Unit 2 Descriptive Statistics

2.1 Introduction

One of the important objectives of statistical analysis is to describe the characteristics of a frequency distribution by determining various numerical measures. To analyze and interpret the main characteristics of a frequency distribution, it is required to determine the numerical measures central value dispersion, skewness, kurtosis, correlation etc. Averages are the representative value of the frequency distribution which give us the gist nature and characteristics of the huge mass of unwieldy numerical data.

After the data have been classified and tabulated, the next step is to analyze it. However, tabular, diagrammatic and graphical approaches are the visual illustration of the unorganized data. These techniques are not capable of describing the quantitative data in detail. Therefore, one of the most important objectives of statistical analysis is to determine various numerical measures which describe the inherent characteristics of a frequency distribution. The first of such measures is “average”. The averages are the measures which condense a huge unwieldy set of numerical data into single numerical values which are representative of the entire distribution.

2.2 Measures of Central Tendency

Averages have typical nature that all other items (values) of the distribution concentrate around the center. Averages are the values in the central part of the frequency distribution which give us an idea about concentration of the values. So, they are also referred as the "Measures of central tendency".

Definition: The single value that can represent whole statistical data is known by central value and its nature is known as measure of central tendency. It lies on the central part of data. For example, Ram is a average student it means that he gets central mark in whole class.

2.3 Various Measures of Central Tendency

The measures of central tendency commonly used in practice are as follows:

1. Mean
   a. Arithmetic Mean (A.M.)
      (i) Simple and; (ii) Weighted
   b. Geometric Mean (G.M.)
   c. Harmonic Mean (H.M.)

2. Median

3. Mode

2.4 Arithmetic Mean

The arithmetic mean is the most popular and widely used measure of central tendency. It is also called simply ‘the mean’ or ‘the average’. It is also considered as an ideal measure of central tendency or the best-known/golden measures of central tendency because it satisfies almost all requisites of ideal measure of central tendency given by Prof. Yule. Arithmetic Mean (A.M.) is the most commonly used measure among all the averages. This is due to the simplicity of its calculation and other advantages. It is used to calculate the average value of quantitative data when the distribution does not have very large and very small items. It is also used to obtain average value of distribution having closed ended class intervals and having non –extreme items.

Definition: Arithmetic mean of a given set of observations is their sum divided by the number of observations. It is denoted by $\bar{X}$ (read as "X bar"), (for sample statistics).

Population mean is denoted by μ, read as mu (for population parameter) and sample mean is denoted by $\bar{X}$.

Arithmetic mean (A.M.) is called an **ideal measure** (or **best measure or golden measure**) of central tendency sine i**t is based on all the observations.**

Uses of Arithmetic Mean

Arithmetic Mean (A.M.) is more suitable average than others while we are dealing with quantitative measures such as average bonus, average income, average sales, average profit, average production, average height, average expenditure, average revenue etc. It gives simple **quantitative information** or **numerical average**.

Note: AM can not be calculated in case of open end class like below 10, above 20. Similarly, it is not representative in highly skewed data (if very small and very large figures are given do not use AM).

2.4.1 Calculation of Arithmetic Mean

a) Individual Series

Individual series is ungrouped data where each and every value of individual item is listed singly after observations. In this ungrouped data, arithmetic mean is calculated as follows:

i) Direct Method

Let $X_1, X_2, X_3, \cdots, X_n$ be the n variate values of a random variable X. Then arithmetic mean is computed by the following formula:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\Sigma X}{n}$$

where, ΣX = the sum of observations

n = the number of observations.

ii) Short-cut Method (or Assumed Mean Method or Change of Origin Method or Deviation Method)

If the number of observation is very large and the values of observations of the given data are also large (i.e. given figure is large in digits), calculation of mean (*A.M.*) by direct method is tedious and time consuming. In this case, we take the deviations of the items from any arbitrary number for computing A.M. This method is known as assumed mean method or short-cut method or deviation method. The formula for calculating the A.M. (mean) by this method is defined by

$$\therefore \quad \bar{X} = A + \frac{\Sigma d}{n}$$

where, A = Assumed mean or arbitrary value

$d = X - A$ = Deviations of the items from the assumed mean. (Origin changed) 'A'

n = Number of observations.

Note: There is no any hard and fast rule for the selection of 'A' but it is better to take a value between highest and lowest values.

iii) Step Deviation Method or Change of Origin and Scale Method or Coding Method

For large value of observations, sometimes values are changed into smaller values by the change of origin and scale. For this, observations are multiplied or divided by a constant and this method is called step deviation method. The formula for calculating A.M.by step deviation method is given by

$$\bar{X} = A + \frac{\Sigma d'}{n} \times h$$

where, $d' = \frac{X - A}{h}$, h = Common factor (Scale change dividing by a factor)

n = Number of observations

Note: There is no any hard and fast rule for the selection of A and h but better to take the value of A between highest and lowest values and to take the value of h is common factor of the values.

Example 2.1 The following are the daily incomes of five persons in a certain locality.

Persons	*A*	*B*	*C*	*D*	*E*
Income (in Rs.)	400	350	450	500	300

Calculate the average income**.**

Solution: **Computation of the average income**

Persons	Income (in Rs) (X)
A	400
B	350
C	450
D	500
E	300
Total	$\sum X = 2000$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2000}{5} = 400$$

Hence, the average income is Rs.400.

Important Note:-

Example:

Individual Series

4,4,4 ,8,8,8,12,16,20

Direct Method:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\Sigma X}{n} = 84/9 = 9.32$$

Short-cut/Deviation/Origin Changed Method

put d = X-A, Where, A is assumed mean
Let, A = 12, Then, d = X-12

Items(X)	d= X-A = X-12
4	-8
4	-8
4	-8
8	-4
8	-4
8	-4
12	0
16	4
20	8
	$\sum d$= -24

$\bar{X} = A + \frac{\Sigma d}{n}$ = 12 + (-24)/9 = 9.32

Step Deviation/Origin and Scale Changed Method:

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

let, A=12 and h= 4 then $d' = \frac{X-12}{4}$

Marks(X)	$d' = \frac{X - 12}{4}$
4	-2
4	-2
4	-2
8	-1
8	-1
8	-1
12	0
16	1
20	2
	$\sum d'$= -6

$\bar{X} = A + \frac{\Sigma d'}{n} \times h = 12 + \frac{(-6)}{9} \times 4 = 9.32$

Example 2.2 The following table gives the monthly income of 10 employees in an office:

Employee	1	2	3	4	5	6	7	8	9	10
Income (Rs.)	4780	5760	6690	7750	4840	4920	6100	7810	7050	6950

Calculate average income

Solution:

Here, taking 7000 as the assumed mean i.e. $A = 7000$

Employee	Income (Rs.) (X)	$d = X - 7000$
1	4780	– 2220
2	5760	– 1240
3	6690	– 310
4	7750	750
5	4840	– 2160
6	4920	– 2080
7	6100	– 900
8	7810	810
9	7050	50
10	6950	– 50
Total		$\Sigma d = -7350$

Mean $(\bar{X}) = A + \frac{\Sigma d}{n} = 7000 + \frac{(-7350)}{10} = 7000 - 735 = 6265$

Hence, the average income is Rs. 6265

Example 2.3 Calculate the average wage by using step deviation method from the following data

Wage (Rs.'00'): 50, 55, 60, 65, 70, and 75

Solution: Taking Assumed mean (A) = 60 and common factor (h) = 5

Wage (Rs. '00') X	$d' = \frac{1}{5}(X - 60)$
50	– 2
55	– 1
60	0
65	1
70	2
75	3
Total	$\Sigma d' = 3$

Here, $n = 6$, $A = 60$ and $h = 5$ (common factor)

Arithmetic mean $(\bar{X}) = A + \frac{\Sigma d'}{n} \times h = 60 + \frac{3}{6} \times 5 = 60 + \frac{15}{6} = 62.5$

Hence, average income = Rs. 62.5 × 100 = Rs. 6,250.

b) Discrete Series

If the data is presented along with their corresponding frequencies, then it is called discrete series (or discrete frequency distribution).

i) Direct method:

Let $X_1, X_2, X_3, \cdots, X_n$ be the variate values of a random variable X and $f_1, f_2, f_3, \cdots, f_n$ be their respective frequencies. Then in discrete series, $A.M.$ $(\bar{X})$ is given by

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + f_3X_3 + \cdots + f_n X_n}{f_1 + f_2 + \cdots + f_n}$$

$$\therefore \quad \bar{X} = \frac{\Sigma fX}{N} \quad \text{where} \quad N = \Sigma f = \text{Total frequency}$$

ii) Short-Cut Method (or Assumed Mean or Coding or Change of Origin Method or deviation method)

The formula for calculating arithmetic mean $(\bar{X})$ using this method is given by

$$\bar{X} = A + \frac{\Sigma fd}{N}$$

where, A = Assumed mean, $N = \Sigma f$ = Total frequency

$d = X - A$ = Deviation of the items from the assumed mean 'A'

iii) Step-Deviation Method (or Coding Method or Change of Origin and Scale Method)

Generally, coding refers to the transformation of data by adding (or subtracting) or multiplying (or dividing) a constant. The addition or subtraction of a constant is called change of origin where as the multiplication or division by a constant is known as change of scale. The method of changing origin as well as scale is also known as step deviation method. Thus, the formula for calculating the arithmetic mean by this method is given by

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h$$

where, A = Assumed mean, $d' = \frac{X - A}{h}$

$N = \Sigma f$ = Total frequency, h = Common factor

Note: In case of individual and discrete series (ungrouped frequency distribution), the step deviation method can be used to calculate A.M. only when 'h' can be taken as common factor from all the items of the given distribution.

Note:

Example:

First make **discrete Series**

4,4,4 ,8,8,8,12,16,20

Direct Method:

Once See in individual series,

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\Sigma X}{n} = \frac{4+4+4+8+8+8+12+16+20}{9} = 84/9 = 9.32$$

Now, go to the discrete series

Marks(X)	No. of Students (f)	f .X
4	3	3× 4 =12
8	3	3× 8 =24
12	1	1×=12
16	1	1×16=16
20	1	1× 20=20
Total	N =9	$\Sigma fX = 84$

$\overline{X} = \frac{\Sigma fX}{N}$ = 84/9 = 9.32

Short-cut/Deviation/Origin Changed Method

Put, d = X-A, Where, A is assumed mean

Let, A = 12, Then, d = X-12

Marks(X)	No. of Students (f)	d = X-12	f .d
4	3	-8	3× (-8) =-24
8	3	-4	3× (-4)=-12
12	1	0	1×0=0
16	1	4	1×4=4
20	1	8	1× 8=8
Total	N =9		$\Sigma f.d = -24$

$\overline{X}$ = A+ $\frac{\Sigma fd}{N}$ = 12 + (-24)/9= 9.32

Step Deviation/Origin and Scale Changed Method:

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

let, A=12 and h= 4 then $d' = \frac{X-12}{4}$

Marks(X)	No. of Students (f)	$d' = \frac{X-12}{4}$	f . d'
4	3	-2	3× (-2) =-6
8	3	-1	3× (-1)=-3
12	1	0	1×0=0
16	1	1	1×1=1
20	1	2	1× 2=2
Total	N =9		$\Sigma f.\boldsymbol{d}' = -6$

$\overline{X} = A + \frac{\Sigma f d'}{N} \times h = 12 + \frac{(-6)}{9} \times 4 = 9.32$

First make Continuous Series

4,4,4,8,8,8,12,16,20

C.I.	No. of students(f)	Mid-Values(X)	f. X
0-5	3	2.5	7.5
5-10	3	7.5	22.5
10-15	1	12.5	12.5
15-20	1	17.5	17.5
20-25	1	22.5	22.5
Total	N=9		Σf.X= 82.5

$\overline{X} = \frac{\Sigma f X}{N}$ = 82.5/9 = 9.16

Short-cut/Deviation/Origin Changed Method

d= X-A = X- 12.5

C.I.	No. of students(f)	Mid-Valus(X)	d=X-12.5	f. d
0-5	3	2.5	-10	-30
5-10	3	7.5	-5	-15
10-15	1	12.5	0	0
15-20	1	17.5	5	5
20-25	1	22.5	10	10
Total	N=9			Σf.d= -30

$\overline{X} = A + \frac{\Sigma f d}{N} = 12.5 + \frac{(-30)}{9} = 9.16$

Step Deviation/Origin and Scale Changed Method

$\boldsymbol{d'} = \frac{\boldsymbol{X-A}}{\boldsymbol{h}} = \frac{\boldsymbol{X-12.5}}{\boldsymbol{5}}$

C.I.	**No. of students(f)**	**Mid-Valus(X)**	$\boldsymbol{d'} = \frac{\boldsymbol{X-12.5}}{\boldsymbol{5}}$	**f.** $\boldsymbol{d'}$
0-5	**3**	**2.5**	**-2**	**-6**
5-10	**3**	**7.5**	**-1**	**-3**
10-15	**1**	**12.5**	**0**	**0**
15-20	**1**	**17.5**	**1**	**1**
20-25	**1**	**22.5**	**2**	**2**
Total	**N=9**			**Σf.** $\boldsymbol{d'}$ **= -6**

$\overline{X} = \text{A} + \frac{\Sigma f d'}{N} \times \text{h} = 12 + \frac{(-6)}{9} \times 5 = 9.16$

Example 2.4 Calculate arithmetic mean for the following frequency distribution.

X	5	10	15	20	25
f	2	4	7	20	25

***Solution*:** Calculation of mean using direct method

X	f	fX
5	2	10
10	4	40
15	7	105
20	3	60
25	1	25
	$N = 17$	$\Sigma fx = 240$

Arithmetic mean, $\overline{X} = \frac{\sum fX}{N} = \frac{240}{17} = 14.12$

Example 2.5 Determine the average wage of 75 employees in a company by changing the origin of data (i.e. short-cut method) from the data given below**:**

Wages (Rs.'00')	10	20	30	40	50	60	70
No. of employees	20	15	12	10	4	8	6

Solution: Calculation of average wage using short-cut method, $A = 40$

Wages (Rs. '00')	$d = X - 40$	f	fd
10	– 30	20	– 600
20	– 20	15	– 300
30	– 10	12	– 120
40	0	10	0
50	10	4	40
60	20	8	160
70	30	6	180
		$N = 75$	$\Sigma fd = -640$

$$\text{Average wage } (\bar{X}) = A + \frac{\Sigma fd}{N} = 40 + \frac{(-640)}{75} = 40 - 8.53 = 31.47$$

Hence, average wage $= 31.47 \times$ Rs.100 = Rs. 3,147

Example 2.6 Following are the marks secured by 40 students. Calculate average marks changing origin and scale (Step-deviation method).

Marks (X)	10	20	30	40	50	60	70	80	90
No. of students	2	4	6	10	7	5	3	2	1

***Solution*:** We have, A = Assumed mean = 50

h = Common factor = 10

Calculation of average marks using step-deviation method

Marks (x)	Frequency (f)	$d' = \frac{X - 50}{10}$	fd'
10	2	– 4	– 8
20	4	– 3	– 12
30	6	– 2	– 12
40	10	– 1	– 10
50	7	0	0
60	5	1	5
70	3	2	6
80	2	3	6
90	1	4	4
	$N = 40$		$\Sigma fd' = -21$

Now, Arithmetic mean $(\bar{X}) = A + \frac{\Sigma fd'}{N} \times h = 50 + \frac{(-21)}{40} \times 10 = 50 - \frac{210}{40} = 44.75$

Hence, average marks = 44.75

Continuous Series (Grouped Data)

When the observations are classified using some short-range values along with class frequency, then it is said to be grouped data or continuous series or continuous frequency distribution. In this case, the midpoint of the class interval is considered as average value of the lower limit and upper limit of each class interval.

The formula for calculating arithmetic mean (A.M.) in continuous series is same as discrete series but in continuous series X is the mid value of class intervals.

a) Direct Method

Let X denotes the mid value of the class intervals and f is their corresponding frequencies. Then

$$\text{A.M. } (\bar{X}) = \frac{\Sigma fX}{N}$$

where, X is the mid-point of the class interval, N = Total frequency

b) Short-cut method or assumed mean method or change of origin method or deviation method

The formula for calculating the arithmetic mean ($A.M.$) by this method is given by

$$\bar{X} = A + \frac{\Sigma fd}{N}$$

Where, A = Assumed mean, X = Mid-value of the class intervals, $N = \Sigma f$ = Total frequency

d = $X - A$ = Deviation of the items from the assumed mean 'A'.

c) Step-deviation Method or Change of Origin & Scale Method or Coding Method

The formula for calculating $A.M.$ by this method is given by

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h$$

where, $d' = \frac{X - A}{h}$, A = Assumed mean, X = Mid points of class intervals, $N = \Sigma f$ = Total frequency

h = Class size or class width or common factor

Example 2.7 **Direct Method**

Compute the A.M. for the following data by using direct method.

Weights (kg)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	4	6	8	5	3

Solution: **Computation of A.M.**

Weights (kg)	No. of students (f)	Mid value (x)	fx
0 – 10	4	5	20
10 – 20	6	15	90
20 – 30	8	25	200
30 – 40	5	35	175
40 – 50	3	45	135
	$N = \Sigma f = 26$		$\Sigma fx = 620$

$$\text{Mean } (\bar{x}) = \frac{\Sigma fx}{\Sigma f} = \frac{620}{26} = 23.85$$

Example 2.8 **Short cut Method**

Compute the arithmetic mean for the following data

Weights (kg)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	4	6	8	5	3

Solution: **Computation of A.M.**

Weights (kg)	No. of students	Mid value (x)	$d = x - A = x - 25$	fd
0 – 10	4	5	– 20	– 80
10 – 20	6	15	– 10	– 60
20 – 30	8	25	0	0
30 – 40	5	35	10	50
40 – 50	3	45	20	60
	$\sum f = 26$			$\sum fd = -30$

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd}{N} = 25 + \left(\frac{-30}{26}\right) = 25 - 1.15 = 23.85$$

Example 2.9 **Step Deviation Method**

Calculate A.M.

Wages	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70
No. of Workers	10	13	18	21	24	28	20	11	8

Solution: Calculation of average wages using changing origin and scale of data

Let Assumed Mean (A) = 47.5, h = 5

Wages (Rs.)	Mid. value (X)	$d' = \frac{X - 47.5}{5}$	f	fd
25 – 30	27.5	– 4	10	– 40
30 – 35	32.5	– 3	13	– 39
35 – 40	37.5	– 2	18	– 36
40 – 45	42.5	– 1	21	– 21
45 – 50	47.5	0	24	0
50 – 55	52.5	1	28	28
55 – 60	57.5	2	20	40
60 – 65	62.5	3	11	33
65 – 70	67.5	4	8	32
			$N = \Sigma f = 153$	$\Sigma fd' = -3$

We have,

$$\text{Mean wage } (\bar{X}) = A + \frac{\Sigma fd'}{N} \times h = 47.5 + \frac{(-3)}{153} \times 5 = 47.5 - 0.098 = 47.40$$

Hence, mean wages is Rs.47.40.

Calculation of A.M. in Case of Open End Classes

The frequency distribution in which the lower limit of first class or upper limit of last class or both are unknown(i.e. not specified), such classes are known as open end classes. In case of open end classes, we cannot find out A.M. unless we make an assumption about the unknown limits. To estimate the lower limit of the first class and upper limit of the last class, the assumption would depend upon the class interval following the first class and preceding the last class.

Note: Better not to find A.M. in open ended classes since mid-value can not be determined unless stated. So, directly find Median.

Example 2.10 Calculate average income from the given income of 1400 workers of a factory.

Income (in Rs.)	Below 1500	1500 – 2000	2000 – 2500	2500 – 3000	3000 – 3500	3500 – 4000	4000or above
No. of Workers	200	225	275	250	180	150	120

Solution: The given distribution has the open-end classes. Therefore, for calculating average income it is necessary to estimate the lower limit of the first class and upper limit of the last class. The limit is estimated according to the width of the second class and class preceding the last class.

Calculation of Average Income

Income in Rs.	No. of workers (f)	Mid-value (x)	$d' = \frac{X - 2750}{500}$	fd'
1000 – 1500	200	1250	– 3	– 600
1500 – 2000	225	1750	– 2	– 450
2000 – 2500	275	2250	– 1	– 275
2500 – 3000	250	2750	0	0
3000 – 3500	180	3250	1	180
3500 – 4000	150	3750	2	300
4000 – 4500	120	4250	3	360
	$N = \Sigma f = 1400$			$\Sigma fd' = -485$

$$A = 2750, \quad h = 500$$

$$\therefore \qquad \bar{x} = A + \frac{\Sigma fd'}{N} \times h = 2750 - \frac{485}{1400} \times 500 = 2576.79$$

Hence, the average income is Rs. 2576.79.

Calculation of A.M. in Case of Cumulative Frequency Distribution

(Less than and more than cumulative frequency distribution)

Example 2.11 The following data shows the life time in hours of 400 tube lights. Find the mean life time.

Life time (in hrs)	Less than 300	Less than 400	Less than 500	Less than 600	Less than 700	Less than 800	Less than 900	Less than 1000	Less than 1100	Less than 1200
No. of lights	0	20	60	116	194	265	324	374	392	400

Solution: Since, the given frequency distribution is in the type of less than cumulative frequency distribution. So, it should be converted into ordinary frequency distribution.

Life time (in hrs.)	No. of tubes (f)	Mid. value (X)	$d' = \frac{X-750}{100}$	fd'
300 – 400	20 – 0 = 20	350	– 4	– 80
400 – 500	60 – 20 = 40	450	– 3	– 120
500 – 600	116 – 60 = 56	550	– 2	– 112
600 – 700	194 – 116 = 78	650	– 1	– 78
700 – 800	265 – 194 = 71	750	0	0
800 – 900	324 – 265 = 59	850	1	59
900 – 1000	374 – 324 = 50	950	2	100
1000 – 1100	392 – 374 = 18	1050	3	54
1100 – 1200	408 – 392 = 8	1150	4	32
	$N = \Sigma f = 400$			$\Sigma fd' = -145$

Here, $A = 750$, $h = 100$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times h = 750 + \frac{-145}{400} \times 100 = 750 - 36.25 = 713.75$$

∴ The mean life time is 713.75 hrs.

Example 2.12 The following table represents the marks of 100 students.

Marks (More than)	20	30	40	50	60	70	80	90
No. of students	100	95	80	60	55	50	30	15

Find the mean marks of all 100 students.

Solution: Since the given frequency distribution is in the type of more than cumulative frequency distribution, it should be converted into ordinary frequency distribution.

Marks	No. of students (f)	Mid value (X)	$d' = \frac{X-55}{10}$	fd'
20 – 30	100 – 95 = 5	25	– 3	–15
30 – 40	95 – 80 = 15	35	– 2	–30
40 – 50	80 – 60 = 20	45	– 1	–20
50 – 60	60 – 55 = 5	55	0	0
60 – 70	55 – 50 = 5	65	1	5
70 – 80	50 – 30 = 20	75	2	40
80 – 90	30 – 15 = 15	85	3	45
90 – 100	15	95	4	60
	$N = \Sigma f = 100$			$\Sigma fd' = 85$

Here, $A = 55$, $h = 10$

We have, $\bar{X} = A + \frac{fd'}{N} \times h = 55 + \frac{85}{100} \times 10 = 55 + 8.5 = 63.5$

∴ Mean marks of the students ($\bar{X}$) = 63.5

Note:

1. For A.M., it is not necessary to convert unequal class intervals into equal class intervals. But in case of unequal class intervals, h is taken as common factor from the mid-value of classes.

2. It is also not necessary to convert inclusive class interval into exclusive class intervals because mid points remain same whether inclusive or exclusive.

2.5 Weighted Arithmetic Mean

While calculating simple arithmetic mean, it is based on the assumption that all the items in the distribution are equally important and valued. But in practice, this may not be so. The relative importance of some items in a distribution is more important than others. So, in such cases, proper weight (priority/merit) is to be given to various items. That is, the weights given to each item being proportional to the importance and value of the item in the distribution. When the weights are assigned for individual items with their relative importance or priorities (or worth), then the arithmetic mean calculated with respect to their priorities is called weighted arithmetic mean..

Let, $W_1, W_2, W_3, \cdots, W_n$ be the weights assigned to the variate values $X_1, X_2, X_3, \cdots, X_n$ according to their importance respectively. Then weighted arithmetic mean, usually denoted by $\overline{X}_w$ and is given by

$$\overline{X}_w = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_n X_n}{W_1 + W_2 + W_3 + \cdots + W_n} = \frac{\Sigma WX}{\Sigma W}$$

where, W = given weight

Example 2.13 An enquiry into the budgets of middle class families in a family gave the following information.

	Food	Rent	Clothing	Fuel	Others
Expenses	10%	15%	20%	15%	25%
Index number	90	100	85	65	137

Compute weighted A.M.

Solution: We have, $\overline{X}_w = \dfrac{\Sigma WX}{\Sigma W}$

Calculation of weighted average

Group	Weights (W)	Index No. (X)	WX
Food	10	90	900
Rent	15	100	1500
Clothing	20	85	1700
Fuel	15	65	975
Others	25	137	3425
	$\Sigma W = 85$		$\Sigma WX = 8500$

$\therefore$ Weighted mean $(\overline{X}_w) = \dfrac{\Sigma WX}{\Sigma W} = \dfrac{85000}{85} = 100$

Example 2.14 The postal service handles six basis types of letters and cards. The mail is given in the table.

Type of mailing	Ounce delivered (in millions)	Price per ounce (Rs.)
Air mail	1500	0.12
First class	77000	0.15
Second class	18100	0.11
Third class	16000	0.05
Registered	1000	0.40
Certified	500	0.45

Find out the average revenue per ounce for these services.

Solution: Here, the average revenue per ounce is given by

$$\bar{X}_W = \frac{\Sigma WX}{\Sigma W}$$

where, X = Price per ounce

W = Weight considered to the ounce delivered

$\bar{X}_W$ = Weighted average price per ounce

Calculation of average revenue

Type of mailing	Ounce delivered in million (W)	Price per ounce (X)	WX
Air mail	1500	0.12	180
First class	77000	0.15	11550
Second class	181000	0.11	19910
Third class	16000	0.05	800
Registered	1000	0.40	400
Certified	500	0.45	225
	$\Sigma_W = 114100$		$\Sigma_{WX} = 33065$

$$\therefore \quad \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{33065}{114100} = 0.2898$$

Thus, the average revenue per ounce is Rs.0.2898.

Combined Mean or Mean of Combined Series

Let, $\bar{X}_1$ and $\bar{X}_2$ be the arithmetic means of two series of sizes n_1 and n_2 respectively. Then the combined mean of two series of size $(n_1 + n_2)$ is denoted by $\bar{X}_{12}$ and is given by

$$\bar{X}_{12} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2} \quad \text{and so on.}$$

For, three sets of data combined mean ($\bar{X}_{123}$) is given by

$$\bar{X}_{123} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + n_3 \bar{X}_3}{n_1 + n_2 + n_3}$$

Example 2.15 From the information given below,

i. Which factory pays larger wage bill?

ii. what is the average daily wage for the workers of two factories?

	Factory A	Factory B
No. of wage earners	$n_1 = 35$	$n_2 = 20$
Average daily wage	$\bar{x}_1$ = Rs. 200	$\bar{x}_2$ = Rs. 250

Solution: Factory *A*: $n_1 = 35$, $\bar{x}_1$ = Rs. 200, $\Sigma x_1 = ?$

So, $\bar{x}_1 = \dfrac{\Sigma x_1}{n_1} \Rightarrow \Sigma x_1 = n_1 \bar{x}_1 = 35 \times 200 =$ Rs.7,000

Factory B: $n_2 = 20$, $\bar{x}_2$ = Rs. 250, $\Sigma x_2 = ?$

So, $\bar{x}_2 = \dfrac{\Sigma x_2}{n_2} \Rightarrow \Sigma x_2 = n_2 \bar{x}_2 = 20 \times 250 =$ Rs.5,000

Hence, factory A pays larger bill.

Combined mean is $\bar{x}_{12} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \dfrac{35 \times 200 + 20 \times 250}{35 + 20} = \dfrac{12000}{55} =$ Rs. 218.18

Example 2.16 In a class of 50 students 15 have failed and their average of marks is 10. The total marks secured by the entire class were 881. Find the average marks of those who have passed.

Solution: Given,

Passed	Failed	Class
$n_1 = 50 - 15 = 35, \bar{x}_1 = ?$	$n_2 = 15,\ \bar{x}_2 = 10$	$n_1 + n_2 = 50, \Sigma x_{12} = 881,\ \bar{x}_{12} = \frac{881}{50} = 18$

Now, $\bar{x}_{12} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \Rightarrow 881 = \dfrac{35\,\bar{x}_1 + 15 \times 10}{50} \Rightarrow \bar{x}_1 = \dfrac{8780}{7} = 1254.28$

The average marks of those who have passed = 21.

Corrected Mean

The formula for the calculation of corrected mean is given by

$$\text{Corrected mean } (\bar{X}_{\text{corrected}}) = \frac{\text{Corrected } \Sigma X}{\text{Corrected } n}$$

where,

Correct ΣX = Incorrect ΣX – Incorrect items + Correct items

Incorrect ΣX = $n\bar{X} = n \times$ Incorrect mean

If any item or items is/are omitted then

Correct ΣX = Incorrect ΣX – Incorrect item,

Correct n = Total items – Number of omitted item/items.

Sometimes any item or items is/are missed, in those cases,

Correct ΣX = Incorrect ΣX + Missed item/items.

Correct n = Total n + Number of missed item/items.

Example 2.17 Mean of 100 items was 50. Later on, it was found that two items were misread as 60 and 8 instead of 192 and 66. Find the correct mean.

Solution: Here, we have,

No. of items, $n = 100$

Mean $(\bar{X}) = 50$

Misread items = 60 and 8

Correct items = 192 and 66

We have, $\bar{X} = \frac{\Sigma X}{n}$

$\Rightarrow \quad 50 = \frac{\Sigma X}{100}$

$\Rightarrow \quad \Sigma X = 5000$

$\therefore$ Incorrect $\Sigma X = 5000$

Correct $n = 100 - 1 - 1 + 1 + 1 = 100$

Correct $\sum X = 5{,}000 - 60 - 8 + 192 + 66 = 5190$

Correct $(\bar{X}) = \frac{\text{Correct } \Sigma x}{\text{Correct } n} = \frac{5190}{100} = 51.9$

$\therefore$ Correct mean $(\bar{X}) = 51.9$

Example 2.18 Arithmetic mean of 150 items is 50. Two items 40 and 60 were left out at the time of calculations. What is the correct mean of all the items?

Solution: Given, mean $(\bar{x}) = 50$, No. of items, $(n) = 150$

Left out items = 40 and 60

Required mean of all the items (Corrected $\bar{x}$) = ?

We have, $\bar{x} = \frac{\Sigma x}{n} \quad \Rightarrow \quad 50 = \frac{\Sigma x}{150} \quad \Rightarrow \quad \Sigma x = 50 \times 150 = 7500$

$\sum$x without including left items = 7500

i.e. Wrong $\Sigma x = 7500$

Correct $n = 150 + 1 + 1 = 152$

$\therefore$ Correct $\sum x = 7500 + 40 + 60 = 7600$

Now, Correct $\bar{x} = \frac{\text{Correct } \Sigma x}{\text{Correct } n} = \frac{7600}{152} = 50$

$\therefore$ Correct mean of all the items $(\bar{x}) = 50$

Example 2.19 The mean marks of 100 students were found to be 65. At the time of checking it was found that the three marks 40, 50 and 55 were incorrect. Find the correct mean if the incorrect marks are omitted (or weeded out).

***Solution*:**

Given, mean $(\bar{x}) = 65$

No. of students, $n = 100$

Incorrect marks = 40, 50 and 55

Correct mean after omitting incorrect marks $(\bar{X})$ = ?

We have, $\bar{X} = \dfrac{\Sigma x}{n}$

$\Rightarrow \quad 65 = \dfrac{\Sigma x}{100}$

$\Rightarrow \quad \Sigma x = 65 \times 100 = 6500$

Σx without omitting incorrect marks = 6500

i.e. incorrect $\Sigma x = 6500$

Correct $n = 100 - 1 - 1 - 1 = 97$

$\therefore \quad$ Correct $\Sigma x = 6500 - 40 - 50 - 55 = 6355$

$$\text{Correct } (\bar{X}) = \frac{\text{Correct } \Sigma x}{\text{Correct } n} = \frac{6355}{97} = 65.515.$$

Relation between A.M., G.M. and H.M.

Arithmetic mean (A.M.), geometric mean (G.M.) and harmonic mean (H.M.) of a series of n observations are connected by the relationship. A.M. ≥ G. M. ≥ H. M. and $(\text{G.M.})^2$ = A.M. × H.M. for unequal.

2.8 Median (M_d)

Median is a **positional average** which divides the whole distribution of data into lower 50% and upper 50%. It is also called middle value of given distribution. It is quite different from mean as the median describes the position of the variable in the distribution. Thus median is the value which divides the distribution of values (arranged in ascending order or descending order) in two equal parts.

The median divides the total number of observations into two equal parts such that 50% of the items lie above median and 50% of the items lie below the median value. Its value depends on the position occupied by a value in the frequency distribution. So, it is also called “positional average”. It is denoted by M_d.

Median is another descriptive statistical measure used for the central values. It is suitable measure of central tendency (or average) for the **qualitative characteristics** such as knowledge, intelligent, beauty, honesty, talent, good, bad, defective, etc. It is also more appropriate/suitable (computable) average (or measure of central tendency) for the open ended data distribution.

Considerations of Computing Median

There are certain conditions in which the median can be calculated. They are as follows:

a. The items (observations) should be arranged in ascending order or descending order according as the magnitude of values.

b. The frequency distribution of the data should be continuous with 'exclusive type class intervals'. For example 0 – 10, 10 – 20, 20 – 30, ···, etc.

c. For the computation of positional average median, the classes in the continuous series may be unequal and open ended.

d. Median is the only average to be used while dealing with qualitative characteristics, still arranged in order of magnitude. For example, computation of average of beauty, honesty, intelligence etc.

Calculation of Median

a. In Individual Series

The items (observations) should be arranged in **ascending order or descending order** according as the magnitude of values.

If the number of observation is odd number, then median is the middle value after arranging the data either in ascending or descending order. Again, if the number of observations is even, there will be two middle values. So, the A.M. of two middle values gives the median.

The formula for calculating the median in case of individual series is given by

$$\text{Median } (Md.) = \text{Value of } \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item}$$

Where, n = Number of observations

b. In Discrete Series:

The steps involved are:

a. Arrange the given data in ascending order of their magnitudes.

b. Obtain the less than cumulative frequency (c.f.)

c. Use the formula, position of Median is given by $\left(\frac{N+1}{2}\right)^{\text{th}}$ item where, $N = \Sigma f$ = Total frequency

d. See the value of $\left(\frac{N+1}{2}\right)^{\text{th}}$ in less than cumulative frequency column and note the value corresponding to the cumulative frequency either equal to or just greater than that of $\left(\frac{N+1}{2}\right)^{\text{th}}$

e. The corresponding value of the variable gives the median.

c. In Continuous Series

The following steps are to be used in finding the median in case of continuous series:

a. Prepare the less than cumulative frequency (c.f.) distribution.

b. Using the formula, the position of median is given by $\left(\frac{N}{2}\right)^{\text{th}}$ item where, $N = \Sigma f$ = Total frequency

c. See cumulative frequency equal to or just greater than the value of $\frac{N}{2}$ and note the corresponding class interval.

d. The corresponding class interval contains the median value and is called the median class.

Then, Median is computed by applying the following formula,

$$M_d = L + \frac{\frac{N}{2} - \text{c.}f.}{f} \times h$$

Where, N = Total frequency

$\frac{N}{2}$ = Position of the median class.

L = Lower limit of median class

f = Frequency of median class

h = Class size of median class or width of the median class.

c.f. = Less than c.f. preceding the c.f. of the median class.

Note: 1. The classes should be exclusive type to calculate the median from continuous series of the data.

2. Median can also be calculated for the distribution having unequal class interval. i.e. For calculation of Median, it is not necessary to be equal class size unless it is stated amend the data.

Example 2.27 Find the median from the set of observations: 30, 40, 25, 18, 27, 26, 35

Solution: Arranging the given data say in ascending order of magnitude:

18, 25, 26, 27, 30, 35, 40

Median (M_d)= Value of $\left(\frac{n+1}{2}\right)^{\text{th}}$ item = Value of $\left(\frac{7+1}{2}\right)^{\text{th}}$ item = Value of 4^{th} item

Since, 4^{th} item is 27 so Median (M_d) = 27

Example 2.28 Find median from the following data: 48, 30, 40, 35,27, 29,38,45

Solution: At first, arranging the given data in the ascending order: 27, 29, 30, 35, 38, 40, 45, 48

Median = Value of $\left(\frac{n+1}{2}\right)^{\text{th}}$ item = Value of $\left(\frac{8+1}{2}\right)^{\text{th}}$ item = Value of $(4.5)^{\text{th}}$ item

Since $(4.5)^{\text{th}}$ item is the average of 4^{th} and 5^{th} items, median is the average of 35 and 38. Thus,

$$\text{Median} = \frac{35+38}{2} = \frac{73}{2} = 36.5$$

In another way,

$$\text{Median} = 4^{\text{th}} \text{ item} + 0.5\,(5^{\text{th}} \text{ item} - 4^{\text{th}} \text{ item}) = 35 + 0.5(38 - 35) = 36.5$$

Example 2.29 The following data gives the daily wages of 80 workers in a firm. Calculate median wage.

Daily wages (in '00'Rs.)	2	7	8	10	15
Number of workers	20	15	12	15	18

***Solution*:** Calculation of median

Daily wages ('00' Rs.) (X)	No. of workers (f)	$c.f.$
2	20	20
7	15	35
8	12	47
10	15	62
15	18	80
	$N = 80$	

$$\text{Median} = \text{Value of } \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item } = \text{Value of } \left(\frac{80+1}{2}\right)^{\text{th}} \text{ item} = \text{Value of } 40.5^{\text{th}} \text{ item}$$

From c.f. table, the c.f. just greater than 40.5 is 47 and its corresponding value of 'X' is 8. So,

Median wage (M_d) = 8 × Rs.100 = Rs.800

Example 2.30 Find the median of the given data:

Weight (in gm)	No. of oranges
40 – 45	14
45 – 50	20
50 – 55	42
55 – 60	54
60 – 65	45
65 – 70	18
70 – 75	7

Solution: **Computation of median**

Weight in gms	No. of oranges (f)	Cumulative frequency (c.f.)
40 – 45	14	14
45 – 50	20	34
50 – 55	42	76
55 – 60	54	130
60 – 65	45	175
65 – 70	18	193
70 – 75	7	200
	$N = 200$	

We have,

$$\text{Postion of Median } = \left(\frac{N}{2}\right)^{\text{th}} \text{ item} = \left(\frac{200}{2}\right)^{\text{th}} \text{item} = 100^{\text{th}} \text{ item}$$

From c.f. table, c.f. just greater than 100 is 130 which corresponds the class (55 – 60). So Median lies in the class (55 – 60). i.e. median class = (55 – 60)

Now we have,

$$L = 55,\ f = 54,\ c.f. = 76,\ h = 5$$

Now,

$$\text{Actual value of Median } (M_d) = L + \frac{\frac{N}{2} - c.f}{f} \times h$$

$$= 55 + \left(\frac{100 - 76}{54}\right) \times 5$$

$$= 55 + 2.222 = 57.222$$

Example 2.31 The following data gives the weekly wages in rupees of 24 workers of a firm.

Wages per week (in Rs.'000')	10–14	15–19	20–24	25–29	30–34
Number of Workers	4	7	8	3	2

Compute Median wage.

***Solution*:** Since, the given class intervals are inclusive type we should first convert the given inclusive class intervals into exclusive class intervals before calculating median by using correction factor.

Correction factor (C_f) = Half of difference of lower limit of the succeeding class and upper limit of preceding class (C_f) = $\frac{15 - 14}{2} = 0.5$

Reconstruction of the given data and calculation of median,

Wages per week (in Rs.'000')	Workers (f)	$c.f.$
9.5 – 14.5	4	4
14.5 – 19.5	7	11
19.5 – 24.5	8	19
24.5 – 29.5	3	22
29.5 – 34.5	2	24
	$N = 24$	

Position of median is given by $\frac{N}{2} = \frac{24}{2} = 12$. From c.f. table, the c.f. just greater than 12 is 19, whose corresponding class is (19.5 – 24.5), so median lies in the class (19.5 – 24.5).

Where, L = Lower limit median class = 19.5

f = Frequency of median class = 8

h = Class width of median class = 5

$c.f.$ = less than $c.f.$ preceding to $c.f.$ of the median class = 11

$$M_d = L + \frac{\frac{N}{2} - c.f.}{f} \times \text{h}$$

or,
$$M_d = 19.5 + \frac{12 - 11}{8} \times 5 = 20.125$$

or, $M_d = 20.125$

∴ Median wage $= 20.125 \times \text{Rs.}1000 = \text{Rs.}20{,}125$

The Partition Values

The values which divide the total number of observations into a number of equal parts are called partition values. Thus, median may also be regarded as a particular partition value because it divides the given data into two equal parts.

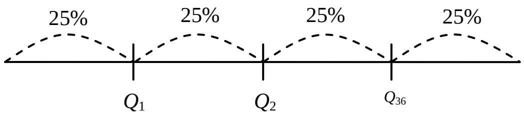
The partition values are generally used for scaling and ranking of test scores in psychological and educational statistics. It is also useful in personnel work and productivity ratings. Partition values are calculated by arranging the data in ascending or descending order of magnitude. Once the quantitative data is ranked according to their magnitude, it is possible to define various boundaries of the set.

Depending upon the equal number of parts, the important amongst these partition values which frequently used are

a. Quartiles   b. Deciles   c. Percentiles

2.8.1 Quartiles:

Quartiles are those variate values, which divide the total observations into four equal parts and each part equal to 25%. Hence, in such cases, it is obvious that there are three quartiles (points) Q_1, Q_2 and Q_3 such that $Q_1 < Q_2 < Q_3$ to divide the data into four equal parts. Q_1 is called the lower quartile (or first quartile) and Q_3 is the upper quartile (or third quartile). 25% values are below Q_1 and 25% values are above Q_3 and rest 50% values lie between Q_1 and Q_3. Since, Q_2 divides the series into two equal parts and hence it is same as median. Generally, quartiles are used in economics and business and also in determining the shape of a distribution.



Calculation of the Quartiles

a) In Individual Series:

After arranging the given data in ascending order of magnitudes, Quartiles can be obtained by the following formula:

$$\text{i}^{\text{th}} \text{ quartile } (Q_i) = \text{Value of } \frac{i\,(n+1)^{\text{th}}}{4} \text{ item.}$$

Where, $i = 1, 2, 3$

n = Number of observations

Example 2.32 Find out the lower and upper quartiles from the given data. 18, 20, 15, 16, 25, 19, 12, 22, 14

***Solution*:** Here, number of observations i.e. $n = 9$,

First the given data are arranged in ascending order.

12, 14, 15, 16, 18, 19, 20, 22, 25

$$\text{Lower quartile } (Q_1) = \text{Value of } \frac{(\text{n}+1)^{\text{th}}}{4} \text{ item} = \text{Value of } \frac{(9+1)^{\text{th}}}{4} \text{ item} = \text{Value of } 2.5^{\text{th}} \text{ item}$$

$$= \text{Value of } 2^{\text{nd}} \text{ item} + 0.5\,(3^{\text{rd}} - 2^{\text{nd}}) \text{ item} = 14 + 0.5\,(15 - 14) = 14.5$$

$\therefore\ Q_1 = 14.5$

Similarly, upper quartile (Q_3) = Value of $3\,\dfrac{(n+1)}{4}^{th}$ item

$$= \text{Value of } 3\,\frac{(9+1)}{4}^{th} \text{ item}$$

$$= \text{Value of } 7.5^{th} \text{ item}$$

$$= \text{Value of } 7^{th} \text{ item} + 0.5(8^{th} - 7^{th}) \text{ item}$$

$$= 20 + 0.5\,(22 - 20) = 21$$

$\therefore\ Q_3 = 21$

b. In Discrete Series:

$$i^{th} \text{ quartile } (Q_i) = \text{value of } \frac{(N+1)}{4}^{th} \text{ item.}$$

$$N = \Sigma f = \text{Total frequency}$$

$$i = 1, 2 \text{ and } 3$$

The steps involved are:

a. Arrange the given data in ascending order of their magnitudes.

b. Obtain the less than cumulative frequency (c.f.)

c. Use the formula, position of i^{th} quartiles = $\dfrac{i(N+1)}{4}^{th}$ item. Where, $N = \Sigma f$ = Total frequency and i = 1, 2 and 3.

d. See the value of $\dfrac{i\,(N+1)}{4}^{th}$ in less than cumulative frequency column and note the value corresponding to the cumulative frequency either equal to or just greater than that of $\dfrac{i(N+1)}{4}^{th}$.

c. The value corresponding to the *c.f.* equal to or just greater than that of the value of $\dfrac{i\,(N+1)}{4}^{th}$ is the quartile.

Example 2.33 Find first and third quartiles from the given data.

X	1	2	3	4	5	6	7
f	2	5	7	10	4	3	2

***Solution*:** Calculation of quartiles

X	f	c.f.
1	2	2
2	5	7
3	7	14
4	10	24
5	4	28
6	3	31
7	2	33
	$N = \Sigma f = 33$	

First quartile (Q_1) = Value of $\frac{(N+1)^{th}}{4}$ item = Value of $\left(\frac{33+1}{4}\right)^{th}$ item = Value of 8.5^{th} item.

The *c.f.* just greater than 8.5 is 14 and the corresponding value of X is 3

$\therefore$ First quartile (Q_1) = 3

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

= Value of $3\left(\frac{33+1}{4}\right)^{th}$ item = value of 25.5^{th} item.

The c.f. just greater than 25.5 is 28 and the corresponding value of X is 5.

$\therefore$ Third quartile (Q_3) = 5

c. In Continuous series or Grouped frequency distribution

The following steps are to be used in finding the i^{th} quartile (Q_i) in case of continuous series:

a. Prepare the less than cumulative frequency (c.f.) distribution.

b. Using the formula, the position of i^{th} quartile is given by $i\left(\frac{N}{4}\right)^{th}$ item, for i = 1, 2 and 3

c. See cumulative frequency equal to or just greater than the value of $i\left(\frac{N}{4}\right)^{th}$ and note the corresponding class interval.

d. The corresponding class interval contains the i^{th} quartile (Q_i) and is called the i^{th} quartile class. Then, i^{th} quartile (Q_i) is computed by applying the following formula,

$$Q_i = L + \frac{\frac{iN}{4} - c.f}{f} \times h, \text{ where } i = 1, 2, 3$$

Where, $\frac{iN}{4}$ = the size for i^{th} quartile's class for i = 1, 2 and 3

L = Lower limit of i^{th} quartile's class

f = Frequency of i^{th} quartile's class

h = Size of class interval of i^{th} quartile's class

c.f. = Preceding c.f. of i^{th} quartile's class.

When, $i = 1$, Q_1 = first (or lower) quartile = $L + \frac{\frac{N}{4} - c.f.}{f} \times h$

When, $i = 2$, Q_2 = second quartile (or middle value or median)

$$= L + \frac{\frac{N}{2} - c.f.}{f} \times h$$

When, $i = 3$, Q_3 = third (or upper) quartile = $L + \frac{\frac{3N}{4} - c.f.}{f} \times h$

Note: 1. The classes should be exclusive type to calculate the quartiles for the continuous series of the data.
2. i^{th} quartile (Q_i) can also be calculated for the distribution having unequal class interval. i.e. For calculation of i^{th} quartile (Q_i), it is not necessary to be equal class size unless it is stated amend the data.

Example 2.34 From the given following data, calculate the first and third quartile.

Wage in Rs.	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of workers	45	85	160	75	35

***Solution*:** Calculation of Quartiles

Wage in Rs.	No. of workers (f)	$c.f.$
0 – 10	45	45
10 – 20	85	130
20 – 30	160	290
30 – 40	75	365
40 – 50	35	400
	$N = 400$	

For Q_1:

The position of Q_1 is given by $\frac{N}{4} = \frac{400}{4} = 100$, the $c.f.$ just greater than 100 is 190.

So, Q_1 lies in the class 10 – 20.

$$L = 10,\ c.f. = 45,\ f = 85 \text{ and } h = 10$$

$$Q_1 = L + \frac{\frac{N}{4} - c.f.}{f} \times h = 10 + \frac{100 - 45}{85} \times 10 = 16.47$$

$\therefore$ First quartile (Q_1) = Rs.16.47

For Q_3 :

The position of Q_3 is given by $\frac{3N}{4} = \frac{3 \times 400}{4} = 300$, the $c.f.$ just greater than 300 is 365.

So, Q_3 lies in the class (30 – 40)

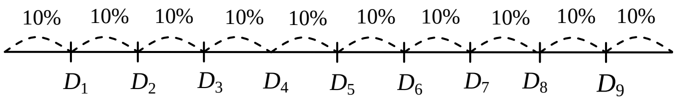
$$L = 30,\ c.f. = 290,\ f = 75 \text{ and } h = 10$$

$$Q_3 = L + \frac{\frac{3N}{4} - c.f.}{f} \times h = 30 + \frac{300 - 290}{75} \times 10 = 31.33$$

$\therefore$ Third quartile (Q_3) = Rs.31.33

2.8.2 Deciles

The variate values which divide the total number of observations into ten equal parts are called deciles and each part equal to 10%. Hence, there are in all nine deciles denoted by $D_1, D_2, \cdots, D_9$ such that $D_1 < D_2 < D_3 < \cdots < D_9$. The number of items that lie between any two consecutive deciles is 10% and the items before D_1 and D_9 are also 10%. D_5 the fifth deciles divides the series into two halves and hence it is the same as median.



Calculation of Deciles

a) In Individual series:

After arranging the given data in ascending order of magnitudes,

Deciles can be obtained by the following formula

$$D_i = \text{value of } \frac{i\,(n+1)}{10}^{\text{th}} \text{ item.}$$

Where, $i = 1, 2, 3, \cdots, 9$

n = Number of observations

Example 2.35 Find the 4th and 7th decile from the following data:

23, 34, 25, 40, 36 and 40

***Solution*:** At first, arranging the data in ascending order

23, 25, 34, 36, 40, 40

Position of 4th decile (D_4) is given by $\frac{4\,(n+1)}{10} = \frac{4(6+1)}{10} = 2.8$

$\therefore$ The value of 4th decile (D_4)

$= 2^{\text{nd}}$ item + 0.8 (3^{rd} item – 2^{nd} item)

$= 25 + 0.8\,(34 - 25)$

$= 25 + 0.72 = 25.72$

Position of 7th decile (D_7) is given by $\frac{7\,(n+1)}{10} = \frac{7(6+1)}{10} = 4.9$

$\therefore$ The value of 7th decile (D_7)

$= 4^{\text{th}}$ item + 0.8 (5^{th} item – 4^{th} item)

$= 36 + 0.8\,(40 - 36) = 36 + 3.2 = 39.2$

b) In Discrete series:

$$D_i = \text{Value of } \frac{i\,(\text{N}+1)}{10}^{\text{th}} \text{ item.}$$

$N = \Sigma f$ = Total frequency

Where, $i = 1, 2, 3, \cdots, 9$

The steps involved are:

a. Arrange the given data in ascending order of their magnitudes.

b. Obtain the less than cumulative frequency (*c.f.*).

c. Use the formula, position of i^{th} deciles = $\frac{i\,(N+1)}{10}^{\text{th}}$ item. where, $N = \Sigma f$ = Total frequency and i = 1, 2, 3, ···, 9.

d. See the value of $\frac{i\,(N+1)^{\text{th}}}{10}$ in less than cumulative frequency column and note the value corresponding to the cumulative frequency either equal to or just greater than that of $\frac{i\,(N+1)^{\text{th}}}{10}$.

e. The value corresponding to the c.f. equal to or just greater than that of the value of $\frac{i\,(N+1)^{\text{th}}}{10}$ is the decile.

Example 2.36 Compute 5$^{\text{th}}$ decile and 9$^{\text{th}}$ decile from the following data:

Income/week (in Rs.000)	1.2	1.5	1.8	2.1	2.4	2.7
Number of employee	7	11	16	10	6	2

***Solution*:** Computation of 5$^{\text{th}}$ decile and 9$^{\text{th}}$ decile

Income/week (in Rs.000)	No. of employee (*f*)	Cumulative frequency (*c.f.*)
1.2	7	7
1.5	11	18
1.8	16	34
2.1	10	44
2.4	6	50
2.7	2	52
Total	$N = 52$	

Position of 5$^{\text{th}}$ decile (D_5) is given by $\frac{5(N+1)^{\text{th}}}{10} = \frac{5(52+1)^{\text{th}}}{10} = 26.5^{\text{th}}$

From *c.f.* table, the c.f. just greater than 26.5 is 34 which corresponds the value 1.8 in income column. Thus 5$^{\text{th}}$ decile (D_5) = 1.8

$\therefore$ 5$^{\text{th}}$ decile (D_5) = 1.8 × Rs.1,000 = Rs.1,800

Position of 9$^{\text{th}}$ decile (D_9) is given by $\frac{9(N+1)^{\text{th}}}{10} = \frac{9(52+1)^{\text{th}}}{10} = 47.7^{\text{th}}$

From c.f. table, the c.f. just greater than 47.7 is 50 which corresponds to the value 2.4 in income column. Thus 9$^{\text{th}}$ decile (D_9) = 2.4

$\therefore$ 9$^{\text{th}}$ decile (D_9) = 2.4 × Rs.1,000 = Rs.2,400

c) In Continuous series or Grouped frequency distribution

Position of i$^{\text{th}}$ decile is given by $i\left(\frac{N}{10}\right)^{\text{th}}$

$$D_i = L + \frac{\frac{i\,N}{10} - c.f.}{f} \times h$$

Where, $i = 1, 2, 3, \cdots, 9$

The steps involved are:

a. Prepare the less than cumulative frequency (c.f.) distribution.

b. Using the formula, the position of i^{th} decile is given by $i\left(\frac{N}{10}\right)^{th}$ item, for $i = 1, 2, 3, \cdots, 9$

c. See cumulative frequency equal to or just greater than the value of $i\left(\frac{N}{10}\right)^{th}$ and note the corresponding class interval.

d. The corresponding class interval contains the i^{th} decile (D_i) and is called the i^{th} decile class. Then, i^{th} decile (D_i) is computed by applying the following formula:

$$D_i = L + \frac{\frac{iN}{10} - c.f.}{f} \times h \quad \text{where, } i = 1, 2, 3, \cdots, 9$$

Where, $\frac{iN}{10}$ = the size for i^{th} decile's class for $i = 1, 2, 3, \cdots, 9$

L = lower limit of i^{th} decile's class

f = frequency of i^{th} decile's class

h = size of class interval of i^{th} decile's class

$c.f.$ = preceding c.f. of i^{th} decile's class.

When, $i = 1$, D_1 = first decile $= L + \frac{\frac{N}{10} - c.f.}{f} \times h$

When, $i = 2$, D_2 = second decile $= L + \frac{\frac{N}{5} - c.f.}{f} \times h$ & so on

Example 2.37 The following frequency distribution is the marks distribution of 20 students in an examination out of 100 full marks.

Marks obtained	0 – 20	20 – 40	40 – 50	50 – 60	60 – 80	80–100
No. of students	2	5	7	3	2	1

Calculate the 2^{nd} decile and 8^{th} decile.

***Solution*:** Calculation of 2^{nd} decile and 8^{th} decile

Marks obtained	**No. of students (f)**	***c.f.***
0 – 20	2	2
20 – 40	5	7
40 – 50	7	14
50 – 60	3	17
60 – 80	2	19
80 – 100	1	20
Total	$N = 20$	

Position of 2^{nd} decile (D_2) is given by

$$= 2\left(\frac{N}{10}\right)^{th} = 2\left(\frac{20}{10}\right)^{th} = 4$$

From $c.f.$ table, the $c.f.$ just greater than 4 is 7 which corresponds to the class interval (20 – 40). Thus 2^{nd} decile lies in the class (20 – 40).

∴ Actual value of 2nd decile (D_2)

$$= i + \frac{\frac{2N}{10} - c.f.}{f} \times h = 20 + \frac{\frac{2 \times 20}{10} - 2}{5} \times 20$$

$$= 20 + \frac{2}{5} \times 20 = 20 + 8 = 28$$

Position of 8th decile (D_8) is given by $8\left(\frac{N}{10}\right)^{th} = 8\left(\frac{20}{10}\right)^{th} = 16$.

From c.f. table, the c.f. just greater than 16 is 17 which corresponds to the class interval (50 – 60). Thus 8th decile lies in the class (50 – 60).

∴ Actual value of 8th decile (D_8)

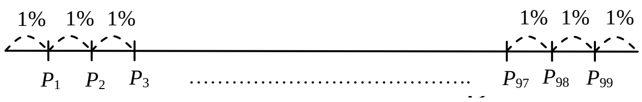
$$= L + \frac{\frac{8N}{10} - c.f.}{f} \times h = 20 + \frac{\frac{8 \times 20}{10} - 14}{3} \times 10$$

$$= 20 + \frac{2}{3} \times 10 = 50 + 6.67 = 56.67$$

∴ 2nd and 8th percentiles are 28 and 56.67.

2.8.3 Percentiles

The variate values which divide the total number of observations into 100 equal parts are called percentiles. There are in all 99 percentiles and are denoted by $P_1, P_2, \cdots, P_{99}$ respectively such that $P_1 < P_2 < P_3 < \ldots < P_{99}$. The area before p_1 and after p_{99} are 1% respectively. Also, the area between any two consecutive percentiles is 1% i.e. it represents $\frac{1}{100}$ part of the population. 50th percentile is same as median.



Calculation of Percentiles

a) In Individual Series

After arranging the given data in ascending order of magnitudes,

Percentiles can be obtained by the following formula

$$P_i = \text{value of } \frac{i\,(n+1)}{100}^{th} \text{ item.}$$

Where, $i = 1, 2, 3, \cdots, 99$

n = No. of observations

Example 2.38 Compute the 45th percentile from the following data

12, 9, 13, 10, 9, 11, 14, 15, 17, 20, 12

***Solution*:** At first, arranging the data in ascending order as below:

9, 9, 10, 11, 12, 12, 13, 14, 15, 17, 20

$$n = 11$$

Position of the 45th percentile is given by

$$\frac{45\,(n+1)^{th}}{100} = \frac{45\,(11+1)^{th}}{100} = 5.4$$

The value of 45th percentile (P_{45})

$$= 5^{th}\text{ item} + 0.4(6^{th}\text{ item} - 5^{th}\text{ item})$$

$$= 12 + 0.4\,(12 - 12) = 12$$

$\therefore$ 45th percentile (P_{45}) = 12

b) In Discrete series:

$$P_i = \text{Value of } \frac{i\,(N+1)^{th}}{100} \text{ item.}$$

$$N = \sum f = \text{Total frequency}$$

Where, $i = 1, 2, 3, \cdots, 99$

The steps involved are:

a. Arrange the given data in ascending order of their magnitudes.

b. Obtain the less than cumulative frequency (*c.f.*)

c. Use the formula, position of i^{th} percentiles = $\frac{i\,(N+1)^{th}}{100}$ item. Where, $N = \Sigma f$ = Total frequency and $i = 1, 2, 3, \cdots, 99$.

d. See the value of $\boldsymbol{\frac{i\,(N+1)^{th}}{100}}$ in less than cumulative frequency column and note the value corresponding to the cumulative frequency either equal to or just greater than that of $\frac{i\,(N+1)^{th}}{100}$.

e. The value corresponding to the *c.f.* equal to or just greater than that of the value of $\frac{i\,(N+1)^{th}}{100}$ is the percentile.

Example 2.39 Compute 15th percentile and 60th percentile from the following data:

wage/day (in Rs.'00')	12	15	18	21	24	27
Number of workers	7	11	16	10	6	2

***Solution*:** Computation of 15th percentile and 60th percentile

Wage (Rs. '00')	No. of workers (f)	Cumulative frequency (*c.f.*)
12	7	7
15	11	18
18	16	34
21	10	44
24	6	50
27	2	52

Total	$N = 52$	

Position of 15^{th} percentile (P_{15}) is given by $\frac{15\,(N+1)}{100}^{th} = \frac{15\,(52+1)}{100}^{th} = 7.95$. From c.f. table, the c.f. just greater than 7.95 is 18 which corresponds to the value 15 in wage column. Thus 15^{th} percentile (P_{15}) = 15

$\therefore$ 15^{th} percentile (P_{15}) = 15 × Rs.100 = Rs.1500

Again,

Position of 60^{th} percentile (P_{60}) is given by $\frac{60(N+1)}{100}^{th} = \frac{60\,(52+1)}{100}^{th} = 31.80^{th}$

From c.f. table, the c.f. just greater than 31.80 is 34 which corresponds to the value 18 in wage column. Thus 60^{th} percentile (P_{60}) = 18

$\therefore$ 60^{th} percentile (P_{60}) = 18 × Rs.100 = Rs.1,800

c) In Continuous series or Grouped frequency distribution

$$P_i = L + \frac{\frac{iN}{100} - c.f.}{f} \times h$$

where, $i = 1, 2, 3, \cdots, 99$

The steps involved are:

i. Arrange the given data in ascending order of their magnitudes.

ii. Obtain the less than cumulative frequency (*c.f.*)

iii. Use the formula, position of i^{th} percentiles = $\frac{i(N)}{100}^{th}$ item. Where, $N = \Sigma f$ = Total frequency and i = 1, 2, 3, ···, 99.

iv. See the value of $\frac{i(N)}{100}^{th}$ in less than cumulative frequency column and note the class interval corresponding to the cumulative frequency either equal to or just greater than that of $\frac{i(N)}{100}^{th}$

The class interval corresponding to the c.f. equal to or just greater than that of the value of $\frac{i(N)}{100}^{th}$ is the percentile class.

The actual value of i^{th} percentile is given by

$$P_i = L + \frac{\frac{iN}{100} - c.f.}{f} \times h$$

Where, $\frac{iN}{100}$ = the size for i^{th} percentile's class

L = lower limit of i^{th} percentile's class

f = frequency of i^{th} percentile's class

h = size of class interval of i^{th} percentile's class

$c.f.$ = preceding c.f. of i^{th} percentile's class.

When, $i = 1$, P_1 = first percentile $= L + \dfrac{\dfrac{N}{100} - c.f.}{f} \times h$

When, $i = 2$, P_2 = second percentile $= L + \dfrac{\dfrac{N}{50} - c.f.}{f} \times h$

When, $i = 50$, $P_{50} = 50^{th}$ percentile or median

$$P_{50} = L + \frac{\frac{N}{2} - c.f.}{f} \times h \text{ and so on.}$$

Example 2.40 From the following distribution of marks of 500 students of a campus, compute 20th percentile and 75th percentile.

Marks	0 – 20	20 – 40	40 – 50	50 – 60	60 – 80	80–100
No. of students	50	100	150	90	60	50

Solution:

Marks	No. of students	c.f.
0 – 20	50	50
20 – 40	100	150
40 – 50	150	300
50 – 60	90	390
60 – 80	60	450
80 – 100	50	500
Total	N = 500	

Position of 20^{th} percentile (P_{20}) is given by $\frac{20N}{100} = \frac{20 \times 500}{100} = 100.$

The c.f. just greater than 100 is 150.

So, P_{20} lies in class 20 – 40.Then, we have

$$L = 20, \; f = 100, \; c.f. = 50, \; h = 20$$

$$P_{20} = L + \frac{\frac{20\,N}{100} - c.f.}{f} \times \mathrm{h}$$

$$P_{20} = 20 + \frac{100 - 50}{100} \times 20 = 20 + 10 = 30$$

$$\therefore \quad P_{20} = 30$$

Again,

The position of 75^{th} percentile is given by $\frac{75\,N}{100} = \frac{3 \times 500}{4} = 375$. From the c.f. table, the c.f. just greater than 375 is 390. So, P_{75} lies in class 50 – 60.

Now, $$P_{75} = L + \frac{\frac{75\,N}{100} - c.f.}{f} \times h$$

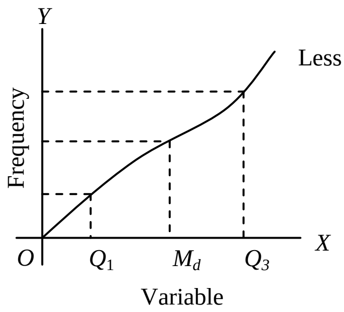
$$P_{75} = 50 + \frac{375 - 300}{90} \times 10 = 50 + 8.33 = 58.33$$

$$\therefore \quad P_{75} = 58.33$$

Note: **a.** Median = $Q_2 = D_5 = P_{50}$

b. All partition values (quartiles, deciles, percentiles) including median can be located with the help of cumulative frequency curve.

For example,



2.9 Mode

Mode is also an important measure of central tendency. Mode is the value (observation) in the series which repeats (occurs) **maximum number of times** or Mode is the value (observation) which has the **highest frequency**. Mode is the most frequently occurring value, whose repetition is maximum i.e. mode is the value, whose frequency is maximum.

There are many situations in which A.M. and median (Md.) fail to reveal the characteristics of data such as most common stock, most common wage, most common height, most common size of shoe, size of T-shirts and other ready-made garments we have to have mode in mind and not the A.M. or Median.

The word "Mode" is derived from the French word **"La Mode**" which is termed as fashionable value of the series. For example, Let us consider the following statements.

i. Average size of the sandal sold in a shopping mall is 40. In the above statement the average does not refer to mean or median. The average referred to the mode which is the most frequent value (observation) in the distribution. From the statement, we mean that there is maximum demand for the sandal of size numbered with 40. Furthermore,

ii. Average height of British is 5' 10".

iii. Average people in a dashain festival spends Rs.3,000.

Computation of Mode

a) In Individual Series

In case of individual series, the mode is the variate value/ observation that accurs maximum number of times. In other words, the most repeated value/observation/item in the data is modal value. For example,

Mode of the data 10, 13, 9, 11, 10, 20, 13, 10, 19, 18 is 10 because it repeats maximum number of times (3 times).

b) In Discrete Series

In case of discrete series, the mode is the value which corresponds to the highest frequency. For example,

Marks obtained (x)	10	20	30	40	50	60	70	80	90
No. of students (f)	3	6	9	20	25	18	7	3	1

In the above table, the highest frequency is 25 which correspond to the marks (value) 50. Therefore mode of the frequency distribution of marks of students is 50.

c) In Continuous Series

In case of continuous series, the class interval in which mode lies on, corresponding to the highest frequency can be obtained by inspection. In other words, the class interval having highest frequency in frequency column is the modal class. Then the mode can be calculated by using the formula as below:

$$\text{Mode } (M_o) = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h \text{ for } \Delta_1 + \Delta_2 \neq 0$$

Where, L = Lowest value of modal class

Δ_1 = Difference between highest frequency and its preceding frequency $(f_1 - f_0)$

Δ_2 = Difference between highest frequency and its succeeding frequency $(f_1 - f_2)$

h = Size of modal class interval/height of modal class interval

Example 2.41 Calculate the mode for the following data.

Class interval	20 – 40	40 – 60	60 – 80	80 – 100
Frequency	12	20	20	14

Solution: Since, 20 is the highest frequency, the class interval (60 –80) is the modal class.

Here,

Lowest value of modal class $(L) = 60$

Highest frequency $(f_1) = 20$

Frequency preceding to the highest frequency $(f_0) = 13$

Frequency succeeding to the highest frequency $(f_2) = 14$

and class size $(h) = 20$

Delta $\Delta_1 = (f_1 - f_0) = 20 - 13 = 7$

$\Delta_2 = (f_1 - f_2) = 20 - 14 = 6$ f1-f0+f1-f2 =2 f1-f0-f2

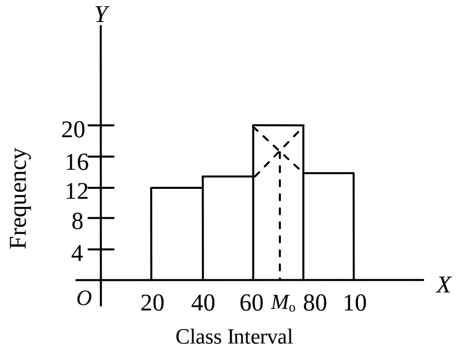
Now,

$$\text{Mode } (M_0) = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

$$= 60 + \frac{(20 - 13)}{(20 - 13) + (20 - 14)} \times 20$$

$$= 60 + \frac{7}{7 + 6} \times 20 = 60 + \frac{140}{13} = 60 + 10.77 = 70.77$$

Graphically, the measurement of mode of continuous series of data can also be computed by presenting the data in histogram as shown as below:



$\therefore$ Mode (M_o) = 70 (Approximately)

While applying the above formula to calculate the mode, the following assumptions must be considered.

i. The frequency distribution must be continuous with exclusive type class intervals without any gaps.
ii. The size (width) of all class intervals must be the same.

The above formula is not practicable in following conditions

i. The maximum frequency is repeated (not unique).
ii. The maximum frequency occurs either in the very beginning or at the end of the distribution.
iii. The difference between maximum frequency and the frequencies preceding and succeeding it is very small.
iv. There are irregularities in the distribution i.e. the frequencies of the variate values (observation) increase or decrease in a haphazard way irregular way/fluctuated.

Relationship between Mean, Median and Mode

a) For a **symmetrical (Normal/Mesokurtic)** frequency distribution, Arithmetic Mean ($\bar{X}$), Median (M_d) and Mode (M_o) are identical i.e., 2,3,4,5,6

$$\bar{X} = M_d = M_o$$

b) For moderately **skewed (asymmetrical)** frequency distribution, Mean ($\bar{X}$) Median (M_d) and Mode (M_o) are not identical i.e., $\bar{X} \neq M_d \neq M_o$. They satisfy the following relationship:

i. Mean – Mode = 3(Mean – Median)
ii. Mode = 3Median – 2Mean
iii. Which was established by Karl Pearson and the relationship is known as an empirical relationship between Mean, Mode and Median.

Computation of Mode Using Empirical Relationship between Mean, Mode and Median

If a frequency distribution is bimodal or multimodal (if the frequency distribution of any variate value have three or more than three modes) the mode is ill-defined. Such situation creates inconvenience in further statistical calculation and statistical analysis. In this case mode is computed by using empirical relationship given by Karl Pearson is

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$\text{i.e. Mode} = 3M_d - 2\bar{X}$$

Note: Types of data based on mode are as follows:

1. No modal data 2. Unimodal data 3. Bimodal data 4. Multimodal data

Some Worked Out Examples

Example 2.42 Compute the arithmetic mean of the following frequency distribution by (a) Direct method (b) Short cut method.

Salary (in Rs.'000'):	10	20	30	40	50	60
Number of officers:	2	3	9	21	11	5

Solution: a) Computation of mean using direct method:

Salary (in Rs.'000') *X*	**No. of officers (*f*)**	***fX***
10	2	20
20	3	60
30	9	270
40	21	840
50	11	550
60	5	300
Total	$N = 51$	$\sum fx = 2040$

We have,

$$\text{Arithmetic Mean } (\overline{X}) = \frac{\sum fX}{N} = \frac{2040}{51} = 40$$

$\therefore$ Arithmetic Mean $(\overline{X}) = 40 \times$ Rs.1000 = Rs.4,000

b) Computation of mean using short– cut method:

Let, assumed mean of the data $(A) = 40$

Salary (in Rs.'000') (*X*)	**Number of officers (*f*)**	$d = X - 40$	fd
10	2	– 30	– 60
20	3	– 20	– 60
30	9	– 10	– 90
40	21	0	0
50	11	10	110
60	5	20	100
Total	$N = 51$		$\sum fd = 0$

We have,

$$\text{Arithmetic Mean } (\overline{X}) = A + \frac{\sum fd}{N} = 40 + \frac{0}{51} = 40$$

Arithmetic Mean $(\overline{X}) = 40 \times$ Rs.1000 = Rs.4,000

Example 2.43 The following data are the monthly salaries (in Rs.'00') of 50 employees in a factory.

30 45 48 55 39 25 31 12 18 21 54 59 51 33 43 44 10
38 19 26 41 35 37 41 46 33 51 37 58 58 17 19 23 26
29 38 57 36 35 44 43 27 19 43 22 31 47 34 31 15

Prepare a frequency table with class intervals 10 –20, 20 – 30, as soon and compute the average salary from the frequency table obtained.

Solution:

Class interval	**Tally bars**	**Frequency (f)**	**Mid value (X)**	fX
10 – 20	~~||||~~ ||||	8	15	120
20 – 30	~~||||~~ ||||	8	25	200
30 – 40	~~||||~~ ~~||||~~ ~~||||~~	15	35	525
40 – 50	~~||||~~ ~~||||~~ |	11	45	495
50 – 60	~~||||~~ |||	8	55	440
		$\sum f = 50$		$\sum fX = 1780$

$$\text{Arithmetic Mean } (\bar{X}) = \frac{\sum fx}{\sum f} = \frac{1780}{50} = 35.6$$

$$\therefore \quad \text{Average salary} = 35.6 \times \text{Rs.}100 = \text{Rs.}3560$$

Example 2.44 Find the average marks of student from the following table:

Marks obtained	0 and less than 10	less than 20	less than 30	less than 40	less than 50
No. of Students	15	22	38	40	50

Solution: The given frequency distribution is less than cumulative frequency distribution so, at first, it is required to convert it into normal (simple) frequency distribution. Construction of normal frequency table and computation of average as below:

Marks	**No. of Students (f)**	**Mid value (X)**	fX
0 – 10	15	5	75
10 – 20	22 – 15 = 7	15	105
20 – 30	38 – 22 = 16	25	400
30 – 40	40 – 38 = 2	35	70
40 – 50	50 – 40 = 10	45	450
	$N = 50$		$\sum fX = 1100$

$$\text{Average Marks } (\bar{X}) = \frac{\sum fX}{N} = \frac{1100}{50} = 22$$

Example 2.45 Find the arithmetic mean from the following data:

Expenditure (Rs '000')	**No. of people**
Above 0	150
Above 100	130
Above 200	100
Above 300	50
Above 400	30
Above 500	12
Above 600	0

Solution: The given frequency table is more than cumulative frequency table. So, at first, we have to reconstruct the normal frequency distribution for the calculation of arithmetic mean.

Computation of arithmetic mean

Expenditure (Rs.'00000')	Number of people (f)	Mid value (X)	fX
0 – 1	20	0.5	10
1 – 2	30	1.5	45
2 – 3	50	2.5	12.5
3 – 4	20	3.5	70
4 – 5	18	4.5	81
5 – 6	12	5.5	66
6 – 7	0	6.5	0
	$N = 150$		**$\sum fX = 284.5$**

Here, Arithmetic Mean $(\bar{X}) = \dfrac{\sum fX}{N} = \dfrac{284.5}{150} = 1.89667$

Arithmetic Mean $(\bar{X}) = 1.89667 \times$ Rs.1,00,000 = Rs.1,89,666.66

Example 2.46 Goals scored by a football striker in 5 matches are 6, 4, 3, 0 and 1. What is the number of goals that the striker must score in successive 6th match in order that the average comes to 4 goals for match?

Solution: Here, the goals scored by a football striker in 5 matches are 6, 4, 3, 0 and 1.
Let, no. of goals scored by him in 6th match be x.
We have,

$$\text{Average goals } (\bar{X}) = \frac{\text{Total sum of goals}}{\text{No. of matches}}$$

$$\Rightarrow \quad 4 = \frac{6+4+3+0+1+x}{6}$$

$$\Rightarrow \quad 24 = 14 + x$$

$$\Rightarrow \quad x = 10$$

$\therefore$ The required number of goals in 6th match is 10.

Example 2.47 Find the missing frequency corresponding to the value 30 if arithmetic mean of given data is 32.

X	10	20	30	40	50	60
f	5	8	–	9	7	1

Solution: Let missing frequency corresponding to 30 be f_1.

X	f	fX
10	5	50
20	8	160
30	f_1	$30f_1$
40	9	360
50	7	350
60	1	60
Total	**$N = 30 + f_1$**	**$\sum fX = 980 + 30f_1$**

Here, $\quad \sum f = 30 + f_1$

$$\sum fx = 980 + 30f_1$$

Now, we have, $\bar{X} = \frac{\sum fX}{\sum f}$

$$\Rightarrow \quad 32 = \frac{980 + 30f_1}{30 + f_1}$$

$$\Rightarrow \quad 960 + 32f_1 = 980 + 30f_1$$

$$\Rightarrow \quad 32f_1 - 30f_1 = 980 - 960$$

$$\Rightarrow \quad 2f_1 = 20$$

$$\therefore \quad f_1 = 10$$

Therefore, the missing frequency corresponding to the value 30 is 10.

Example 2.48 Calculate the average income from the given income distribution of 1400 workers of a factory.

Income in Rs.	No. of workers
Below 1500	200
1500 – 2000	225
2000 – 2500	275
2500 – 3000	250
3000 – 3500	180
3500 – 4000	150
4000 and above	120
Total	**N = 1400**

Solution: Since only average is asked and we know that simple average represents the arithmetic mean. So, we have to calculate arithmetic mean for the given data. Since the data are given in continuous series and lower limit of the first class and upper limit of the last class are unknown. So we take 1000 and 4500 as lower limit of the first class and upper limit of the last class respectively fixing the class interval of size 500 for our convenience. Then

Taking assumed mean (A) = 2750 and Common factor (h) = 500

Income in Rs.	**Mid value (X)**	$d^1 = \frac{x - 2750}{500}$	**No. of workers (f)**	fd^1
1000 – 1500	1250	– 3	200	– 600
1500 – 2000	1750	– 2	225	– 450
2000 – 2500	2250	– 1	275	– 275
2500 – 3000	2750	0	250	0
3000 – 3500	3250	1	180	180
3500 – 4000	3750	2	150	300
4000 – 4500	4250	3	120	360
			$N = 1400$	$\sum fd^1 = -485$

We know,

$$\text{Arithmetic mean } (\bar{X}) = A + \frac{\sum fd^1}{N} \times h$$

$$= 2750 + \frac{-485}{1400} \times 500 = 2576.786$$

Hence, the required average income for given distribution is Rs. 2576.786.

Example 2.49 The average of 100 observations is 72. Later, it was discovered that two observations 85 and 63 were misread as 58 and 36. Find the correct average of 100 observations.

Solution: Here, $n = 100$, $\bar{X} = 72$

Sum of observations $(\sum x) = n \times \bar{X}$

$= 100 \times 72 = 7200$

Now, Correct $\sum X = 7200 - 58 - 36 + 85 + 63$

$= 7254$

$\therefore$ Correct average $(\bar{X}) = \dfrac{\text{Correct } \sum X}{n} = \dfrac{7254}{100} = 72.54$

Example 2.50 Calculate the average daily wages for the workers of two firms together.

	Firm *A*	**Firm *B***
No. of workers	100	200
Average daily wage (Rs)	200	150

Solution: Here,

No. of workers in Firm A (n_A) = 100
No of workers in firm B (n_B) = 200
Average daily wage in firm A $(\bar{X}_A)$ = Rs.200
Average daily wage in firm A $(\bar{X}_B)$ = Rs.150
We have,

$$\text{Combined Mean } (\bar{X}_{AB}) = \frac{n_A \cdot \bar{X}_A + n_B \cdot \bar{X}_B}{n_A + n_B}$$

$$= \frac{100 \times 200 + 200 \times 150}{100 + 200}$$

$$= \frac{20{,}000 + 30{,}000}{3{,}00} = \frac{50{,}000}{300} = 166.67$$

Therefore, Average daily wages for the workers of two firms is Rs.166.67.

Example 2.51 A student obtained the following marks out of 100 in S.L.C Examination. English 65, Mathematics 90, Science 75 and Accountancy 95. Find the students weighted arithmetic mean if weights 1, 2, 3 & 4 respectively allocated to the subjects.

Solution: Computation of weighted mean

Subject	Obtain Marks (X)	Weight (w)	wX
English	65	1	65
Mathmatics	90	2	180
Science	75	3	225
Accountancy	95	4	380
		$\sum w = 10$	$\sum wX = 850$

We have,

Weighted arithmetic mean $(\bar{X}_w) = \dfrac{\sum wX}{\sum w} = \dfrac{850}{10} = 85$

b) The marks obtained by two candidates A and B for a scholarship test is given below where as the weights of various subjects were different.

Subjects	Weights	Marks obtained by	
		A	*B*
Statistics	4	63	60
Accountancy	2	65	64
English	2	60	66
G.K.	2	60	50

If the performance of the candidates is the criteria for awarding scholarship, who should be awarded by the scholarship?

Solutions: Computation of Weighted Arithmetic Means

Subjects	Candidate *A*			Candidate *B*		
	Marks (X)	Weight (w)	wX	Marks (X)	Weight (w)	wX
Statistics	63	4	252	60	4	240
Accountancy	65	2	130	64	2	128
English	60	2	120	66	2	132
G.K.	60	2	120	50	2	100
Total		$\sum w = 10$	$\sum wX = 622$		$\sum w$=10	$\sum wX = 600$

For candidate *A*,

Average score of candidate *A*, $\bar{X}_w(A) = \frac{\sum wX}{\sum w} = \frac{622}{10} = 62.2$

Average score of candidate *B*, $\bar{X}_w(B) = \frac{\sum wX}{\sum w} = \frac{600}{10} = 60$

Since $\bar{X}_w(A) > \bar{X}_w(B)$, it indicates that *A*'s performance is better than *B*'s performance in the test. So, candidate *A* should be awarded with the scholarship.

Example 2.52 The annual profit rate of a firm in 4 different years is recorded as 5%, 5.5%, 6% and 6.5% respectively. What is the average growth rate of profit of the firm per annum?

Solution: Let x (%) be the annual profit rate of the firm. The appropriate average is geometric mean. So, Calculation of geometric mean as below:

x (%)	$\log X$
5	0.699
5.5	0.740
6	0.778
6.5	0.813
Total	$\sum \log x = 3.03$

We have, $\quad$ G.M. $= \text{Antilog}\left(\frac{\sum \log X}{n}\right)$

$= \text{Antilog}\left(\frac{3.03}{4}\right)$

$= \text{Antilog}\ (0.7575) = 5.72$

$\therefore$ The average growth rate of profit of the firm is 5.72%

Example 2.53 Find the geometric mean from the following data.

X	2	4	6	8	10
f	10	15	12	8	9

Solution: Computation of Geometric Mean

X	f	$\log X$	$f.\log X$
2	10	0.301	3.01
4	15	0.602	9.03
6	12	0.778	9.336
8	8	0.903	7.224
10	9	1.00	9.00
	$\sum f = 54$		$\sum f.\log X = 37.6$

$$\text{Geometric Mean (G.M.)} = \text{Antilog}\left(\frac{\sum f \cdot \log X}{\sum f}\right) = \text{Antilog}\left(\frac{37.6}{54}\right)$$

$$= \text{Antilog}\ (0.696) = 4.966$$

Example 2.54: From the following table, compute *A.M.* and *G.M.*, hence verify that *A.M.* > *G.M.*

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	2	4	6	3	1

Solution: Calculation of *A.M.* and *G.M.*

C.I.	f	Mid value (X)	fX	$\log X$	$f \cdot \log X$
0 – 10	2	5	10	0.699	1.398
10 – 20	4	15	60	1.176	4.704
20 – 30	6	25	150	1.398	8.388
30 – 40	3	35	105	1.544	4.632
40 – 50	1	45	45	1.653	1.653
	$\sum f = 16$		$\sum fX = 370$		$\sum f.\log X = 20.775$

$$\text{Arithmetic Mean (A.M.)} = \frac{\sum fX}{\sum f} = \frac{370}{16} = 23.125$$

$$\text{Geometric Mean (G.M.)} = \text{Antilog}\left(\frac{\sum f \log X}{\sum f}\right)$$

$$= \text{Antilog}\left(\frac{20.775}{16}\right)$$

$$= \text{Antilog}\ (1.298) = 19.86$$

Hence, *A.M.* > *G.M.*

Example 2.55: Calculate the harmonic mean from the following data.

X	10	20	30	40	50
f	7	8	6	5	7

Solution: Computation of harmonic mean as below:

X	f	$\frac{1}{X}$	$f \cdot \frac{1}{X}$
10	7	0.1	0.70
20	8	0.05	0.40
30	6	0.033	0.2
40	5	0.025	0.125
50	7	0.02	0.14
	$N = \sum f = 33$		$\sum f \cdot \frac{1}{X} = 1.565$

$$\text{Harmonic Mean } (H.M.) = \frac{N}{\sum f \cdot \frac{1}{X}} = \frac{33}{1.565} = 21.08$$

Example 2.56: Find Q_1, D_3 and P_{65} from the given data: 8, 6, 5, 4, 10, 15, 3,16

Solution: Here, the number of observation, i.e. $n = 8$

First, the data are arranged in ascending order: 3, 4, 5, 6, 8, 10, 15, 16.

$$\begin{aligned}
Q_1 &= \text{Value of } \left(\frac{1(n+1)}{4}\right)^{\text{th}} \text{ item} \\
&= \text{Value of } \frac{(8+1)^{\text{th}}}{4} \text{ item} \\
&= 2.25^{\text{th}} \text{ item} \\
&= 2^{\text{nd}} \text{ item } + 0.25\ (3^{\text{rd}} - 2^{\text{nd}}) \text{ item} \\
\therefore Q_1 &= 4 + 0.25\ (5 - 4) = 4.25 \\
\text{Similarly, } D_3 &= \text{Value of } \left(\frac{3(n+1)}{10}\right)^{\text{th}} \text{ item} \\
&= \text{Value of } \frac{3(8+1)^{\text{th}}}{10} \text{ item} \\
&= \text{Value of } 2.7^{\text{th}} \text{ item} \\
&= \text{Value of } 2^{\text{nd}} \text{ item } + 0.7\ (3^{\text{rd}} - 2^{\text{nd}}) \text{ item} \\
\therefore D_3 &= 4 + 0.7\ (5 - 4) = 4.7 \\
P_{65} &= \text{Value of } \left(\frac{65\ (\text{n}+1)}{100}\right)^{\text{th}} \text{ item} \\
&= \text{Value of } \frac{65(8+1)^{\text{th}}}{100} \text{ item} \\
&= \text{Value of } 5.85^{\text{th}} \text{ item} \\
&= \text{Value of } 5^{\text{th}} \text{ item } + 0.85\ (6 \text{ item} - 5^{\text{th}} \text{ item}) \\
&= 8 + 0.85\ (10 - 8) = 9.7 \\
\therefore P_{65} &= 9.7
\end{aligned}$$

Example 2.57: Find upper quartile and upper decile from the given data. Also obtain P_{77}.

X	1	2	3	4	5	6	7	8	9	10	11
f	2	5	8	10	12	8	6	4	3	2	1

Solution: Calculation of partition values

X	f	Less than $c.f.$
1	2	2
2	5	7
3	8	15
4	10	25
5	12	37
6	8	45
7	6	51
8	4	55
9	3	58
10	2	60
11	1	61
	$N = 61$	

For Q_3,

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

$$= \text{Value of } 3\left(\frac{61+1}{4}\right)^{\text{th}} \text{item}$$

$$= \text{Value of } 46.5^{\text{th}} \text{ item.}$$

The value in c.f. just greater than 46.5 is 51.

So, upper quartile i.e. $Q_3 = 7$.

Similarly, For upper decile (D_9)

$$= \text{Value of } 3\left(\frac{N+1}{10}\right)^{\text{th}} \text{ item}$$

$$= \text{Value of } 3\left(\frac{61+1}{10}\right)^{\text{th}} \text{ item}$$

$$= \text{Value of } 55.8^{\text{th}} \text{ item}$$

The value of c.f. just greater than 55.8 is 58.

So upper decile i.e. $D_9 = 9$.

For P_{77}

$$P_{77} = \text{Value of } 3\left(\frac{77+1}{100}\right)^{\text{th}} \text{ items}$$

$$= \text{Value of } \frac{3 \times 78^{\text{th}}}{100} \text{ item}$$

$$= \text{value of } 47.74^{\text{th}} \text{ item.}$$

The value of c.f. just greater than 47.74 is 51.

$$\therefore P_{77} = 7$$

Example 2.58: The given following data, calculate the number of workers getting wage between first and third quartile.

Wage in Rs.(less than)	10	20	30	40	50
Number of workers	45	85	160	75	35

Solution: Since, the given frequency distribution of income is in the less than cumulative frequency distribution form. So, it should be converted into ordinary (simple) frequency distribution.

Calculation of Quartiles

Wage in Rs.	No. of workers (f)	$c.f.$
0 – 10	45	45
10 – 20	85	130
20 – 30	160	290
30 – 40	75	365
40 – 50	35	400
	$N = 400$	

For Q_1

The position of Q_1 is given by $\frac{N}{4} = \frac{400}{4} = 100$, the $c.f.$ just greater than 100 is 130.

So, Q_1 lies in the class 10 – 20.

$\therefore L = 10, c.f. \;=\; 45, f = 85$ and $h = 10$.

$$\text{Now,} \qquad Q_1 = L + \frac{\frac{N}{4} - c.f.}{f} \times h$$

$$= 10 + \frac{100 - 45}{85} \times 100 = 10 + 6.47 = \text{Rs.}16.47$$

For Q_3

The position of Q_3 is given by $\frac{3N}{4} = \frac{3 \times 400}{4} = 300$, the *c.f.* just greater than 300 is 365.

So, Q_3 lies in the class 30 – 40.

$\therefore L = 30, c.f. \;=\; 290, f = 75$ and $h = 10$

$$\text{Now,} \qquad Q_3 = L + \frac{\frac{3N}{4} - c.f.}{f} \times h$$

$$= 30 + \frac{300 - 290}{75} \times 10 = 30 + 1.33 = \text{Rs.}31.33$$

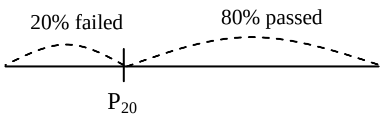
$\therefore$ The number of workers getting wage between Rs.16.47 and Rs.31.33

$$= \frac{20 - 16.67}{10} \times 85 + 16 + \frac{31.33 - 30}{10} \times 75 = 55.98$$

Example 2.60: From the following distribution of marks of 500 students of a campus, find the minimum pass mark if only 20% of the students had failed and also find the minimum marks obtained by the top 25% of the students**.**

Marks	0 – 20	20 – 40	40 – 50	50 – 60	60 – 80	80–100
No. of students	50	100	150	90	60	50

Solution: If 20% students had failed, it means 80% students had passed. The minimum pass mark obtained by the students is given by P_{20}.



$$P_{20} = D_2 = ?$$

$$P_{20} = L + \frac{\frac{20N}{100} - c.f.}{f} \times h$$

Calculation of partition values

Marks	No. of students	*c.f.*
0 – 20	50	50
20 – 40	100	150
40 – 50	150	300
50 – 60	90	390
60 – 80	60	450
80 – 100	50	500
Total	$N = 500$	

Now, the position of P_{20} is given by $\frac{20N}{100} = \frac{20 \times 500}{100} = 100$

The c.f. just greater than 100 is 150.
So, P_{20} lies in class 20 – 40. L = 20, f =100, c.f. = 50, h= 20

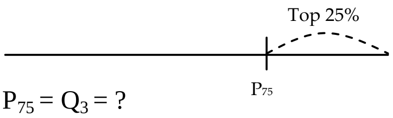
Now, $$P_{20} = L + \frac{\frac{20N}{100} - c.f}{f} \times h$$

$$P_{20} = 20 + \frac{100 - 50}{100} \times 20 = 30$$

$\therefore$ The required minimum pass mark = 30

For the top 25%:

The minimum marks obtained by top 25% of the students is given by P_{75}.



$P_{75} = Q_3 = ?$

For P_{75}:

The position of P_{75} is given by $\frac{75\,N}{100} = \frac{3 \times 500}{4} = 375.$

The c.f. just greater than 375 is 390. So, P_{75} lies in class 50–60.

Now, $$P_{75} = L + \frac{\frac{75N}{100} - c.f.}{f} \times h$$

$$P_{75} = 50 + \frac{375 - 300}{90} \times 10$$

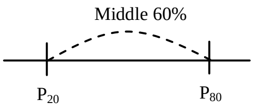
$$= 58.33 \approx 58$$

Therefore, the minimum marks obtained by the top 25% students = 58.

Example 2.61: The table given below represents the daily wage distribution of 130 workers. Find out the range of income of the middle 60% workers.

Wage (Rs. per week)	More than 70	More than 85	More than 100	More than 115	More than 130	More than 145	More than 160	More than 175
No. of workers	130	122	109	79	44	26	14	5

Solution: The limits of income of the middle 60% workers are given by P_{20} and P_{80}



Calculation of partition values P_{20} and P_{80}

Wage (Rs. per week)	f	$c.f.$
70 – 85	130 – 122 = 8	8
85 – 100	122 – 109 = 13	21
100 – 115	109 – 79 = 30	51
115 – 130	79 – 44 = 35	86
130 – 145	44 – 26 = 18	104
145 – 160	26 – 14 = 12	116
160 – 175	14 – 5 = 9	125
More than 175	5	130
	$N = 130$	

For P_{20}

The position of P_{20} is given by $\frac{20N}{100} = \frac{20 \times 130}{100} = 26$. The c.f. just greater than 26 is 51. So, P_{20} lies in class interval (100 – 115). Then, we have,

$L = 100,\ f = 30,\ c.f. = 21,\ h = 15$

$$P_{20} = L + \frac{\frac{20N}{100} - c.f.}{f} \times h$$

$$= 100 + \frac{26 - 21}{30} \times 15 = 100 + 2.50 = \text{Rs.}102.50$$

Similarly, For P_{80},

The position of P_{80} is given by $\frac{80N}{100} = \frac{80 \times 130}{100} = 104$

The c.f. just equal to 104 So, P_{80} lies in class 130 – 145

$L = 130,\ f = 18,\ c.f. = 86,\ h = 15$

$$P_{80} = L + \frac{\frac{80N}{100} - c.f.}{f} \times h$$

$$= 130 + \frac{104 - 86}{18} \times 15 = 130 + \frac{18 \times 15}{18}$$

$$= 130 + 15 = 145$$

Hence,

The range of income of the middle 60% workers = $P_{80} - P_{20}$

$= 145 - 102.50 \qquad = \text{Rs.}42.5.$

2.10 Measurement of Dispersion /Variation/Deviation/Difference of Data

Central tendency just measures the central value of the given items which is supposed to represent mass of the data. However, it may not be true if the values of the items highly scattered or spread. In other words, averages or measures of central tendency give us the idea of concentration of the items (observations) about the central part of the distribution data. But measures of central tendency do not give the information or characteristics of the distribution of how spread or scatter the data around the average value. Two or more than two distributions having same average may differ in the scatteredness or variability of the observations from the central value (averages). Thus measure of dispersion has been discussed. Measure of dispersion provides the information regarding the amount of variability or inequality or deviation or scatteredness of the data from the average (central value). The study of scatteredness or variation of the data from the averages of a distribution is called the study of 'measures of dispersion'.

The study of averages is not enough to analyze the nature of a given frequency distribution. A further analysis of the distribution is necessary if we are to know how representative the average is.

For example, the following are the marks obtained by two students in 5 different tests.

Tests:	I	II	III	IV	V	Mean	Median
Student A:	20	30	40	50	60	40	40
Student B:	0	10	40	60	90	40	40

Here we see that mean and median marks of both students A and B are same. The difference of the marks from the average marks of B is more than that of the difference of the marks from the average marks of A.

This means A's performance is more consistent than that of B. Therefore the measure of central value (average) alone does not present the whole characteristics of the distribution of data unless the variation or the dispersion of the individual values is considered. Hence, a measure of dispersion or variation is another important aspect of statistical analysis.

Thus, only the measures of central tendency are inadequate to describe (or characterize) the distribution perfectly and sufficiently. Therefore, the measures of central tendency must be supported and supplemented by some other measures. One of such measure is dispersion or variability. The definitions of the prominent statisticians about the dispersion are given below:

"*Dispersion or spread is the degree of the scatter or variation of the variables about a central value*."

-B.C. Brooks & W.F.L. Dick

"*Dispersion is a measure of the extent to which the individual items vary*." **- L.R. Connor**

"*A measure of dispersion or variation describes the degree of scatter shown by the observations and is usually measured as an average deviation about some central value or by on order statistic*."

- John I. Griffin

"*The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data.*" - **Spiegel**

Thus dispersion measure variability of statistical data from central value mean, median and mode.

The Objectives of Measures of Dispersion are as Follows:

i) To find out the reliability of an average.

ii) To control the variation of the data from central value of the distribution.

iii) To provide the comparison of variability of two or more than two distributions.

iv) To facilitate the use of other statistical measures for further analysis of data as regression and correlation analysis for further analysis of distribution.

v) To help in devicing a system of quality control.

Characteristics of Good (Ideal) Measures of Dispersion (Requisites of an Ideal/Good Measure of Dispersion)

The characteristics of good (ideal) measures of dispersion are as follows:

i) It should be rigidly defined.

ii) It should be easy to calculate and understand.

iii) It should be based on all observations.

iv) It should be amenable and suitable for further mathematical treatment.

v) It should be least affected by fluctuations of sampling.

vi) It should not be affected much by extreme values of the distribution.

2.10.1 Absolute and Relative Measures of Dispersion

Absolute Measures of Dispersion

The measures of dispersion which are dependent to original units of measurement of data are said to be absolute measures of dispersion. The absolute measures of dispersion can be used only for comparing the variability or dispersion of two or more distributions having same unit or scale. If the distributions are given in different units or scales, then for comparison, the absolute measures of dispersion cannot be used.

For example, Income x: Rs. 800

Income y: Rs. 300

$\therefore$ Absolute measure of dispersion = Rs. 800 – Rs. 300 = Rs. 500

Relative Measures of Dispersion

The measures of dispersion which are pure numbers, independent of units of measurement of the data are said to be relative measures of dispersion. Thus, the relative measures of dispersion will not have any unit of measurement and are obtained by taking a ratio or percentage of an absolute measure of dispersion to a suitable average. i.e. the ratios of two absolute values.

$\therefore$ Relative value = $\frac{\text{Rs.}800}{\text{Rs.}300} = \frac{8}{3}$

Therefore, for comparing the variability of two or more than two distributions even if they are measured in the different units and scales, for convenience results, the relative measures of dispersion instead of the absolute measures of dispersion are computed.

The example of absolute measure of dispersion and relative measure of dispersion are listed as below:

Absolute Measure	Relative Measure
1. Range 2. Quartile deviation or Semi-inter-quartile range 3. Mean deviation or Average deviation 4. Standard deviation 5. Variation	1. Coefficient of range 2. Coefficient of quartile deviation. 3. Coefficient of mean deviation. 4. Coefficient of standard deviation 5. Coefficient of variation

2.10.2 Types of Measurement of Dispersion

1. (a) Range (b) Coefficient of Range
2. Quartile Deviation (Semi-inter quartile Range)
3. Average Deviation
4. Standard Deviation
5. (a) Variance (b) Co-efficient of Variation
6. Lorenz curve

Among the above measure of dispersions range is the simplest, average (mean) deviation is better and standard deviation is the best.

Range

The simplest method of studying and measuring the dispersion is range. It is defined as the difference between extreme values of the distribution. In other words, the difference between largest (maximum) item/observation and smallest (minimum) item/observation of the distribution is range. It is generally denoted by R. Thus

Range (R) = L – S

In case of continuous series of the data, the range is obtained as the difference between the upper limit of the highest class and the lower limit of lowest class.

Range is absolute measure of dispersion. So the relative measure of dispersion corresponding to the range to compare two distributions is known as co-efficient of range which is defined by

Co-efficient of range = $\frac{L - S}{L + S}$ =500-500/500+500=0/1000= 0 =0 x100% =0%

Office-A 500, 500, 500, 500, 500 L = 500, S=500 coefficient= 0 or 0%

office-B 0, 0, 2500, 0, 0 L= 2500 S=0 Coeff= 2500-0/2500+0=2500/2500= 1 or 1x100%=100%

5000, 92000, 2000 mean= 33000

8000 10000 12000 mean = 10000

Deviation less भए - more homogeneous/equitable/stable/consistent/representative

Where, L is the largest item and S is the smallest item.

Quartile Deviation/Semi-inter-Quartile Range

Quartile deviation is the measure of dispersion based on the partition values (quartiles) of the data. The difference between third quartile (Upper quartile) and first quartile (Lower quartile) of the data is called inter-quartile range.

Inter-quartile range = $Q_3 - Q_1$. (L-S)

Quartile deviation is half of inter-quartile range and thus is called ***semi-inter-quartile range***. It is denoted by Q.D.

$\therefore$ Quartile Deviation (Q.D.) = $\frac{1}{2}(Q_3 - Q_1)$

It is also expressed as Q.D. = $\frac{1}{2}[(Q_3 - M_d) + (M_d - Q_1)]$

So, quartile deviation gives the average of deviations of the quartiles taken from median of the data. Since, in a distribution, 25% of the observations lie below Q_1 and 25% observations lie above Q_3, so 50% of the observations lie between Q_1 and Q_3. Therefore, for a symmetrical distribution, $M_d \pm$ Q.D. covers exactly 50% of the observations.

Quartile deviation is absolute measure of dispersion. So relative measure of dispersion corresponding to the Q.D. is co-efficient of Q.D. and is defined as

Co-efficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Note: Quartile deviation is the most suitable for the frequency distribution of the data having open end class intervals.

Mean Deviation

Absolute value/**modulus value**/ scalar value/magnitude/Numerical value

$|+2| = 2 \qquad |-2| = 2$

Mean of .... = $\frac{\Sigma.....}{n}$

Mean of **X** = $\frac{\Sigma X}{n}$

Mean of **deviation** = $\frac{\Sigma\, deviation}{n} = \frac{\Sigma|d|}{n}$ Absolute d or

Where, modulus d is

d = X-$\bar{X}$

d= X- Md

d=X-Mo

Eg.
2,3,4,5,6

Items(X)	f	d = X-$\bar{X}$ =X-4	$\lvert d \rvert$	f $\lvert d \rvert$
2		2-4=-2	2	
3		3-4=-1	1	
4		4-4=0	0	
5		5-4=1	1	
6		6-4=2	2	
20			$\lvert d \rvert$ =6	

$\bar{X}$=4

M.D. = $\frac{\Sigma|d|}{n}$ =6/5 =1.2 (Average deviation) 2, 3, 4, 5, 6

Coefficient of M.D.=M.D./mean=1.2/4= 0.3

Range and Quartile deviation are not the better measure of dispersion as all the items are not included in both cases and they do not show the variations of the items or observations from an average. Thus, they completely ignore the composition of the distribution. So, to overcome both these drawbacks, the other measure of dispersion is developed, which is known as mean deviation or average deviation or mean absolute deviation. Generally, it is denoted by M.D.

Mean deviation is also called average deviation. Mean deviation is defined as the average (arithmetic mean) of the positive deviations (differences) of the items taken from either of averages (Mean or Median or Mode).

Computation of Mean Deviation

Mean deviation is also called average deviation. Mean deviation from mean, median and mode are computed by using the following formula.

Series	**M.D. from Mean**	**M.D. from Median**	**M.D. from Mode**
Individual	$M.D. = \frac{\sum\lvert X - \bar{X}\rvert}{n}$	$M.D. = \frac{\sum\lvert X - M_d\rvert}{n}$	$M.D. = \frac{\sum\lvert X - M_0\rvert}{n}$
Discrete	$M.D. = \frac{\sum f\lvert X - \bar{X}\rvert}{N}$	$M.D = \frac{\sum f\lvert X - M_d\rvert}{N}$	$M.D = \frac{\sum f\lvert X - M_0\rvert}{N}$
Continuous	$M.D. = \frac{\sum f\lvert M - \bar{X}\rvert}{N}$	$M.D. = \frac{\sum f\lvert M - M_d\rvert}{N}$	$M.D = \frac{\sum f\lvert M - M_o\rvert}{N}$

Mean deviation is an absolute measure of dispersion. So relative measure of dispersion corresponding to the mean deviation is defined as follows:

$$\text{Co-efficient of M.D. from Mean} = \frac{\text{M.D. from Mean}}{\text{Mean}}$$

Co-efficient of M.D. from Median = $\dfrac{\text{M.D. from Median}}{\text{Median}}$

Co-efficient of M.D from Mode = $\dfrac{\text{M.D. from Mode}}{\text{Mode}}$

Standard Deviation

Standard deviation is defined as the positive square root of average of the squares of the deviations of the items from their arithmetic mean. It is also termed as the root mean squared deviation from mean of the data. It was first suggested by Karl Pearson in 1893. It is usually denoted by σ (small sigma) of Greek Alphabet. Since it satisfies most of the requisites of a good measure of dispersion, so it is regarded as the best measure of dispersion (or ideal measure of dispersion). It is more powerful and most widely used measure of dispersion than others.

Computation of Standard Deviation

Positive square root of mean(A.M.) of square of deviations of items taken from A.M.

Standard deviation is computed by using the following methods.

$\sqrt{\frac{\sum(X-\overline{X})^2}{n}} = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$ **Note**: d=X-A and $d' = \frac{x-A}{h}$

Series	Direct method	Short-cut method	Step Deviation Method
Individual	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n}} =$	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$	$S.\ D.(\sigma) = \sqrt{\frac{\Sigma d'^2}{n} - \left(\frac{\Sigma d'}{n}\right)^2} \times h$
Discrete	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma f\,(X - \overline{X})^2}{N}}$	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$	$S.D.(\sigma) = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h$
Continuous	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma f(M - \overline{X})^2}{N}}$	$S.D.\ (\sigma) = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$	$S.D.(\sigma) = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h$

Where $d = (X - A)$ or, d = M – A and A= Assumed Mean,

$d' = \dfrac{X - A}{h}$, h = size of class interval (or common factor/multiplier)

Standard deviation is absolute measure of dispersion. Thus, the relative measure of dispersion corresponding to standard deviation is co-efficient of standard deviation is defined as

Co-efficient of S.D. = $\dfrac{\text{S.D }(\sigma)}{\overline{X}}$

By definition of standard deviation, the above formula is derived as below:

Individual Series

$$\textbf{S.}\text{D.}\ (\sigma) = \sqrt{\frac{\sum(X - \overline{X})^2}{n}} = \sqrt{\frac{\sum X^2}{n} - (\overline{X})^2} = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

Discrete Series

$$\text{S.D.}\ (\sigma) = \sqrt{\frac{\sum f(X - \overline{X})^2}{N}} = \sqrt{\frac{\sum fX^2}{N} - (\overline{X})^2} = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2}$$

Continuous Series

$$\text{S.D.}\ (\sigma) = \sqrt{\frac{\sum f(M - \overline{X})^2}{N}} = \sqrt{\frac{\sum f M^2}{N} - (\overline{X})^2} = \sqrt{\frac{\sum f M^2}{N} - \left(\frac{\sum f M}{N}\right)^2},$$

where M = mid value of the class interval.

Example 2.63: Find standard deviation of the following series by using

(a) Actual mean method.

(b) Direct method

(c) Short cut method (or assumed mean method)

X: 4, 6, 8, 14, 18

Solution:

a) Actual mean method

X	d= (X – $\overline{X}$)= X-10	$(X - \overline{X})^2$
4	–6	36
6	–4	16
8	–2	4
14	4	16
18	8	64
$\sum X = 50$	$\sum(X - \overline{X}) = 0$	$\sum(X - \overline{X})^2 = 136$

Now, Mean ($\overline{X}$) = $\dfrac{\sum X}{n} = \dfrac{50}{5} = 10.$

Standard Deviation $(\sigma) = \sqrt{\dfrac{\sum(X - \overline{X})^2}{n}}$

$$= \sqrt{\frac{1}{5} \times 136} = \sqrt{27.2} = 5.21$$

b) Direct method

X	$\mathbf{X^2}$
4	16
6	36
8	64
14	196
18	324
$\sum X = 50$	$\sum X^2 = 636$

Standard Deviation (σ) = $\sqrt{\dfrac{\sum X^2}{n} - \left(\dfrac{\sum X}{n}\right)^2}$

$$= \sqrt{\frac{636}{5} - \left(\frac{50}{2}\right)^2}$$
$$= \sqrt{27.2} = 5.21$$

$\bar{X} = \frac{\sum X}{n} = \frac{50}{5} = 10$

Coefficient of S.D. = S.D./mean= 5.21/10 = 0.521x 100%= 52.100%

Coefficient of Variation(C.V.) = $\frac{S.D.(\sigma)}{\bar{X}} \times 100\% = 52.1\%$

Variance = Square of S.D. = σ^2

c) Short cut method

Let assumed mean (A) = 8

X	d = X – 8	d^2
4	–4	16
6	–2	4
8	0	0
14	6	36
18	10	100
	$\sum d = 10$	$\sum d^2 = 156$

Standard Deviation (σ) = $\sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

$$= \sqrt{\frac{156}{5} - \left(\frac{10}{5}\right)^2}$$
$$= \sqrt{27.2} = 5.21.$$

Example 2.64: Find the S.D. of the following data.

12, 13, 15, 16, 18

What will be the value of S.D. if each item is increased by 2? Also what will be the value of S.D. if each element is multiplied by 2?

Solution: Case I: Let assumed mean of the data be 15. i.e. A = 15

Calculation of the S.D. of the given data:

Variable (X)	d = X – 15	d^2
12	–3	9
13	–2	4
15	0	0
16	1	1
18	3	9
	$\sum d = -1$	$\sum d^2 = 23$

S.D. (σ) = $\sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

$$= \sqrt{\frac{23}{5} - \left(\frac{-1}{5}\right)^2}$$
$$= \sqrt{4.6 - 0.04}$$
$$= \sqrt{4.56} = 2.135$$

Case II: If each item is increased by 2. The original data becomes
14, 15, 17, 18, 20
Let assumed mean of the data be 17. i.e. A = 17.
Calculation of the S.D. of the given data:

Variable (X)	d = X – 17	d^2
14	–3	9
15	–2	4
17	0	0
18	1	1
20	3	9
	$\sum d = -1$	$\sum d^2 = 23$

Standard Deviation (σ) $= \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

$= \sqrt{\frac{23}{5} - \left(\frac{-1}{5}\right)^2}$

$= \sqrt{4.6 - 0.04}$

$= \sqrt{4.56}$

$= 2.135$

Case III: If each element is multiplied by 2, the original data becomes
24, 26, 30, 32, 36

Let assumed mean of the data be 30 and common factor be 2 i.e. A = 30, h = 2, Then $d' = \frac{X - A}{h}$.

Calculation of S.D. of the data:

Variable (X)	$d' = \frac{X-30}{2}$	d'^2
24	–3	9
26	–2	4
30	0	0
32	1	1
36	3	9
	$\sum d' = -1$	$\sum d'^2 = 23$

S.D.(σ) $= \sqrt{\frac{\sum d'^2}{n} - \left(\frac{\sum d'}{n}\right)^2} \times h$

$= \sqrt{\frac{23}{5} - \left(\frac{-1}{5}\right)^2} \times 2$

$= \sqrt{4.6 - 0.04} \times 2$

$= \sqrt{4.56} \times 2$

$= 2.14 \times 2$

$= 4.28$

Conclusion: SD is affected by change of scale but not by change of origin.

Example 2.65: Find S.D. of the following data.

Variable (X)	10	14	16	18	20
Frequency (f)	3	5	7	6	4

Solution: Let assumed mean of the data be 16 and common factor (multiplier) be 2. i.e. A = 16 and h = 2. Then, we have

$$d' = \frac{X - A}{h}.$$

Calculation of S.D. of the data:

Value (X)	**Frequency (f)**	**$d' = \frac{X - 16}{2}$**	**fd'**	**fd'^2**
10	3	–3	–9	27
14	5	–1	–5	25
16	7	0	0	0
18	6	1	6	6
20	4	2	8	16
	$N = \sum f = 25$		$\sum fd' = 0$	$\sum fd'^2 = 4$

$$\text{S.D. } (\sigma) = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{54}{5} - \left(\frac{0}{25}\right)^2} \times 2$$

$$= \sqrt{2.16 - 0} \times 2 = 2.939$$

Example 2.66: Find out the standard deviation from the following distribution (using step deviation method).

Wages	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
No. of workers	2	3	4	3	2

Solution: Let assumed mean be 50 and size of class interval be 20. i.e. A = 50 and h = 20, then $d' = \frac{X–A}{n}$.

Computation of standard deviation of the data:

Wages	**No. of workers (f)**	**Mid Value (X)**	**$d' = \frac{X - 50}{20}$**	**fd'**	**fd'^2**
0 - 20	10	10	–2	–20	40
20 - 40	12	30	–1	–12	12
40 - 60	15	50	0	0	0
60 - 80	8	70	1	8	8
80 - 100	5	90	2	10	20
	$N = \sum f = 50$			$\sum fd' = –14$	$\sum fd'^2 = 80$

$$\text{Standard Deviation} (\sigma) = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{80}{25} - \left(\frac{-14}{25}\right)^2} \times 20$$

$$= \sqrt{3.2 - 0.3136} \times 20$$

$$= 1.699 \times 20$$

$$= 33.98$$

Variance

The square of standard deviation is known as variance. It is denoted by σ^2 or μ_2. Variance is computed by squaring standard deviation. It is clear that variance is always positive and measured in terms of square units of given data. The concept of variance is very much useful in an advanced statistical work and has very important applications in inferential statistics.

Coefficient of S.D. = $\frac{S.D.}{Mean} = \frac{\sigma}{\bar{X}}$

The relative measure of dispersion corresponding to the variance is **co-efficient of variation** which is usually denoted by C.V. and is given by

$$C.V. = \frac{\sigma}{\bar{X}} \times 100$$

$$C.V. = \frac{S.D.}{Mean} \times 100\%$$

C.V. is the most widely used relative measure of dispersion in comparing two or more than two distributions. In comparison between two or more distributions, the distribution with lower C.V. is supposed to be more homogeneous or more consistent or more uniform or more regular or more equitable or more representative or more stable less variable

According to Karl Pearson, coefficient of variation is the "percentage variation in the mean". It is a relative measure of dispersion, so it is independent of units of measurement. It is always expressed in percentage. Therefore, C.V. can be used to compare two or more than two distributions with regard to their variability, consistency, uniformity, homogeneity, equitability, stability etc.

Coefficient of variation is applicable for the comparison of variability of two or more than two distributions (series) as follows

Less C.V. is considered as	**More C.V. is considered as**
More consistent	Less consistent
More homogeneous (less heterogeneous)	Less homogeneous (more heterogeneous)
More uniform	Less uniform
More stable	Less stable
More representative to mean	Less representative to mean
More equitable	Less equitable
Less variable	More variable
Less disparity	More disparity
more regular	less regular

Note: The relation between different measures of dispersion is 4S.D. = 5M.D. = 6Q.D. and Range = 6S.D. (Approximately).

Properties of Standard Deviation

Property 1: Standard deviation is independent of change of origin but not of scale.
Property 2: For identical observation, standard deviation is zero.

Eg: For 4, 4, 4, 4, 4, 4, Standard Decviation is zero.

Property 3: For any discrete distribution standard deviation is not less than mean deviation from mean.

i.e. $\sqrt{\frac{1}{N}\Sigma f\left(X-\overline{X}\right)^2} \geq \frac{1}{N}\Sigma f\left|X-\overline{X}\right|$

Property 4. Combined standard deviation

Let $\overline{X}_1$ and $\overline{X}_2$ be the means of first and second series, σ_1^2 and σ_2^2 be the variances of first and second series with n_1 and n_2 number of observations respectively. Then the combined standard deviation of these two series taken together is given by

$$\sigma_{12} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

Where, $d_1 = \overline{X}_1 - \overline{X}_{12}$, $d_2 = \overline{X}_2 - \overline{X}_{12}$

$$\overline{X}_{12} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$$

Similarly, the combined standard for three series is given by

$$\sigma_{123} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_3\sigma_3^2 + n_1 d_1^2 + n_2 d_2^2 + n_3 d_3^2}{n_1 + n_2 + n_3}}$$

Where, $d_1 = \overline{X}_1 - \overline{X}_{123}$

$d_2 = \overline{X}_2 - \overline{X}_{123}$

$d_3 = \overline{X}_3 - \overline{X}_{123}$

$$\overline{X}_{123} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3}$$

Example 2.67: Calculate the combined mean and standard deviation from the following information

	Factory A	**Factory B**
No. of workers	100	500
Daily Mean wage (Rs.)	50	60
Standard deviation	10	11

Solution: Here, $N_1 = 100$, $N_2 = 500$, $\overline{X}_1 = 50$, $\overline{X}_2 = 60$, $\sigma_1 = 10$ and $\sigma_2 = 11$

``Now,

Combined Mean ($\overline{X}_{12}$) $= \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$

$= \frac{100 \times 50 + 500 \times 60}{100 + 500}$

$$= \frac{35000}{600} = \text{Rs. } 58.3.$$

And

$d_1 = \bar{X}_1 - \bar{X}_{12} = 50 - 58.3 = -8.3$

$d_2 = \bar{X}_2 - \bar{X}_{12} = 60 - 58.3 = 1.7$

$$\text{Combined S.D. } (\sigma_{12}) = \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_2^2 + d_2^2)}{N_1 + N_2}}$$

$$= \sqrt{\frac{100[10^2 + (-8.3)^2] + 500[11^2 + (1.7)^2]}{100 + 500}}$$

$$= 11.46$$

Example 2.68: An analysis of monthly wages paid to the workers in two factories M and N belonging to the same industry give the following information.

	factory M	**factory N**
No. of workers:	600	700
Average monthly wage (Rs):	182	178.50
Variance of distribution of wage:	78	98

i) Which factory M or N has a larger wage bill?

ii) In which factory M or N is there more uniformity in distribution of wages?

Solution: For Factory M

Here, n = 600, $\bar{X}_M = 182$, $\sigma_M^2 = 78$ i.e. $\sigma_M = \sqrt{78} = 8.83$

Total wage bill = $\sum X_M = n.\bar{X}_M = 600 \times 182 = \text{Rs.}1,09,200$

Co-efficient of variance $(C.V_M) = \frac{\sigma_M}{\bar{X}_M} \times 100 = \frac{8.83}{182} \times 100 = 4.85\%$

For Factory N

Here, n = 700, $\bar{X}_N = 178.50$, i.e. $\sigma_N = \sqrt{98} = 9.89$

Total wage bill = $\sum X_N = n . \bar{X}_N = 700 \times 178.50 = \text{Rs. } 1,24,950$

Co-efficient of variance $(C.V_N) = \frac{\sigma_N}{6_N} \times 100\% = \frac{9.89}{178.50} \times 100 = 5.54\%$

i) Since $\sum X_N > \sum X_M$, so, factory N has a larger wage bill by Rs. (1,24,950 – 1,09,200) = Rs.15,750

ii) Since $C.V_M < C.V_N$, so there is more uniformity in distribution of wages in factory M.

Example 2.69 Compute appropriate measures of dispersion from the following table.

Marks	Below 10	10-20	20-30	30-40	40-50	50 and Above
Frequency	5	8	7	12	28	20

Solution: Since the given frequency distribution has open end classes. So measure of dispersion based on partition values (quartiles) is appropriate i.e. Quartile Deviation is appropriate.

Here, we construct the continuous frequency distribution as shown below and calculate the Q.D.

Marks	Frequency (f)	c.f.
Below 10	5	5
10 – 20	8	13
20 – 30	7	20
30 – 40	12	32
40 – 50	28	60
50 and Above	20	80
Total	**N = 80**	

Position of Q_1 is given by $\frac{N}{4} = \frac{80}{4} = 20$. From c.f. table, the c.f. just greater than 20 is 32. Thus Q_1 lies in the class interval (30 - 40).

Now, $Q_1 = L + \frac{\frac{N}{4} - c.f}{f} \times h = 30 + \frac{20 - 20}{12} \times 10 = 30$

And position of Q_3 is given by $\frac{3N}{4} = 3 \times \frac{80}{4} = 60$. From c.f. table, the c.f. just greater than 60 is 80. Thus Q_3 lies in the class interval (50 - 60).

Now, $Q_3 = L + \frac{\frac{3N}{4} - c.f}{f} \times h = 50 + \frac{(60 - 60)}{20} \times 10 = 50$

Therefore, Quartile Deviation (Q.D.) = $\frac{Q_3 - Q_1}{2} = \frac{50 - 30}{2} = 10$

Example 2.70 The following are the runs by two cricketers in 10 matches.

Matches	1	2	3	4	5	6	7	8	9	10
Cricketer A:	24	37	27	30	31	34	36	26	29	33
Cricketer B:	22	40	38	24	26	36	34	28	30	27

If the consistency of performance is the criterion for selecting in National team, which cricketer should be selected?

Solution: To test the consistency of performance, the C.V. should be calculated.

For cricketer A, Assumed Mean (A_1) = 30, , $n_1 = 10$

X_1	$d_1 = X_1 - A_1$	d_1^2
24	-6	36
26	-4	16
27	-3	9
29	-1	1
30	0	0
31	1	1
33	3	9
34	4	16
36	6	36
37	7	49
Total	$\sum d_1 = 7$	$\sum d_1^2 = 173$

$$\text{Mean } (\overline{X}_A) = A_1 + \frac{\sum d_1}{n_1}$$

$$= 30 + \frac{7}{10} = 30.7$$

$$\text{Standard deviation } (\sigma_A) = \sqrt{\frac{\sum d_1^2}{n_1} - \left(\frac{\sum d_1}{n_1}\right)2}$$

$$= \sqrt{\frac{173}{10} - \left(\frac{7}{10}\right)2}$$

$$= \sqrt{17.3 - 0.49}$$

$$= \sqrt{16.81} = 4.1$$

$$C.V_A = \frac{\sigma_A}{\overline{X}_A} \times 10\% = \frac{4.1}{30.7} \times 100 = 13.35\%$$

For cricketer B, Assumed Mean (A_2) = 30, n_2 = 10

$\mathbf{X_2}$	$\mathbf{d_2 = X_2 - A_2}$	d_2^2
22	-8	64
24	-6	36
26	-4	16
27	-3	9
28	-2	4
30	0	0
34	4	16
36	6	36
38	8	64
40	10	100
Total	$\mathbf{\sum d_2 = 5}$	$\mathbf{\sum d_2^2 = 345}$

$$\text{Mean } (\overline{X}_B) = A_2 + \frac{\sum d_2}{n_2} = 30 + \frac{5}{10} = 30.5$$

$$\text{Standard deviation } (\sigma_B) = \sqrt{\frac{\sum d^2_2}{n_2} - \left(\frac{\sum d_2}{n_2}\right)2}$$

$$= \sqrt{\frac{345}{10} - \left(\frac{5}{10}\right)2}$$

$$= \sqrt{34.5 - 0.25}$$

$$= \sqrt{34.25} = 5.85$$

$$C.V_B = \frac{\sigma_B}{\overline{X}_B} \times 100\% = \frac{5.85}{30.5} \times 100 = 19.18\%$$

Since, $C.V_A < C.V_B$. So performance of A is more consistent than performance of B. Therefore, cricketer A should be selected.

Example 2.71: A sample of 500 cars of each of two makes X and Y is taken and average running life in years is recorded.

Life (no. of years)	No. of cars	
	Make X	**Make Y**
0 – 2	80	60
2 – 4	120	100
4 – 6	170	200
6 – 8	100	120
8 – 10	30	20

If prices of car are same, which makes car should be preferred by the buyer?

Solution: To decide, which make should be preferred by the buyer. We should calculate C.V. of each make.

For make X

Calculation of Mean and Standard Deviation

Life (no. of yrs)	**Mid value (m)**	**No. of cars (f)**	**d = m – A**	**fd**	**fd²**
0 – 2	1	80	–4	–320	1280
2 – 4d	3	120	–2	–240	480
4 – 6	5	170	0	0	0
6 – 8	7	100	2	200	400
8 – 10	9	30	4	120	480
		$N = 500$		**$\sum fd = 240$**	**$\sum fd^2 = 2640$**

Here, Assumed Mean (A) = 5, $N = 500$

$$\text{Mean } (\bar{X}_x) = A + \frac{\sum fd}{N} = 5 + \frac{-240}{500} = 5 - 0.48 = 4.52$$

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= \sqrt{\frac{2640}{500} - \left(\frac{-240}{500}\right)^2} = 2.247$$

$$\text{C.V}_X = \frac{\sigma_X}{\bar{X}_x} \times 100\%$$

$$= \frac{2.247}{4.52} \times 100\% = 49.71\%$$

For Make Y

Life (no. of yrs)	Mid value (m)	f	$d = m - A$	fd	fd^2
0 – 2	1	60	–4	–240	960
2 – 4	3	100	–2	–200	400
4 – 6	5	200	0	0	0
6 – 8	7	120	2	240	480
8 – 10	9	20	4	80	320
		$N = 500$		$\sum fd = -120$	$\sum fd^2 = 2160$

Here, Assumed Mean (A) = 5, $N = 500$

$$\text{Mean } (\bar{X}_Y) = A + \frac{\sum fd}{N} = 5 + \left(\frac{-120}{500}\right)$$

$$= 5 - 0.24 = 4.76$$

$$\text{Standard deviation } (\sigma_Y) = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$

$$= \sqrt{\frac{2160}{500} - \left(\frac{-120}{500}\right)^2} = 2.06$$

$$C.V_Y = \frac{\sigma_Y}{\bar{X}_Y} \times 100\% = \frac{2.06}{4.76} \times 100\% = 43.28\%$$

Here, we see that $C.V_Y < C.V_X$, so makes Y is more consistent in life duration than that of X. Thus makes Y is preferred by the buyer.

Example 2.72: Following are the marks obtained by two students A and B in 10 test of 100 marks each.

Test 1	1	2	3	4	5	6	7	8	9	10
Marks of A	54	90	86	58	62	82	78	66	70	64
Marks of B	58	85	64	70	73	79	82	61	67	76

(a) Who is better?

(b) Who is intelligent?

(c) If the consistency of performance is the criteria for awarding a prize. Who should get the prize?

Solution: Arrange the marks of A and B according to ascending order

For student *A*			**For student *B***		
X	$d = X - 70$	d^2	***X***	$d = X - 70$	d^2
54	– 16	256	58	– 12	144
58	– 12	144	61	– 9	81
62	– 8	64	64	– 6	36
64	– 6	36	67	– 3	9
66	– 4	16	70	0	0
70	0	0	73	3	9
78	8	64	76	6	36
82	12	144	79	9	81
86	16	256	82	12	144
90	20	400	85	15	256
Total	$\Sigma d = 10$	$\Sigma d^2 = 1380$		$\Sigma d = 15$	$\Sigma d^2 = 796$

(a) $\bar{X}_A = A + \frac{\Sigma d}{n} = 70 + \frac{10}{10} = 71.$

$\bar{X}_B = A + \frac{\Sigma d}{n} = 70 + \frac{15}{10} = 71.5.$

Since, $\bar{X}_A < \bar{X}_B$, therefore, B is better.

(b) For student *A*

$$M_d(A) = \text{value of } \left(\frac{n+1}{2}\right)^{th} \text{ items}$$

$$= \text{ value of } \left(\frac{10+1}{2}\right)^{th} \text{ item}$$

$$= \text{ value of } (5.5)^{th} \text{ item}$$

$$\therefore \; M_d(A) = \frac{66+70}{2} = 68 \text{ marks}$$

Similarly, for student *B*

For student *B*,

$$M_d\,(B) = \text{ value of } \left(\frac{n+1}{2}\right)^{th} \text{ item}$$

$$= \text{ value of } \left(\frac{10+1}{2}\right)^{th} \text{ item}$$

$$= \text{ value of } (5.5)^{th} \text{ item}$$

$$M_d(B) = \frac{70+73}{4} = 71.5$$

Since, $M_d(A) < M_d(B)$

Hence, B is more intelligent than A

(c) For student A

$$\text{S.D. } \sigma(A) = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} = \sqrt{\frac{1380}{10} - \left(\frac{10}{10}\right)^2}$$

$$= \sqrt{137} = 11.70.$$

For student B

$$\text{S.D. } \sigma(B) = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} = \sqrt{\frac{796}{10} - \left(\frac{10}{10}\right)^2}$$

$$= \sqrt{77.35} = 8.79.$$

For student A, C.V. (A) $= \dfrac{\sigma(A)}{\bar{X}_A} \times 100\ \%$

$$= \frac{11.70}{71} \times 100\% = 16.48\%$$

For student B, C.V. (B) $= \dfrac{\sigma(B)}{\bar{X}_B} \times 100\ \%$

$$= \frac{8.79}{71.5} \times 100\% = 12.29\%.$$

Conclusion: Since C.V. (student B) < C.V.(student A), so performance of student B is more consistent than that of student A. Therefore, if the consistency of performance is the criteria for awarding a prize, student B should get the price.

Example 2.73: The mean and standard deviation of 100 items are found to be 40 and 10 respectively. On consequent investigation, two items were wrongly taken as 30 and 70 instead of 3 and 27. Find the correct mean and correct standard deviation.

Solution:

Here, $n = 100$, $\overline{X} = 40$, S.D. $(\sigma) = 10$

We have $\sum X = n.\overline{x} = 100 \times 40 = 4000$

And S.D. $(\sigma) = \sqrt{\frac{\sum X^2}{n} - (\overline{X})^2}$

or, $10 = \sqrt{\frac{\sum x^2}{100} - 40^2}$

or, $100 = \frac{\sum x^2}{100} - 1600$

or, $100 + 1600 = \frac{\sum x^2}{100}$

$\therefore \sum x^2 = 100 \times 1700 = 170000$

According to question, two items 30 and 70 are wrongly taken instead of 3 and 27. So,

Corrected $\sum x = 4000 - 30 - 70 + 3 + 27 = 2930$

and corrected $\sum x^2 = 170000 - 30^2 - 70^2 + 3^2 + 27^2 = 164938$

Now,

Corrected Mean $= \frac{\text{corrected } \sum X}{n} = \frac{3930}{100} = 39.3$

& Corrected S.D. $(\sigma) = \sqrt{\frac{\text{corrected } \sum X^2}{n} - (\overline{X})^2}$

$= \sqrt{\frac{164938}{100} - (39.3)^2} = 10.24$

"Quality Never Says Sorry"

“Better go meticulously through the note”

“Stay safe, stay at home”

Exercise 2.1

Numerical and practical problems:

1. The production of paddy in five places is as follows:

Place	*A*	*B*	*C*	*D*	*E*
Production ('000' tones)	500	520	480	570	610

Find the arithmetic mean.

2. The following table gives the basic salaries of the persons employed in a factory. Calculate the average basic salary by using
   a. Direct method
   b. Short-cut method
   c. Step deviation method

Salary (Rs.)	1100	1300	1500	1700	1900	2100	2300
No. of persons	5	7	10	15	13	16	14

3. Calculate the mean of the marks obtained by the students from the data given below:

Marks Group	0 – 10	10–20	20–30	30–40	40–50	50–60	60–70
No. of students	4	8	11	15	12	7	3

4. Calculate the mean from the following data using direct method.

Marks	10–20	10–30	10–40	10–50	10–60	10–70	10–80	10–90
No. of students	4	16	56	97	124	137	146	150

5. Calculate average wage of the workers. Assume that none are earning more than Rs.200.

Daily wages (Rs.) more than	20	40	60	80	100	120	140	160	180
No. of workers	100	94	85	74	60	40	25	15	7

6. a) Income of employees of a factory is given in the following tale. The total income of the 10 employees in the highest income group is Rs.3,000. Compute the arithmetic mean of the income.

Income (Rs.)	0 –50	50 –100	100–150	150–200	200–250	250&over
Frequency	90	150	100	80	70	10

   b) From the following data of income distribution, calculate the arithmetic mean. It is given that the total income of the person in the highest group is Rs.435 and none is earning less than Rs.20

Income (Rs.) below	30	40	50	60	70	80	80 & above
No. of persons	16	36	61	76	87	95	5

7. A factory pays its workers on a piece rate basis and also a bonus to each worker on the basis of individual output in each month. The rate of bonus payable is as follows:

Output (unit)	Below 75	75–79	80–84	85–89	90–94	95–99	100 and over
Bonus (Rs.)	35	45	50	60	70	80	100

The individual output of 50 workers is given below:

94	83	78	76	88	86	93	80	91	82
89	97	92	84	82	80	85	83	98	103
90	87	81	99	86	95	81	88	88	87

84	97	80	75	93	101	82	82	89	72
80	71	87	77	98	83	72	75	83	85

Calculate the average bonus per worker for the month.

8. The following are the weekly production of the product X in units of 60 workers in a manufacturing company:

23	48	51	64	82	19	33	50	39	72	35	88
77	25	39	52	48	64	49	57	41	72	62	49
32	54	67	46	55	52	82	44	75	56	51	63
59	69	53	57	75	85	68	55	52	45	40	57
20	42	46	51	50	16	62	56	54	40	55	75

The management has decided to give bonus of Rs.5, 10, 15, 20 and 25 to each worker in the respective output group of 40 or over weekly output. Find the average bonus received by the workers.

9. The mean of the following distribution is 1.46. Find the missing frequencies.

Value	0	1	2	3	4	5	Total
Frequency	46	?	?	25	10	5	200

10. 100 Salesman were appointed in various places of Kathmandu valley and the following data were compiled from their sales reports.

Sales (in Rs.'000')	4–8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
No. of salesman	11	13	16	14	–	9	–	6	4

If the average sale is believed to be Rs.19.92, find the missing frequencies.

11. a) The following table represents the weekly wages of the workers in a firm. Calculate the average weekly wage per worker.

Wages (Rs.)	100–120	120–140	140–160	160–180	180–200	200–220
Total hours	180	140	156	153	195	143
Average no. of hrs worked per week	15	10	12	9	15	13

b) Find the mean from the following distribution of average number of teachers of the schools with different income groups.

Income (central value)	1200	1400	1600	1800	2000
No. of schools	4	6	7	6	2
Average No. of teachers per school	10	13	15	14	9

12. A cold store sells five different products. Find the average profit per unit of the quantity sold from the following information.

Product	Profit per units (Rs.)	Quantity sold (unit)
A	4	150
B	9	50
C	6	250
D	2	450
E	12	100

13. A professor has decided to use different weights in different evaluations of a group of student s in a year. The weights assigned are as follows:

Homework: 20%, Mid-term: 25%, Final paper: 35%, Term paper: 10% and Presentation: 10%

From the data on five students, compute the average marks of the students in the final examination. All marks of the students are given out of 100 full marks. Who should get the scholarship, if best performance is the criteria to award the scholarship?

Students	Home work	Mid term	Final paper	Term paper	Presentation
A	85	87	90	94	89
B	78	91	92	88	84
C	94	86	89	93	88
D	82	84	93	88	79
E	95	82	88	92	90

14. The following data are related to the pass percentage of the students of the three universities. Comment on the performance of the universities using the weighted average.

Program	Universities					
	Tribhuvan		Pokhara		Kathmandu	
	Pass%	No. of students	Pass%	No. of students	Pass%	No. of students
BA	74	600	73	600	73	500
B. Sc.	70	800	65	500	65	300
BBS	58	500	75	700	74	200
MA	81	400	82	300	71	300
M. Sc.	73	300	66	700	60	300
MBS	81	500	76	200	83	400

15. From the information given below, find
    a. Which factory pays larger amount as daily wage?
    b. What is the average daily wage for the workers of the two factory?

	Factory A	Factory B
No. of workers	250	200
Average daily wage (in Rs.)	20	25

16. The mean wage of 100 workers in a factory, running two shifts of 60 and 40 workers is Rs.80. The mean wage of 60 workers working in the morning shift is Rs.40. Find the mean wage of the 40 workers working in the afternoon shift.
17. The number of workers I each section of a factory and their average of daily wages are given below. Find the average of the daily wages of all the workers in the factory.

Section	Number of workers	Average of daily wages
A	50	113
B	60	120
C	70	150

18. The average scores of a group of students in a test was 52. The top 20% of them secured a mean score of 80 and the lowest 25% a mean score of 31. Find the mean score of the remaining students.
19. The pass result of 40 students who took up a class test is given below:

Marks	40	50	60	70	80	90
No. of students	8	10	9	6	4	3

If the average mark of all the 50 workers was 51.6, find out the average marks of the students who failed?

20. The mean mark obtained by 150 students in a class is 60. The mean mark of boys is 70 and that of girls is 55. Find the number of boys and girls in the class.
21. In a class of 50 students, 10 have failed and their average marks is 2.5. The total marks secured by entire class were 281. Find the average marks of those students who have passed.
22. The arithmetic mean of 100 items was 40. Later on, it was found that an item 53 was misread as 83. Find the correct mean when the wrong item is omitted.
23. The mean salary paid to 1000 employees of an establishment was found to be Rs.180.40. Later on, it was discovered that the salary of two employees was wrongly entered as Rs.297 and Rs.165. Their correct salaries were Rs.197 and Rs.185 respectively. Find the correct mean salary.
24. Find the geometric mean of the following statistical data:
    a. 10, 110, 120, 50, 52, 80, 37, 60
    b. 125, 130, 75, 10, 45, 0.5, 0.4, 500, 1505
25. Calculate the geometric mean for the following data.

x	12	13	14	15	16	17
f	5	4	4	3	2	1

26. Find the geometric mean for the following data.

Marks obtained	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of students	5	7	15	25	8

27. Find the average growth rate of population, which is increased by 20% in the first decade, by 25% in the second decade, and by the 44% in the decade.
28. A machine was purchased for Rs.50,000 in 2015 A.D. Depreciation on the diminishing balance was charged by the 30% in the first year, 25% in the second year and 15% per annum during the next three years. Find the average rate of depreciation.
29. a) Find the harmonic mean of 4, 6 and 10.
    b) Compute the harmonic mean for the following data.

x	2	4	6	8
f	2	3	3	2

30. a) A car driver covers a distance of 200 kilometers from Kathmandu to Pokhara at the rate of 50 km/hr. In return journey, he covers the distance at the rate of 100 km/hr. Find the average speed of the journey to and info.
    b) Cities A, B and C are equidistant from each other. A motorist travels from A to B at 30km/hr, from B to C at 40/km/hr and from C to A at 50km/hr. Determine his average speed for the entire trip.
    c) A man travelled by car for 3 days. He covered 480 kms each day. On the first day, he drove for 10 hours at 48 km/hr. On the second day he drove for 12 hours at 40 km/hr and on last day he drove for 15 hours at 32 km/hr. What was his average speed?
31. Find the geometric mean and harmonic mean from the following data:

Class interval	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	30	75	70	60	15

32. a) Compute AM, GM and HM of the following observations and verify that AM > GM > HM
    10, 12, 14, 16, 18, 20
    b) If for two observations, the arithmetic mean is 25 & harmonic mean is 9, what is the geometric mean of the observations?
    [**Hint:** Use the relationship between Am, GM and HM. i.e. $GM^2 = AM \times HM$]
33. Find the median from the following statistical data.
    a) 15, 10, 5, 13, 12, 1, 15, 9, 8, 18

b) 40, 50, 30, 20, 25, 35, 30, 30, 20, 30

c)

x	10	11	12	13	14	15	16
f	5	7	9	11	9	7	5

d)

Profit (Rs.'00')	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of shops	3	5	20	10	5

34. Compute median from the following data:

C.I.	55 – 64	65 – 74	75 – 84	85 – 94	95 – 104	105 – 114	115 – 124	125 – 134	135 – 144
f	1	2	9	22	33	22	8	2	1

35. Monthly salary (in Rs.'00') of 30 employees was recorded as follows.

131	125	147	130	139	137	128	148	125	132
145	131	130	149	147	127	148	130	145	141
129	129	147	137	134	127	125	133	145	128

Prepare the frequency distribution and find the median.

36. Calculate the median from the following data.

Daily wages (in Rs.)	Less than 100	100 – 200	200 – 300	300 – 400	400 – 500
No. of workers	40	89	148	64	39

Give reason that why median is more appropriate than mean for this data?

37. a. A manufacturing company has 1000 employees. 10% of them earn less than Rs.500 per day, 200 earn between Rs.500 and Rs.999, 30% of the employees earn between Rs.1,000 and Rs.1499, 250 employees earn between Rs.1,500 and Rs.1,999 and rest of them earn Rs.2,000 and above. Calculate the suitable average wage. Give reason.

b. Calculate the appropriate measures of central tendency from the following distribution and support your choice.

Monthly Income (Rs.)	Below 1000	1000–1999	2000–2999	3000–3999	4000–4999	5000 and above
No. of families	50	500	555	100	30	15

c. Calculate the appropriate average marks from the following marks distribution. Give reason for choice.

Marks up to	15	25	35	45	55	65	75
No. of students	7	15	28	57	92	101	104

38. From a batch of 13 students who had appeared in an examination, 4 students were failed. The marks of passed students were 43, 57, 45, 61, 75, 64, 53, 50 and 40. Calculate the median marks of all students.

39. a. The monthly expenditure in rupees for a group of families is as follows:

Expenditure (in Rs.)	100–200	200–300	300–350	350–400	400–500
No. of families	8	?	20	12	5

Median of the expenditure is known to be Rs.317.50. Determine the number of families having expenditure between Rs.200 – Rs.400.

b. The expenditure of 1000 families is given below:

Expenditure (Rs.)	40 – 59	60 – 79	80 – 99	100 – 119	120 – 139
No. of families	50	–	500	–	50

If the median expenditure is Rs.87 obtain the missing frequencies

40. Find the mode from the following.

a) 4, 7, 3, 5, 3, 5, 5, 9

b) 12, 13, 14, 14, 15, 15, 15, 15, 16, 16, 16, 17, 17, 19

41. The shoe sizes of 50 customers visiting a shop were as follows:

Shoe size	6	7	8	9	10
No. of customers	4	17	22	5	2

Calculate the modal shoe size.

42. Calculate the modal weight from the following data.

Weight (in kg)	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
No. of persons	18	37	45	27	15	8

43. a) A random sample survey was conducted by ABC shoe company to determine the size of the shoe it should produce so that the size can be fit for majority of people in a population. The following information was obtained from the survey.

Size of foot (inches)	4 – 5	5 – 6	6 – 7	7 – 8	8 – 9	9 – 10
No. of people	10	35	70	35	12	6

What size of the shoe should the company produce to fulfill the objective?

b) Sigma gets a pocket money allowance Rs. 120 per day. Thinking that this was rather less, she asked her friends about their allowances and obtained the following data which includes her allowance also.

120	180	100	50	250	200	200	220	150	100
100	150	130	200	180	100	150	100	180	150
120	150	100	150	100	120	180	200	50	80

She presented this data to her father and asked for an increase in her allowance as she was getting less than average amount. Her father countered pointing out that her allowance was actually more than the average amount. Reconcile these statements with reason.

44. The frequency distribution of the marks obtained by 60 students of a class in a college is given below:

Marks	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64
No. of students	3	5	12	18	14	6	2

Find the value of mode.

45. Find out the mean and mode from the following data.

Marks	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60
No. of students	4	19	35	55	67	70

46. Compute the mode from the following data.

Mid value	10	20	30	40	50	60	70
Frequency	7	12	17	29	31	5	3

47. a) Mr. Shrestha is the director of the student financial aid office at a campus. He has used available data on summer earnings of all the students who have applied to his office for financial aid. The following frequency distribution is given below:

Summer earnings (Rs.)	0–500	500–1000	1000–1500	1500–2000	2000–2500	2500–3000	3000 and above
No. of students	231	304	400	296	123	68	23

i. Find the modal value for the Shrestha's data.

ii. If the student's aid is restricted to those summer earnings were at least modal summer earnings, how many of the applicants qualify?

b) A cement company sells his production in different cities trough the appointed dealers. The sales of his production in the last year are given in the following table:

Sales (in '00' bags)	0 – 500	500 – 1000	1000 – 1500	1500 – 2000	2000 – 2500	2500 & above
No. of dealers	40	48	60	52	35	22

i. Find the value of most usual sales.

ii. Find the number of dealers selling more that the usual sales.

iii. Calculate the amount of money to be awarded to the dealers at Rs. 5000 each dealer whose annual sale is more than the most usual sale.

48. In the marks distribution of 100 students given below, frequencies corresponding to two groups are missing from the table. However, the mode is known to be 24. Find the missing frequencies.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	14	?	27	?	15

49. The weekly expenditure of 1000 families is given below:

Expenditure (in Rs.'00')	40 – 59	60 – 79	80 – 99	100–119	120–139
No. of families	50	–	500	–	50

The median for the distribution is Rs. 8700. Calculate the missing frequencies. Also calculate the mode of the distribution.

50. From the following data, compute the mean, median, and modal output of all the workers of 50 factories in an industrial city.

Mid-value of output	15	20	25	30	35	40
No. of factories	8	11	14	8	6	3
Average No. of workers	100	120	125	125	150	200

51. a) In a moderately skewed distribution, if an arithmetic mean = 24.6, and the mode = 26.1. Find the value of median.

b) In a moderately symmetrical distribution, the value of mode and median are 20 and 24 respectively. Find the value of mean.

c) If mean = 40 and mode = 30 then find median.

52. Compute the first quartile, sixth decile and 82^{th} percentile from the following data.

82 56 90 50 120 75 75 80 130 65

53. Calculate the quartiles (upper quartile and lower quartile), 7^{th} decile and 60^{th} percentile from the following information.

C.I.	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
F	1	3	11	21	43	32	9

54. From the following distribution of marks of 250 students of a campus, find the minimum pass marks if only 20% of the students has failed and also find the minimum marks obtained by the top 25% of the students.

Marks	0 – 20	20 – 40	40 – 50	50 – 60	60 – 80	80 – 100
No. of students	25	50	75	45	30	25

55. The marks distribution of 50 students in a subject is given below:

Marks more than	0	10	20	30	40	50
No. of students	50	46	40	20	10	3

If 60% of the students passed this test, find the minimum marks obtained by the pass candidate.

56. From the following distribution of income of 1500 persons, find
    a. Limits of central 60% of the persons.
    b. Lowest income of richest 60% of the persons
    c. Highest income of poorest 60% of the persons

Income (in Rs.'000')	0 – 5	5–10	10–15	15–20	20–25	25–30	30–35
No. of persons	100	150	400	550	200	75	25

57. The following table shows the wage distribution in a certain factory.

Weekly wage	No. of employees	Weekly wages (Rs.)	No. of employees
20–40	8	120–140	35
40–60	12	140–160	18
60–80	20	160–180	7
80–100	30	180–200	5
100–120	40		

Determine:
    a. The wage limits for he middle 50% of the wage earners.
    b. The percentage of workers who earned between Rs. 75 and Rs.125.

Income (in Rs.'000')	0 – 5	5–10	10–15	15–20	20–25	25–30	30–35
No. of persons	100	150	400	550	200	75	25

Answers

Numerical and Practical Problems:

1. 536000 tones **2.** Rs.1,820 **3.** 34.333 **4.** 46.333 marks
5. Rs.110 **6.** a) Rs.117.50 b) Rs. 48 **7.** Rs.59
8. Rs 12 **9.** 76, 38 **10.** 10, 17 **11.** a) Rs.159.50
b) Rs. 1576.62 **12.** $\bar{X}_w$ = Rs. 4.65
13. $\bar{X}_{wA} =$ 88.55, $\bar{X}_{wB}$ = 87.75, $\bar{X}_{wC}$ = 89.55, $\bar{X}_{wD}$ = 86.65, $\bar{X}_{wE}$ = 88.50. C should get the scholorship
14. $\bar{X}_{wT}$ = 72.32, $\bar{X}_{wP}$ = 71.60, $\bar{X}_{wK}$ = 71.6, Tribhuwan University
15. a) Equal b) Rs. 22.22 **16.** Rs.140
17. Rs.129.72 **18.** 51.36 **19.** 21 **20.** 50 boys and 100 girls.
21. 6.4 marks **22.** 39.7, 39.57 **23.** Rs.180.32 **24.** a) 52.84
b) 35.17 **25.** 13.71 **26.** 25.64 marks **27.** 28.02%
28. 19.08% **29.** a) 5.77 b) 4.44 **30.** a) 66.67km/hr

b) 38.3 km/hr c) H.M. = 37.8678 km/hr

31. GM = 31.19, HM = 29.04 **32.** a) AM = 15.5, GM = 12, HM = 9.93

b) 15 **33.** a) 12 b) 30 c) 13

d) Rs.4,675 **34.** 99.34 **35.** Rs.13,375 **36.** Rs.241.22

37. a) Rs.1,332.83 b) Rs.2,134.635

c) 43.28,the frequency distribution contains open ended class interval. **38.** 45

39. (a) 54, (b) 263, 137 **40.** a) 5 b) 15

41. 8 **42.** 53.07kg **43.** a) M_o = 6.5inch

b) Saraswati computed arithmetic mean and her father computed Mode.

44. 47.5 marks **45.** 30, 33.33 **46.** 45.71

47. a) i) M_o = Rs.1,240 ii) 727 b) i) Rs.1,30,000 ii) 133

iii) 6,65,000 **48.** 23, 21 **49.** 263, 137, M_0 = Rs.8740

50. Mean = 26.32, Median = 25.54 & Mode = 24.32 **51.** a) $M_d = 25.1$ b) Mean = 26

c) 36.67 **52.** $Q_1 = 62.75$, $D_6 = 81.20$ $P_{82} = 120.20$

53. $Q_1 = 67.14$, $Q_3 = 83.44$, $D_7 = 81.56$, $P_{60} = 78.37$

54. 30, 58.33 **55.** 25 marks **56.** a) Rs.10,625 and Rs.20,000

b) Rs.14,375 c) Rs.17, 273

57. (a) the limits for the central 50% of the employees are 82.5 and 132.14. (b) 48%

Exercise 2.2

Multiple Choice Questions

Circle (O) the correct answer.

1. Which is more appropriate central tendency to find the average of profit?
   (a) Arithmetic mean (b) Median (c) Mode (d) All
2. Mean is a measure of
   (a) location (central value) (b) dispersion
   (c) correlation (d) none of the above
3. Which of the following is a measure of central value?
   (a) Median (b) Standard deviation
   (c) Mean deviation (d) Quartile deviation
4. Which of the following represents median?
   (a) First quartile (b) Fiftieth percentile (c) Sixth decile (d) None of the above
5. If a constant value 50 is subtracted from each observation of a set, the mean of the set is:
   (a) increased by 50 (b) decreased by 50 (c) is not affected (d) zero
6. If a constant 5 is added to each observation of a set, the mean is:
   (a) increased by 5 (b) decreased by 5
   (c) 5 times the original mean (d) not affected

7. If each observation of a set in multiplied by 10, the mean of the new set of observations:

(a) remains the same (b) is ten times the original mean

(c) is one-tenth of the original mean (d) is increased by 10

8. If each value of a series is multiplied by 10, the median of the coded values is:

(a) not affected (b) 10 times the original median value

(c) one-tenth of the original median value (d) increased by 10

9. If each value of a series is multiplied by 10, the mode of the coded values is:

(a) not affected (b) one-tenth of the original modal value

(c) 10-ties of the original modal value (d) 100-times of the original modal value

10. If each observation of a set is divided by 2, then the mean of new values:

(a) is two times the original mean, (b) is decrease by 2

(c) is half of the original mean (d) remains the same

11. Which of the following relations among the location parameters does not hold?

(a) Q_2 = Median (b) P_{50} = Median (c) D_5 = Median (d) D_6 = Median

12. Harmonic mean is better than other means if

(a) speed or rates (b) heights or lengths

(c) binary values like 0 and 1 (d) ratios or proportions

13. The correct relationship between A.M., G.M and H.M. is:

(a) A.M. = G.M. = H.M. (b) G.M. ≥ A.M. ≥ H.M.

(c) H.M. ≥ G.M. ≥ A.M. (d) A.M. ≥ G.M. ≥ H.M.

14. Extreme value have no effect on:

(a) average (b) median (c) geometric mean (d) harmonic mean

15. Geometric mean of two numbers $\frac{1}{16}$ and $\frac{4}{25}$ is:

(a) $\frac{1}{10}$ (b) $\frac{1}{100}$ (c) 10 (d) 100

16. Expenditure during first five months of a year is Rs. 96 per month and during last seven months is Rs. 120 per month. The average expenditure per month during whole year is:

(a) Rs. 108 per month (b) Rs. 110 per month (c) Rs. 100 per month (d) Rs. 216 per month

17. Average strength of eleven members = 11.0. Average strength of the first six members = 10.5. Average strength of the last six members = 11.5.

The average strength of the sixth member is:

(a) 10.5 (b) 11.5 (c) 11.0 (d) 10.0

18. The average of the 7 number 7, 9, 12, x, 5, 4, 11 is 9. The missing number x is:

(a) 13 (b) 14 (c) 15 (d) 8

19. The mean proportion of 0.16 and 0.01 is:

(a) 0.4 (b) 0.17 (c) 0.085 (d) 0.04

20. A train covered the first 5 km of its journey at a speed of 30 km/h and next 15 km at a speed of 45 km/h. the average speed of the train was:

(a) 30 km/h (b) 40 km/h (c) 32 km/h (d) 42 km/h

21. The second of the two samples has 50 item with mean 15. If the whole group has 150 items with mean 16, the mean of the first sample is:

(a) 18.0 (b) 15.5 (c) 16.5 (d) none of the above

22. For a group of 100 candidates, the mean was found to be 40. Later on it was discovered that a value 45 was misread as 54. The correct mean is:

(a) 40.50 (b) 39.85 (c) 39.80 (d) 39.91

23. A distribution consists of three groups having 40, 50 and 60 items with means 20, 26 and 15 respectively. The mean of the distribution is:

(a) 20 (b) 18

Answer

1. a	2. a	3. a	4. b	5. b	6. a	7. b	8. b	9. c	10. c	11. d
12. a	13. d	14. b	15. a	16. b	17. c	18. c	19. d	20. b	21. c	22. d
23. a	24. c	25. d	26. d	27. a	28. d	29. b	30. c			

"Quality Never Says Sorry"

“Better go meticulously through the note”

“Stay safe, stay at home”

***