65}{21}$$

$$= 490 + 3.0952$$

$$= 493.0952$$

$$\approx 493.1$$

Thus, maximum quantity of milk is 493.1 ml in 20% bags having least milk content.

For percentile P_{55} :

Class of P_{55} = class of the $55\left(\frac{n}{100}\right)$th observation

= class of the $55\left(\frac{90}{100}\right)$th observation

= class of the 49.5th observation

It can be found from the cumulative frequencies that the 49th and the 50th observations are in the class 495-500.

Taking $L = 495,\ cf = 26,\ f = 33,\ c = 5,$

$$P_{55} = L + \frac{55\left(\frac{n}{100}\right) - cf}{f} \times c$$

$$= 495 + \frac{49.5-26}{33} \times 5$$

$$= 495 + \frac{23.5}{33} \times 5$$

$$= 495 + \frac{117.5}{33}$$

$$= 495 + 3.5606$$

$$= 498.5606$$

$$\approx 498.56$$

Thus, maximum quantity of milk is 498.56 ml in 55% bags having least milk content.

Illustration 27 : We have the following information from a survey of 100 customers of a bank on their number of visits to the bank during a month.

No. of visists	**0**	**1**	**2**	**3**	**4**	**5**	**6**
No. of customers	12	22	40	15	6	4	1

Find the first quartile, 4th decile and 95th percentile.

No. of visits	0	1	2	3	4	5	6
No. of customers	12	22	40	15	6	4	1
Cumulative frequency *cf*	12	34	74	89	95	99	100

Here, $n = 100$

First qurtile Q_1 = Value of the $\left(\frac{n+1}{4}\right)$th observation

= Value of the $\left(\frac{100+1}{4}\right)$th observation

= Value of the 25.25th observation

According to cumulative frequencies, the 25th and the 26th observations have value 1. $\therefore$ $Q_1 = 1$
Thus, maximum number of visits is 1 among 25% least visiting customers.

4th Decile D_4 = Value of the $4\left(\frac{n-1}{10}\right)$th observation

= Value of the $4\left(\frac{100+1}{10}\right)$th observation

= Value of the 40.4th observation

According to cumulative frequencies, the 40th and the 41st observations have value 2. $\therefore$ $D_4 = 2$

Thus, maximum number of visits is 2 among 40% least visiting customers.

95th Percentile P_{95} = Value of the 95 $\left(\frac{n+1}{100}\right)$th observation

= Value of the 95 $\left(\frac{100+1}{100}\right)$th observation

= Value of the 95.95th observation

According to cumulative frequencies, the 95th observation has value 4 and the 96th observation has value 5.

$\therefore$ P_{95} = Value of the 95th observation + 0.95 (Value of the 96th observation – Value of the 95th observation)

= 4 + 0.95(5 – 4)

= 4 + 0.95

= 4.95

Thus, maximum number of visits is 4.95 $\approx$ 5 among 95% least visiting customers.

Illustration 28 : The following table shows data about monthly travelling expenses (in ₹) in a sample of 50 local bus commuters.

Monthly expenses	**350 - 500**	**500 - 650**	**650 - 800**	**800 - 950**	**950 and above**
No. of commuters	7	13	16	9	5

(1) Find the limits for the expense of central 50% of the commuters.

(2) Find the minimum expense of the highest spending 15% commuters.

(3) Find the maximum expense in the lowest spending 10% commuters.

We will use the concepts of positional averages to answer the above questions. For this we find the cumulative frequencies for this continuous frequency distribution.

Monthly expenses(₹)	**350 - 500**	**500 - 650**	**650 - 800**	**800 - 950**	**950 and above**
No. of commuters	7	13	16	9	5
Cumulative frequeny *cf*	7	20	36	45	50

Here, $n = 50$

(1) For any variable, central 50% observations have values between Q_1 and Q_3. Hence we will find Q_1 and Q_3.

Class of Q_1 = class of the $\left(\frac{n}{4}\right)$th observation

= class of the $\left(\frac{50}{4}\right)$th observation

= class of the 12.5th observation

It can be found from the cumulative frequencies that the 12th and the 13th observations are in the class 500-650.

Taking $L = 500$, $cf = 7$, $f = 13$, $c = 150$,

$$Q_1 = L + \frac{\left(\frac{n}{4}\right) - cf}{f} \times c$$

$$= 500 + \frac{12.5 - 7}{13} \times 150$$

$$= 500 + \frac{5.5}{13} \times 150$$

$$= 500 + \frac{825}{13}$$

$$= 500 + 63.4615$$

$$= 563.4615$$

$$\approx 563.46$$

Class of Q_3 = class of the $3\left(\frac{n}{4}\right)$th observation

= class of the $3\left(\frac{50}{4}\right)$th observation

= class of the 37.5th observation

It can be known from the cumulative frequencies that the 37th and the 38th observations are in the class 800 - 950.

Taking $L = 800$, $cf = 36$, $f = 9$, $c = 150$,

$$Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

$$= 800 + \frac{37.5 - 36}{9} \times 150$$

$$= 800 + \frac{1.5}{9} \times 150$$

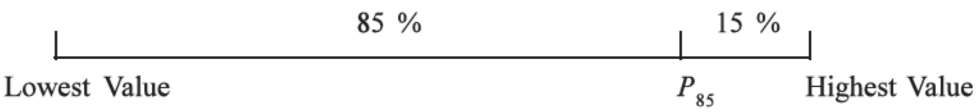
$$= 800 + \frac{225}{9}$$

$$= 800 + 25$$

$$= 825$$

Thus, central 50% commuters have monthly travelling expenses between ₹ 563.46 and ₹ 825.

(2) Here we want a value so that 15% observations should be above it, which means 85% observations should have value below it. Hence we find P_{85} .



$$\text{Class of } P_{85} = \text{Class of the } 85\left(\frac{n}{100}\right)\text{th observation}$$

$$= \text{Class of the } 85\left(\frac{50}{100}\right)\text{th observation}$$

$$= \text{Class of the } 42.5\text{th observation}$$

It can be known from the cumulative frequencies that the 42nd and the 43rd observations are in the class 800 - 950.

Taking $L = 800$, $cf = 36$, $f = 9$, $c = 150$,

$$\text{Class of } P_{85} = \text{Class of } L + \frac{85\left(\frac{n}{100}\right) - cf}{f} \times c$$

$$= 800 + \frac{42.5 - 36}{9} \times 150$$

$$= 800 + \frac{6.5}{9} \times 150$$

$$= 800 + \frac{975}{9}$$

$$= 800 + 108.3333$$

$$= 908.3333$$

$$\approx 908.33$$

Thus, minimum expense is ₹ 908.33 among highest spending 15% commuters.

(3) To find a value so that 10% observations are below it, we find D_1.

| 10 % | 90 % |

Lowest value D_1 Highest value

Class of D_1 = class of the $\left(\frac{n}{10}\right)$th observation

= class of the $\left(\frac{50}{10}\right)$th observation

= class of the 5th observation

It can be found from the cumulative frequencies that the 5th observation is in the class 350 – 500.

Taking $L = 350$, $cf = 0$, $f = 7$, $c = 150$,

$$D_1 = L + \frac{\left(\frac{n}{10}\right) - cf}{f} \times c$$

$$= 350 + \frac{5-0}{7} \times 150$$

$$= 350 + \frac{5}{7} \times 150$$

$$= 350 + \frac{750}{7}$$

$$= 350 + 107.1429$$

$$= 457.1429$$

$$\doteq 457.14$$

Thus, the maximum expense among the lowest spending 10 % commuters is ₹ 457.14.

EXERCISE 3.4

1. Find all quartiles for the data given below about marks scored by 15 students in class test.

8, 6, 7, 0, 2, 4, 6, 5, 5, 4, 8, 9, 3, 6, 7

2. The following table shows data about the distance travelled (in km) by a salesman on different days. Find median, Q_3, D_8, P_{62} and interpret them.

Distance travelled (km)	0–100	100–200	200–300	300–400	400–500	500–600
No. of days	5	18	24	7	5	1

3. The following table gives ages of 80 students selected from a college.

Age (years)	17	18	19	20	21	22	23
No. of students	11	14	22	15	8	6	4

Find median age. Also find Q_1, D_4, P_{32} for age and interpret them.

4. Use of following data to find the median salary of employees in a firm. Also find the lower limit for the richest 20% employees.

Salary (thousand ₹)	5 or more	10 or more	15 or more	20 or more	25 or more	30 or more
No. of employees	120	117	106	76	31	12

5. The following table shows the monthly expense for entertainment in a group of 100 students. Find the median of this expense.

Expense (₹)	Less than 200	200 - 400	400 - 600	600 - 700	700 - 800	800 and above
No. of students	8	23	40	17	7	5

6. The following data indicate records of hospital stays (in days) of 30 patients admitted to a hospital.

1, 10, 2, 6, 3, 4, 15, 1, 5, 9, 2, 4, 3, 1, 10,
7, 3, 5, 4, 2, 4, 8, 5, 3, 1, 9, 6, 2, 3, 7

Find the median stay. Further convert this information in a continuous frequency distribution (inclusive type) by taking classes of equal length starting from 1 –3. Find the median from the frequency distribution and compare it with your earlier answer.

*

3.5 Mode

We have earlier studied the mean and the median as the measures of central tendency. We shall now study 'mode' as another measure which is extensively used in business and commercial fields.

3.5.1 Meaning :

The value which gets repeated maximum number of times or the value occurring with maximum frequency in the given data is called as **mode.** It is denoted by M_o.

It is very often used in business to give a representative value for a set of data. For example, see the following statements :

(1) On an average 3 languages are known to the students of this school.
(2) The average height of the men in our country is 1.7 m.
(3) The average daily production of our company is 50 items.
(4) The average daily overtime put in by the workers of a factory is 3 hours.

The value which is repeated most number of times is considered in the calculation of the average in these situations. As per the first statement, it is implied that most of the students know three languages. Thus we can say that mode is used as an average here.

Calculation of mode :

For raw data and for discrete frequency distribution :

In these cases the mode can be found simply by inspection. We can find the mode as a value among the observations which is repeated maximum number of times or the one which has maximum frequency.

Illustration 29 : The numbers of books purchased by each of the 15 persons from a book store are as follows.

1, 0, 2, 2, 3, 4, 2, 7, 2, 2, 5, 4, 2, 1, 2.

Find the modal value for the number of books purchased.

We can see that the value of 2 is repeated 7 times which is more than the number of repetitions of any value of the other observations. Hence mode $M_o = 2$.

Thus, the mode of the number of books purchased is 2.

Illustration 30 : TV sets assembled by a TV manufacturing company in a month are tested. The following table shows the numbers of defects per TV set. Find the mode for the number of defects.

No. of defects	0	1	2	3	4	5
No. of TV sets	45	22	18	10	6	4

An inspection of frequencies shows that the observation 0 has the maximum frequency 45.

Hence $M_o = 0$

Thus the mode for the number of defects in TV sets is 0.

Note : According to the definition, the value of mode in this illustration is 0. But it cannot be taken as a measure of central tendency for the data as the value of mode is in the beginning of the data.

Mean or Median should be chosen as the measures of central tendency for such data or the value of mode should be found using empirical formula based on the values of mean and median which will be discussed in the later part of this chapter.

Illustration 31 : The number of trips made by 24 taxi drivers in a day are shown in the following data. Find the modal number of trips.

No. of trips	1	2	4	5	6	7
No. of drivers	3	7	4	7	2	1

The maximum frequency is 7 which is the frequency for the observations 2 and 5. Here, we say that this distribution has two modes $M_o = 2$ and $M_o = 5$.

Thus, the modes for the number of trips by the taxi driver are 2 and 5.

Note : Such a distribution is called as a bimodal distribution. Similarly there can be distributions with more than two modes.

Illustration 32: The number of patients arriving at a clinic each hour during working hours of a day are recorded as follows.

3, 5, 4, 2, 7, 8

Find the mode for number of patients.

As all the values are appearing only once, we can't find the most common observation.

Hence the mode for the number of patients cannot be found from the given data using the definition.

For continuous frequency distribution :

When the data are converted into a frequency distribution with classes, the exact values of the observations are not available.

Similar to median, for mode also, the class containing mode is found first and the value of mode is found using it.

The class having maximum frequency is called as modal class of the frequency distribution. The mode is further obtained using the following formula.

$$\text{Mode } M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

Where L = lower boundary point of the modal class

f_m = frequency of the modal class

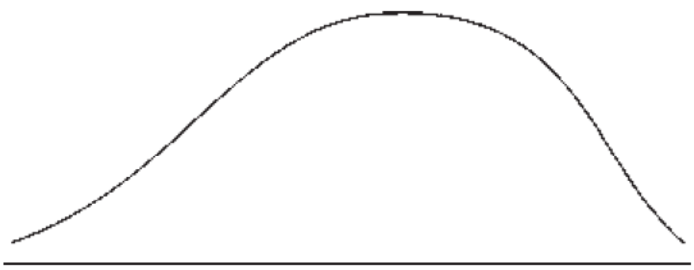
f_1 = frequency of the class prior to modal class

f_2 = frequency of the class succeeding to modal class

c = class length of the modal class

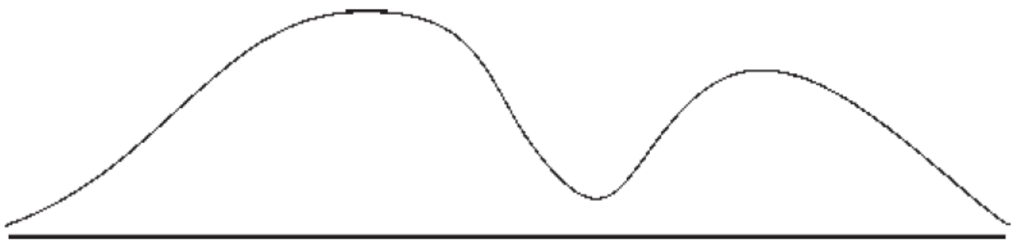
Note : The above formula can be used only if the distribution has classes of equal class length. Moreover, the formula can be used only in those cases where the maximum frequency is only for one class.

The frequency distribution in which the frequencies increase initially and then start decreasing after attaining the maximum frequency is called as a **regular frequency distribution**. Such distributions are also called as **unimodal distributions** as the distribution has only one mode. The frequency curve of such distributions is as follows :



Frequency curve of regular distribution

For bimodal distribution, the frequencies increase and then decrease but then again increase and decrease. Such a frequency distribution is called as an **irregular frequency distribution** whose frequency curve is as follows.



Frequency curve of irregular distribution

Illustration 33 : The following table gives output of workers in a factory. Find the modal output.

Output (no. of items)	150 - 160	160 - 170	170 - 180	180 - 190	190 - 200	200 - 210	210 - 220	220 - 230
No. of workers	4	5	19	33	48	22	12	6

The maximum frequency is 48 for the class 190-210. Hence the modal class is 190-200.

Taking, $L = 190, f_m = 48, f_1 = 33, f_2 = 22, c = 10,$

Mode $\quad M_0 = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$

$= 190 + \frac{48-33}{2(48)-33-22} \times 10$

$= 190 + \frac{15}{96-33-22} \times 10$

$= 190 + \frac{150}{41}$

$= 190 + 3.6585$

$= 193.6585$

$\fallingdotseq 193.66$

Thus, the modal output is 193.66 items.

Illustration 34 : The number of cold drink bottles sold by a shopkeeper on different days is shown in the following table. Find the mode for the number of cold drink bottles sold.

No. of bottles	0 - 3	4 - 7	8 - 11	12 - 15	16 - 19	20 - 23
No. of days	3	11	16	20	18	12

The maximum frequency is 20 for the class 12 - 15. Hence the modal class is 12 - 15. This is an inclusive type of frequency distribution and hence we will take the boundary points of this class as 11.5 - 15.5.

Taking, $\quad L = 11.5, f_m = 20, f_1 = 16, f_2 = 18, c = 4,$

Mode $\quad M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$

$= 11.5 + \frac{20-16}{2(20)-16-18} \times 4$

$= 11.5 + \frac{4}{40-16-18} \times 4$

$= 11.5 + \frac{16}{6}$

$= 11.5 + 2.6666$

$= 14.1666$

$\fallingdotseq 14.17$

Thus, mode for sale of cold drink bottles is 14.17.

Empirical formula for mode :

We have observed that mode is not well defined in many cases. The noted statistician Karl Pearson established a relation between mean, median and mode by studying their values for different data sets. He observed that for data that are not evenly distributed around average, difference between mean and mode is approximately 3 times the difference between mean and median.

That is, (Mean – Mode) = 3 (Mean – Median)

The following formula is obtained to find the mode using this relation:

Mode = 3 (Median) – 2(Mean)

This is written in notations as $M_o = 3M - 2\overline{x}$

This formula to find mode is called as an empirical formula because the value obtained from observation and not from the theory. The value of mode found using this formula can be negative if the frequency distribution is not evenly distributed around the average.

This formula for mode is used in the following situations :

- Each observation appears just once in raw data.
- More than one observation in a frequency distribution appears with highest frequency.
- The continuous distribution has classes of unequal length.
- The frequency distribution is a mixed distribution that is a part of it is discrete and the rest is continuous.
- The right or left end of the curve of the frequency distribution is too extended.

Activity

Find the mean and median for the data given in Illustration 34 and verify the empirical formula.

Illustration 35 : The time between placing an order and its delivery was noted for a certain wholesaler as follows. Find the mode for this time.

Time (hours)	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 55
No. of orders	2	5	7	5	6	7	3

We can see that the maximum frequency is 7 which appears for two classes. Hence we shall use the empirical formula for finding the mode.

The following calculations are carried out to find mean and median :

Time (hrs.)	No. of orders f	Mid-value x	$d = \frac{x-A}{c}$ $A = 37.5$ $c = 5$	fd	Cumulative frequency cf
20 - 25	2	22.5	– 3	– 6	2
25 - 30	5	27.5	– 2	– 10	7
30 - 35	7	32.5	– 1	– 7	14
35 - 40	5	37.5	0	0	19
40 - 45	6	42.5	1	6	25
45 - 50	7	47.5	2	14	32
50 - 55	3	52.5	3	9	35
Total	$n = 35$			**6**	

Mean $\bar{x} = A + \frac{\Sigma fd}{n} \times c$

$= 37.5 + \frac{6}{35} \times 5$

$= 37.5 + \frac{30}{35}$

$= 37.5 + 0.8571$

$= 38.3571$

≈ 38.36

Median class $=$ Class containing the $\left(\frac{n}{2}\right)$th observation

$=$ Class containing the $\left(\frac{35}{2}\right)$th observation

$=$ Class containing the 17.5th observation

It can be known from the cumulative frequencies that, the 17th and the 18th observations lie in the class 35 - 40.

Taking $L = 35$, $cf = 14$, $f = 5$, $c = 5$,

$$\text{Median } M = L + \frac{\frac{n}{2} - cf}{f} \times c$$

$= 35 + \frac{17.5 - 14}{5} \times 5$

$= 35 + 3.5$

$= 38.5$

Using the empirical formula, $M_o = 3M - 2\bar{x}$

$= 3(38.5) - 2(38.36)$

$= 115.5 - 76.72$

$= 38.78$

Thus, the mode for time between placing of an order and delivery is 38.78 hours.

Illustration 36 : The following table shows the number of visits to the dentist of the persons having toothache. Find the mode for number of visits.

No. of visits	1	2	3	4 - 7	7 - 10	10 - 15
No. of patients	7	11	18	9	4	1

The given distribution is mixed distribution and the class lengths are unequal. Thus, we will use the empirical formula for finding the mode. The calculations for mean and median are shown in the following table.

No. of visists	No. of patients f	Mid-value x	fx	Cumulative Frequency cf
1	7	1	7	7
2	11	2	22	18
3	18	3	54	36
4–7	9	5.5	49.5	45
7–10	4	8.5	34	49
10–15	1	12.5	12.5	50
Total	$n = 50$		**179**	

Note : The values of x are not too large. Hence we need not use the short cut method.

Mean $\bar{x} = \frac{\sum fx}{n} = \frac{179}{50} = 3.58$

Median = Value of the $\left(\frac{n}{2}\right)$th observation

= Value of the $\left(\frac{50}{2}\right)$th observation

= Value of the 25th observation

It can be known from the cumulative frequencies that the 25th observation has value 3.

$\therefore M = 3$ visits

Using the empirical formula, $M_o = 3M - 2\bar{x}$

$= 3(3) - 2(3.58)$

$= 9 - 7.16$

$= 1.84$

Thus, the mode for the number of visits to the dentist will be 1.84.

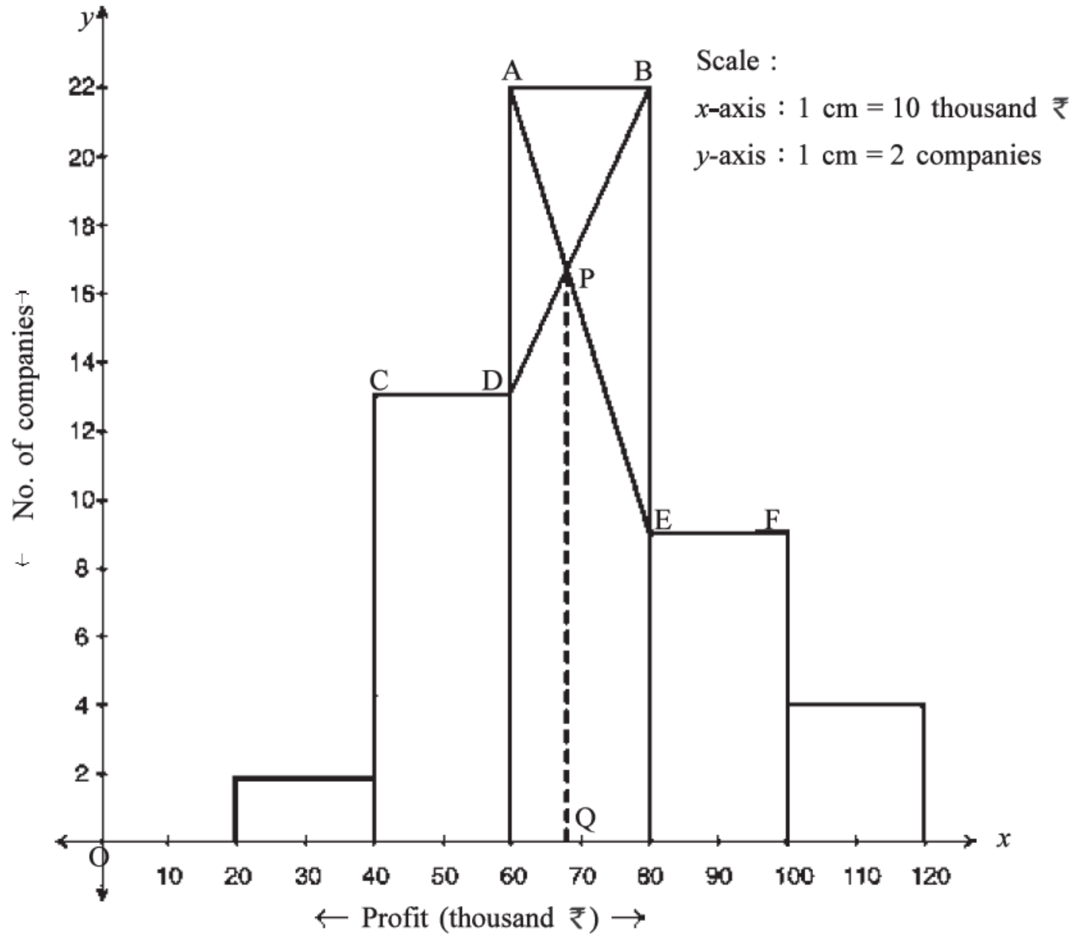
3.5.2 Graphical Method

The commonly used formula for mode (page no. 101) can not be used in a continuous frequency distribution having classes with unequal length. The mode can be found graphically for a continuous frequency distribution having classes with equal or unequal lengths. The histogram of the frequency distribution is used here. This method can be used only for unimodal distributions.

We shall understand this method with the help of the following data which are about the profits (in thousand ₹) of 50 small scale industrial units.

Profit (thousand ₹)	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120
No. of compaines	2	13	22	9	4

The histogram for this frequency distribution is as follows :



To find the mode, we first consider the rectangle with maximum length which corresponds to the class interval with maximum frequency. In this case, it is the interval 60 - 80. Suppose the upper side of this rectangle is shown as AB. Now we consider the two rectangles adjacent to this rectangle. They will be the rectangles for the class intervals 40 - 60 and 80 - 100 respectively. We shall denote the upper sides of these rectangles as CD and EF respectively. Now we will draw a line segments AE joining points A and E as well as a segment BD. The point of intersection of AE and BD will be denoted by P. The point where a perpendicular drawn from the point P on x- axis meets x-axis will be denoted by Q. The distance of the point Q from the origin O will give us the value of mode. It can be seen from the above histogram that OQ = 68 (thousand ₹).

Hence, the mode of profits of these companies is ₹ 68 thousand.

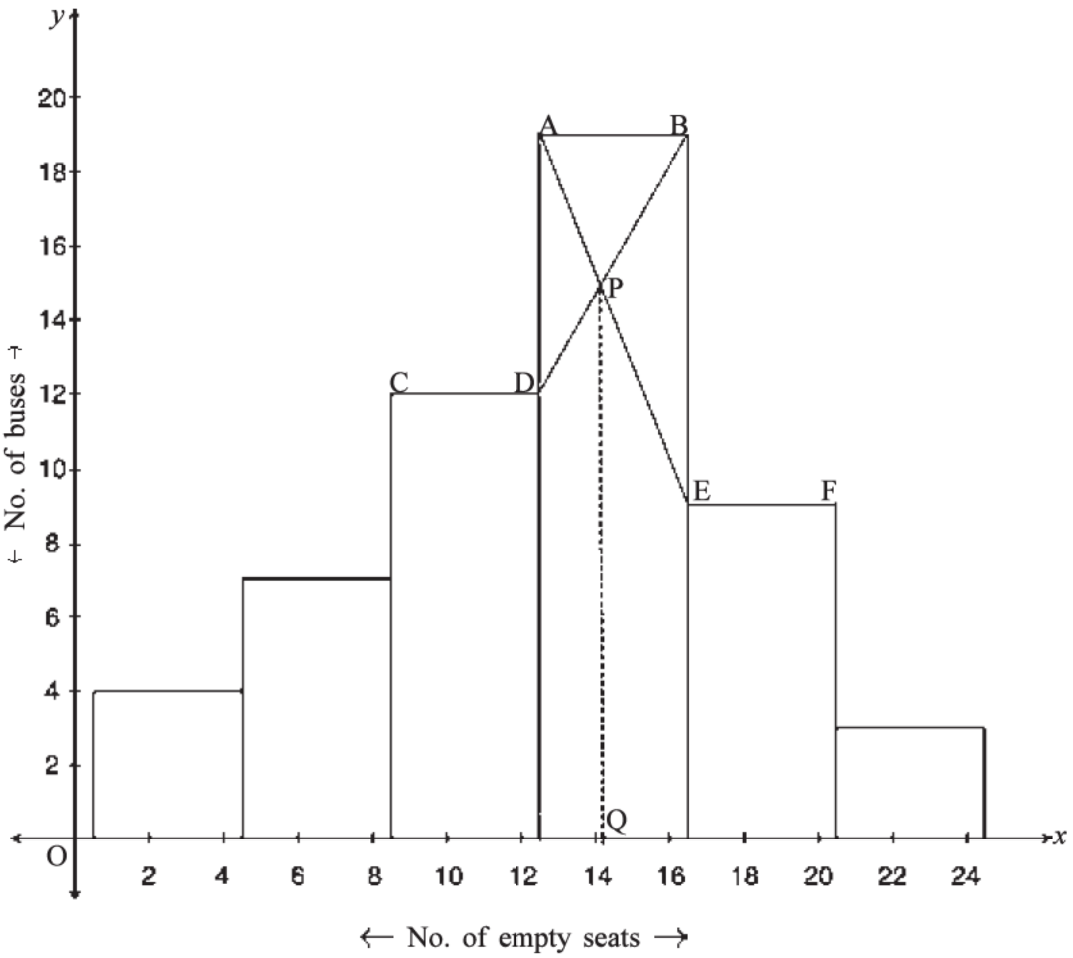
Illustration 37 : The following data show the number of empty seats per bus for the buses leaving from a depot. Find the modal number of empty seats using the graphical method.

No. of empty seats x	1 - 4	5 - 8	9 - 12	13 - 16	17 - 20	21 - 24
No. of buses	4	7	12	19	9	3

We can see that this is a continuous distribution of inclusive type. Therefore, first we find the class boundaries to draw the histogram.

No. of empty seats x	0.5–4.5	4.5–8.5	8.5–12.5	12.5–16.5	16.5–20.5	20.5–24.5
No. of buses	4	7	12	19	9	3

The histogram for this distribution is drawn below :



Following the method explained above, we find that OQ = 14.2. Hence modal number of empty seats for the buses leaving the depot is 14.2.

Illustration 38 : The distribution of the share prices of a company on different days is given in the following table. Find the mode for the share prices using the graphical method.

Share price (₹)	**200 - 210**	**210 - 220**	**220 - 240**	**240 - 260**	**260 - 300**
No. of days	4	13	36	16	8

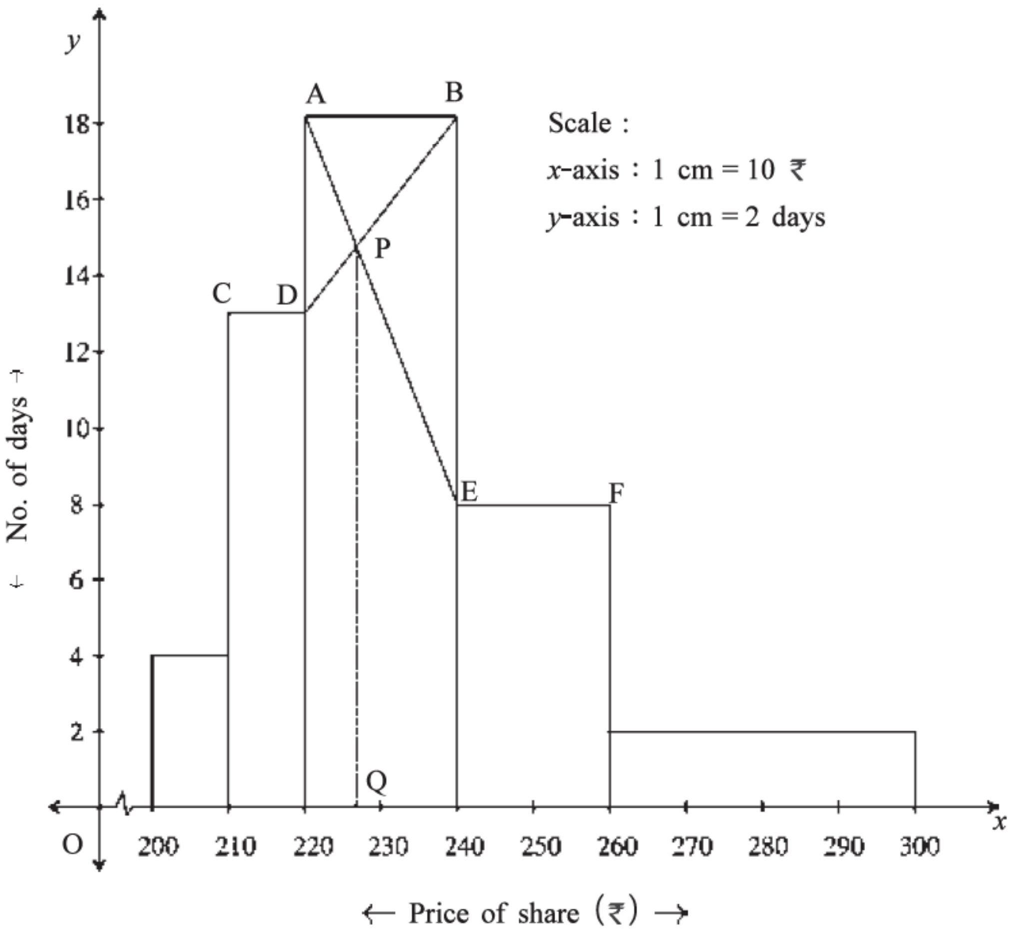
We have to draw the histogram to find the mode by graphical method. This frequency distribution does not have equal class lengths and hence we first find the proportionate frequencies with respect to the smallest class length which are shown in the following table :

Share price	**Class length**	**Frequency**	**Proportionate frequency**
200 - 210	10	4	$\frac{4}{10} \times 10 = 4$
210 - 220	10	13	$\frac{13}{10} \times 10 = 13$
220 - 240	20	36	$\frac{36}{20} \times 10 = 18$
240 - 260	20	16	$\frac{16}{20} \times 10 = 8$
260 - 300	40	8	$\frac{8}{40} \times 10 = 2$

Proportionate frequency of each class = $\frac{\text{class frequency}}{\text{class length}}$ × minimum class length

Here the minimum class length is 10.

The histogram drawn using the proportionate frequencies is as follows :



Using the graphical method, OQ = 227 ₹

Hence, the mode for share prices is ₹ 227.

Advantages and disadvantages of mode :

Advantages :

(1) It is easy to understand and calculate.

(2) It can be found merely by inspection.

(3) It is not affected by too large or too small values.

(4) Its value can be found using graph.

Disadvantages :

(1) It is not rigidly defined.

(2) There can be more than one mode for the given variable whereas sometimes the mode cannot be found.

(3) It is not based on all observations.

(4) It has less stability in sampling as compared to mean.

(5) It is not suitable for further mathematical calculations.

EXERCISE 3.5

1. The IQ levels of students in a class are given below. Find the modal value of IQ level of students.
   146, 134, 143, 144, 138, 145, 153, 138, 138, 146, 140, 135.
2. The following table gives the number of cakes sold each day at a bakery. Find the mode for sale of cakes.

No. of cakes	10	12	13	16	17	18
No. of days	5	9	25	16	10	7

3. The distribution of ages of 48 persons in an old age home is given below. Which formula will be appropriate to find the mode ? Why ? Find the modal age of the persons in the old age home using the formula you have chosen.

Age (years)	50 - 60	60 - 65	65 - 70	70 - 85	85 - 100
No. of persons	6	10	19	9	4

4. Comment on the mode for the following data showing the time taken (in seconds) for 8 competitors in a running race.
   25.2, 26.5, 28.6, 32.1, 29.0, 29.3, 31.3, 27.8
5. The table below shows the data about weights of 86 apples from a garden. Find the mode for the Weight of apples.

Weight of apple (gram)	120 - 130	130 - 140	140 - 150	150 - 160	160 - 170	170 - 180	180 - 190
No. of apples	8	13	19	23	10	8	5

   Also find the mode of this distribution graphically.
6. The data about the monthly house rent paid by 50 families is given in the following table :

House rent (thousand ₹)	0–5	5–10	10–20	20–30	30–50
No. of families	1	7	14	16	12

   Find the mode for house rent using the graphical method.

*

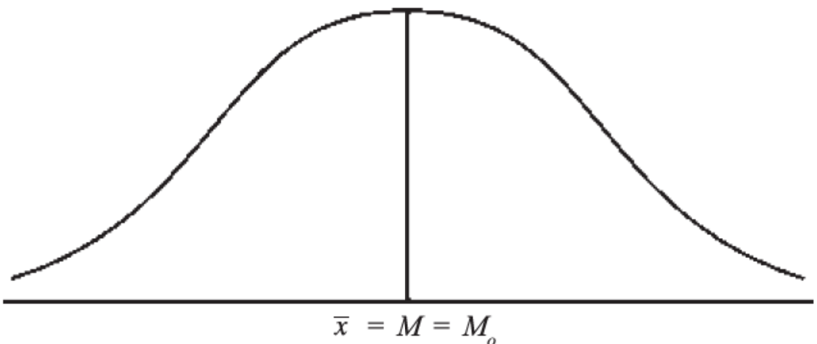
Some results for the measures of central tendency :

(1) If all the observations in the given data have same value, all the measures of central tendency have the same value.

For example, if the ages (in years) of 5 students selected from a class are 15, 15, 15, 15, 15 years, all the measures of central tendency have the value 15.

Thus, $\bar{x} = M = M_o = G = \bar{x}_w = 15$.

(2) Mean, median and mode have the same value if the data are evenly distributed around average.



(3) If the variable is multiplied by a non-zero constant b and a constant a is added to it, we get a variable $y = bx + a$

We saw in case of mean that using mean $\overline{x}$ for x, we can find the mean for the variable y as $\overline{y} = b\overline{x} + a$.

Similarly, if we have the median or mode of x, we can find the median or mode of the variable y respectively.

Median of $y = b$ (median x) $+ a$

Mode of $y = b$ (mode of x) $+ a$

Illustration 39 :

(1) The mean of a variable x is 25. Find the mean of the variable obtained by first subtracting 3 from x and then dividing it by 2.

(2) The relation between price (p) of an item and its demand (d) is $d = 50 - 2p$. If the median of price is ₹ 11, find the mean of demand.

(3) The mode of salaries of employees of a company is ₹ 8500. The company has decided to deduct 2% of each employee's salary for welfare fund. Find the mode for the amount of this fund.

(1) Here $y = \frac{x-3}{2}$. As $\overline{x} = 25$,

$$\overline{y} = \frac{\overline{x}-3}{2}$$

$$= \frac{25-3}{2}$$

$$= \frac{22}{2} = 11$$

Thus, the mean of y is 11.

(2) $d = 50 - 2p$ and median of price (p) is 11.

$\therefore$ median of demand (d) $= 50 - 2$ (median of p)

$= 50 - 2\ (11)$

$= 50 - 22$

$= 28$

Thus, the median of demand is 28 items.

(3) The mode of salaries (x) of employees is ₹8500

Amount of welfare fund (y) $= 2\ \% \times x$

$= 0.02\ x$

$\therefore$ mode for the amount of welfare fund (y) $= 0.02$ (mode of x)

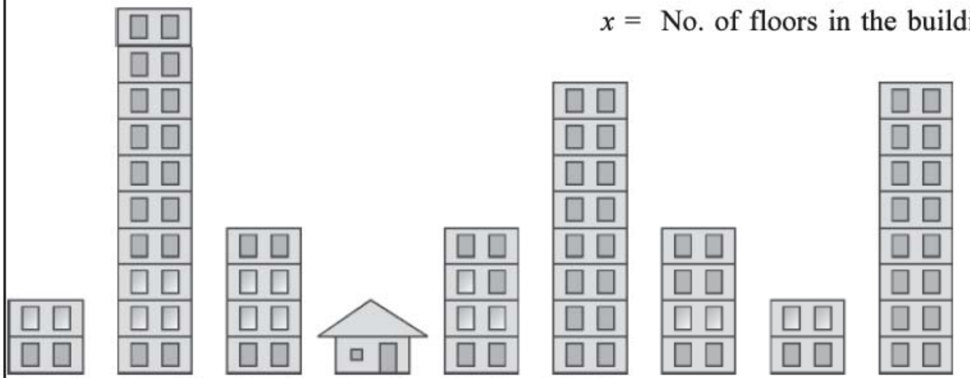
$= 0.02 \times 8500$

$= 170$

Thus, the mode for the amount deducted for welfare fund is ₹170.

Study the following pictures for a general comparison of mean, median and mode :

x = No. of floors in the building



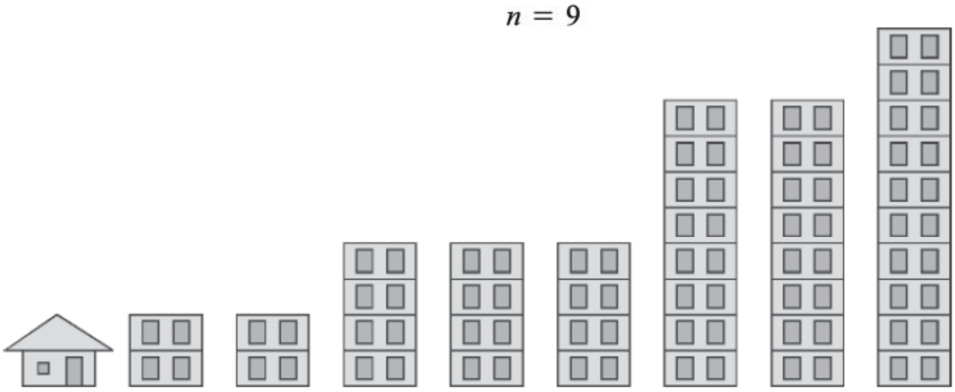
$x = 2$ $x = 10$ $x = 4$ $x = 1$ $x = 4$ $x = 8$ $x = 4$ $x = 2$ $x = 8$

There is no need to arrange the observations in any order to find mean.

$$\bar{x} = \frac{2+10+4+1+4+8+4+2+8}{9} = \frac{43}{9} = 4.78$$

x = No. of floors in the building

$n = 9$



$x = 1$ $x = 2$ $x = 2$ $x = 4$ $x = 4$ $x = 4$ $x = 8$ $x = 8$ $x = 10$

The observations are arranged in the ascending order of their values to find median.

M = value of the $\left(\frac{n-1}{2}\right)$ th observation = value of the 5th observation = 4

x = No. of floors in the building

$x = 1$, $f = 1$; $x = 10$, $f = 1$; $x = 2$, $x = 2$, $f = 2$; $x = 8$, $x = 8$, $f = 2$; $x = 4$, $x = 4$, $x = 4$, $f = 3$

The observations having same values are arranged together to find mode.

M_o = value having maximum frequency = 4

3.6 Comparative study of mean, median and mode

We have discussed advantages and disadvantages of various measures of central tendency. It is obvious from them that any particular average cannot be suitable for all types of practical problems. Each average has some specific applications and also has certain limitations.

Among all the averages, mean satisfies most of the requisites of a good average and hence it is used in most situations of data analysis. The most important property of mean is its compatibility for further algebraic computations. The advanced statistical methods applied for studying various characteristics of a population or for comparing two populations use mean as a representative value for the variable under consideration. These points make mean an optimum measure of central tendency.

However, mean cannot truly represent the entire set of data when the data are not evenly distributed around average. Many variables studied in agriculture, social sciences and in business activities are not found to be evenly distributed. Median is a better measure of central tendency in these situations. Median is used as an average in qualitative data like education, skill and consumer satisfaction.

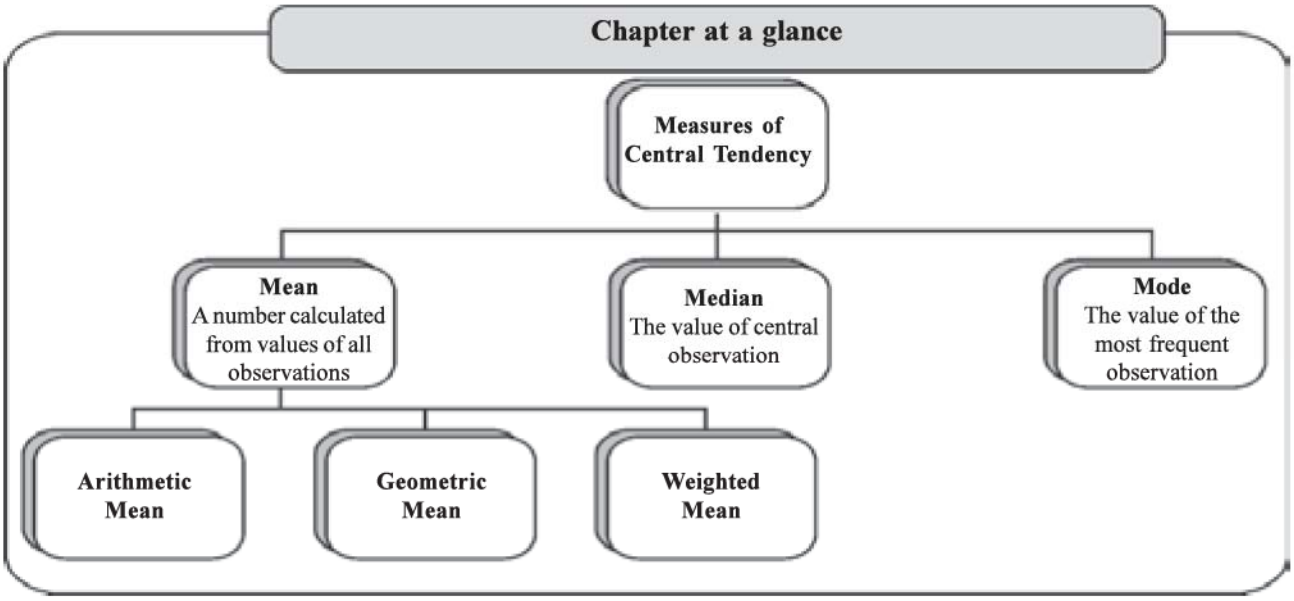
Mode is particularly used in business and commerce fields. For qualitative data also, mode is found to be useful as an average. While deciding the dishes served at a restaurant and their flavour, the choice and taste of maximum customers is taken into consideration which is an example of Mode. Mode is extensively used for finding the average by readymade garment manufacturers and in foot wear industry.

Thus, the selection of average depends upon the following factors :

(1) The nature of data (2) The nature of variable involved (3) The purpose of study (4) The type of classification used (5) The use of average for further statistical analysis

Summary

- The observations of any variable are concentrated around a certain value which is called as a measure of central tendency or average.
- Mean is the most popular average.
- Quartiles, Deciles, Percentiles are called as positional averages.
- Arithmetic mean is the sum of observations divided by the number of observations.
- If arithmetic means of two or more groups of observations are known, the combined mean can be found for the entire data.
- Weighted mean is found by assigning weights to the observations proportional to their importance.
- Geometric mean is the n th root of the product of n positive observations.
- Median is the middle most observation in the ordered data.
- Quartiles, deciles and percentiles divide the data in 4, 10 and 100 parts respectively.
- Mode is the most frequent observation for the given data.
- $M_o = 3M - 2\overline{x}$ is called as the empirical relation between mean, median and mode.



List of formulae :

(1) **Mean :**

Type of data		Short cut method
Raw data	$\bar{x} = \frac{\Sigma x}{n}$	$\bar{x} = A + \frac{\Sigma d}{n}$
Grouped data	$\bar{x} = \frac{\Sigma fx}{n}$	$\bar{x} = A + \frac{\Sigma fd}{n} \times c$

(2) **Combined mean :** $\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k}{n_1 + n_2 + \dots + n_k}$

(3) **Weighted mean :** $\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\Sigma wx}{\Sigma w}$

(4) **Geometric Mean :** $G = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$

Median and other positional averages :

	Positional average	Raw data and Discrete frequency distribution	Continuous frequency distribution
(5)	Median	M = value of the $\left(\frac{n+1}{2}\right)$ th observation	$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$
(6)	j th Quartile	Q_j = value of the $j\left(\frac{n+1}{4}\right)$ th observation	$Q_j = L + \frac{j\left(\frac{n}{4}\right) - cf}{f} \times c$
(7)	j th Decile	D_j = value of the $j\left(\frac{n+1}{10}\right)$ th observation	$D_j = L + \frac{j\left(\frac{n}{10}\right) - cf}{f} \times c$
(8)	j th Percentile	P_j = value of the $j\left(\frac{n+1}{100}\right)$ th observation	$P_j = L + \frac{j\left(\frac{n}{100}\right) - cf}{f} \times c$

9. Mode :

Raw data	Discrete frequency distribution	Continuous frequency distribution
M_o = value of observation which gets repeated Maximum number of times	M_o = value of observation with maximum frequency	$M_o = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$

10. Empirical formula : $M_o = 3M - 2\overline{x}$

EXERCISE 3

Section A

Find the correct option for the following multiple choice questions :

1. Which average gets most affected by too large or too small values ?

(a) Arithmetic mean (b) Median (c) Mode (d) Geometric mean

2. Which of the following will give us the value of median ?

(a) D_7 (b) Q_1 (c) P_{45} (d) P_{50}

3. In which of the following situations, mean cannot be found ?

(a) class lengths are unequal, (b) there are open ended class intervals,

(c) the number of class intervals is more than 5, (d) inclusive type of classes are used

4. For any set of observations, which of the following is true ?

(a) $\overline{x} \le G$ (b) $\overline{x} = G$ (c) $\overline{x} \ge G$ (d) $\overline{x} > G$

5. Which of the following results is true for the data that are evenly distributed around average ?

(a) $\overline{x} = M = M_o$ (b) $\overline{x} > M > M_o$

(c) $\overline{x} < M < M_o$ (d) $\overline{x} < M > M_o$

6. If the mean of 10 observations is 15, what is the sum of observations ?

(a) 25 (b) 150 (c) 5 (d) 1.5

7. If $\Sigma(x-9)=0$ for data having 5 observations, then what is the value of mean ?

(a) $\overline{x}=0$ (b) $\overline{x}=5$ (c) $\overline{x}=9$ (d) $\overline{x}=45$

8. What is the mode of observations 7, 9, 9, 1, 7, 9, 4, 9, 1 ?

(a) 1 (b) 4 (c) 7 (d) 9

9. In a set of 50 observations, what is the median ?

(a) value of 25th observation (b) value of 26th observation

(c) value of 25.5th observation (d) value of 26.5th observation

10. What is the geometric mean of 4 and 9 ?

(a) 4 (b) 6 (c) 6.5 (d) 36

11. If mean for a variable is 15 and its median is 20, what is the mode using empirical formula ?

(a) 30 (b) 5 (c) 35 (d) 17.5

12. The median of 10 observations is 14. What will be the median of observations obtained when each observation gets doubled ?

(a) 10 (b) 28 (c) 7 (d) 1.4

13. All the observations in a data are of same value 16. What will be their mode ?

(a) 8 (b) 2 (c) 16 (d) 4

14. Which of the following statements is false ?

(a) The quartiles divide the data in 4 equal parts.

(b) The mean divides the data in 2 equal parts.

(c) The percentiles divide the data in 100 equal parts.

(d) The deciles divide the data in 10 equal parts.

15. The lengths (in meters) of 6 pipes manufactured by a company are as follows :

1.05, 1.15, 0.98, 1.12, 0.89, 0.95

Which of the following statements is true ?

(a) Mode = 1m (b) Mode = 1.15 m (c) Mode = 0.98 m (d) Mode cannot be found

Section B

Answer the following questions in one sentence:

1. State any one advantage of mean.

2. If observations have varying importance, which average should be used ?

3. Name any two positional averages.

4. State the empirical relation between mean, median and mode.

5. State the condition under which geometric mean cannot be found.

6. Define mode.

7. State the name of the statistician who gave the empirical formula between mean, median and mode.

8. Median of 10 observations is 55. If the value of the largest observation increases from 100 to 110, find the new median.

9. Mean of variable x is 9. What is the mean of the variable $y = x + 4$?

10. Find the modal value of the variable having the following frequency distribution :

x	5	10	15	20	25
f	12	48	23	10	2

11. Arithmetic mean of two numbers is 5. If one number is 6, find the other number.

12. Find first quartile for variable with observations 15, 4, 7, 20, 2, 7, 13.

13. Which average can be obtained if the continuous frequency distribution has open ended classes ?

14. If $Q_3 = 25.75$ for a variable, then find P_{75}.

15. The median of daily demand of a vendor is 15. If he sells each item for ₹ 10, find the median of his revenue ?

Section C

Answer the following questions :

1. Define weighted mean.
2. Explain what is meant by measure of central tendency.
3. State the advantages of mode.
4. Explain the combined mean.
5. Which type of data have median as a better measure of central tendency than mean ?
6. What are the factors to be considered while choosing an appropriate average ?
7. The mean and mode of a variable are 5.5 and 6.4 respectively. Find its median.
8. Geometric mean of two numbers is 8. If one number is 4, find the other number.
9. Mean weekly production (x) of factory is 81units. Find the mean production cost if cost is given by $y = 3x + 50$
10. The median of observations $a - 5$, $a + 1$, $a + 2$, $a - 3$, a is 10. Find a.
11. The mean of marks in Mathematics of 40 students in class is 76, whereas the same for the other class of 50 students is 85. Find the mean of marks in Mathematics of students in both the classes together.
12. The number of vehicles per family in families residing in a certain area are given in the following table. Find the median for the number of vehicles.

No. of vehicles	0	1	2	3	4	**Total**
No. of families	2	4	9	7	3	25

13. Find the weighted mean of variable x from the following data.

Variable x	1500	800	200
Weight w	5	4	1

Section D

Answer the following questions :

1. State the characteristics of an ideal average.
2. Define geometric mean and state its advantages.
3. Explain the use of mode as a measure of central tendency.
4. Explain the positional averages briefly.
5. Compare mean and median as the measures of central tendency.
6. Which average is called as optimum average ? why ?
7. The economy growth rates of a state for four consecutive years are 2%, 2.5%, 3%, 4% respectively. Find the average growth rate using an appropriate average.
8. Find D_7 and P_{15} from the following data about daily sales of a mobile phone shop and interpret them.

No. of phones	4	6	7	8	10	12
No. of days	3	9	15	23	8	2

9. The mean perfume content in the bottles filled by a perfume manufacturer's machine should be between 29.6 ml and 30.4 ml. The 7 bottles tested had the following perfume contents (in ml):

30.2, 28.9, 29.2, 30.1, 29.4, 31.3, 31.4

Is the machine working properly ?

10. The mean of marks scored by 34 boys in a class is 57. The mean of marks of all 60 students of the class is 59. Find the mean marks scored by the girls.

11. The mean of 50 observations is 35. Later on, it was known that the value of one observation was taken as 50, which was wrong. Find the mean of remaining observations by excluding the wrong observation.

12. 3 students from a group of 18 students failed in the examination for the subject of Economics. The marks obtained by the 15 students who passed are as follows:

42, 65, 53, 75, 43, 50, 68, 57, 79, 48, 51, 61, 55, 70, 64. Find the median marks of all 18 students.

13. The mean daily sale of a company is 126.2 units. The sales on 10 days after adopting a new advertising strategy are as follows : 156, 125, 162, 153, 130, 124, 127, 142, 149, 121. Can we say that mean sale has increased by new advertising strategy ?

Section E

Solve the following :

1. The following table shows the number of units of electricity consumption of different families :

No. of units	Below 200	200 - 300	300 - 400	400 - 500	500 and above
No. of families	7	13	24	16	10

Find the median consumption.

2. The weekly profit and loss information of a vendor is available as follows. Find the modal profit.

Profit (thousand ₹)	–2 - 0	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10
No. of weeks	4	8	14	6	2	1

3. The number of bags of wheat sold in a grocer's shop each day are shown in the following table :

No. of bags	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 and above
No. of days	9	17	32	24	10	5	3

Find Q_1 and D_4 for the number of bags sold.

4. The heights of students of a college are given in the following table. Find the mean height of the students :

Height (cm)	150 - 155	155 - 160	160 - 165	165 - 170	170 - 175	175 - 180	180 - 185
No. of students	8	10	20	17	15	4	1

5. The monthly income (in thousand ₹) of 130 persons living in a certain area is as follows :

Income (thousand ₹)	Less than 4	4 - 8	8 - 12	12 - 20	20 - 28	28 - 36
No. of persons	6	14	31	35	28	16

Find the median of income.

6. The data about population (in thousands) of 70 villages in a district is given in the following table :

Population (thousands)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of villages	6	18	22	15	9

Find the mode for the population using graphical method.

7. The marks obtained by 60 students in an examination are as follows. Find the mean marks of the students.

Marks	0 - 10	10 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of students	3	8	20	16	9	4

8. A survey was conducted for 50 employees in an office regarding their computer usage time. The details are shown in the following table :

Time (hours)	5 - 5.5	5.5 - 6	6 - 6.5	6.5 - 7	7 - 7.5	7.5 - 8	8 - 8.5	8.5 - 9
No. of employees	1	3	5	11	15	9	4	2

Find the quartiles Q_1 and Q_3 for the time for the usage of computer.

Section F

Solve the following :

1. The data about marks scored by 55 students from a school are given below.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of students	4	7	11	14	9	7	3

(i) If 30% students failed in the examination, what are the passing marks ?

(ii) If top 5% students are to be selected for scholarship, find the lowest marks among them.

2. Two brands of tyres are to be compared for their mean life. The following data are available.

Life (thousand km.)	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of tyres of brand A	4	7	10	5	3	1
No. of tyres of brand B	5	8	15	9	6	2

On the basis of mean, which brand of tyres is better ?

3. The distribution of sale of cars of a company on different days is as follows. Find the mode for the number of cars sold using an appropriate formula.

No. of cars	0 - 10	10 - 15	15 - 20	24	26	28
No. of days	8	14	16	11	4	2

4. The wheat crop grown per acre by farmers in different parts of a state is given below :

Wheat crop per acre (quintals)	20 - 25	25 - 30	30 - 40	40 - 50	50 - 60
No. of farmers	12	23	45	29	7

Find the mean and median for the wheat crop per acre.

5. The distribution of age of 150 spectators in a theatre is as follows :

Age (years)	15 - 20	20 - 25	25 - 30	30 - 40	40 - 50	50 - 60	60 - 80
No. of spectators	6	13	19	52	34	18	8

Find the mode for age of spectators using graphical method.

6. A producer believes that the mode of his daily production is 70. The distribution of production from the data obtained after making some changes in the design of the produced units is as follows :

No. of units	60 - 64	65 - 69	70 - 74	75 - 79	80 - 84	85 - 89	90 - 94
No. of days	5	7	10	8	5	3	2

Is there any change in the mode of the number of items produced ?

7. The data for sales of oil tins of two companies sold in a shop are as follows, which show the sales of 40 days.

No. of oil tins		2 - 5	6 - 9	10 - 13	14 - 17	18 - 21	22 - 25
No. of days	Company X	1	3	17	9	6	4
	Company Y	5	9	20	3	2	1

If median is used to compare the sales, which company can be said to have higher sale ?

8. The distribution of age (in complete years) at the time of marriage of 50 married men is as follows :

Age (years)	21 - 23	24 - 26	27 - 29	30 - 32	33 - 35
No. of men	6	21	15	6	2

Find the mode of their age at the time of marriage using graphical method.

C. G. Khatri
(1931 - 1989)

Prof. C.G. Khatri obtained his Ph.D. degree (1960) in Statistics from the MS University of Baroda. He was a Professor and head of the department of Statistics, Gujarat University, Ahmedabad.

Dr. Khatri did original work on multivariate distribution theory, matrix algebra, especially on g-inverses, linear models, estimation of variance components and location parameters in linear models, design of experiments, characterization of distributions and optimality of certain functions of matrix arguments.

He has authored or co–authored several books and about two hundred research publications in prestigious journals.

•