Chapter 23

Measures of Central Tendency

23.1 INTRODUCTION

The term 'average' is very commonly used in day to day conversation. For example, we often talk of an average student in a class, average price, average speed of a vehicle etc. The most important objective of statistical analysis is to get one single value that represent or describes the entire data. Such a single value is called average or central value.

23.2 DEFINITION OF AVERAGE

"Average is an attempt to find one single figure to describe whole of figures". —**Clark**

23.3 PROPERTIES OF A GOOD AVERAGE

An average value should satisfy the following properties :

1. It should be easy to understand.
2. It should be simple to calculate.
3. It should be based on all observations.
4. It should be rigidly defined.
5. It should not be affected by extreme observations.
6. It should have sampling stability.

23.4 TYPES OF AVERAGES

1. Mathematical averages.
2. Positional averages.

Mathematical Averages

The following are two important types of mathematical averages :

(*a*) Arithmetic mean—simple and weighted.
(*b*) Geometric mean
(*c*) Harmonic mean.

Positional Averages

(*a*) Median
(*b*) Mode
(*c*) Quartiles, deciles, percentiles, etc.

23.5 SIMPLE ARITHMETIC MEAN

FORMULAE

1. Case of Individual Observations

Let the observations be $X_1, X_2, \ldots\ldots, X_n$

Then arithmetic mean or A.M. of these observations denoted by $\overline{X}$ is given by

$$\overline{X} = \frac{X_1 + X_2 + \dots\dots\dots + X_n}{n}$$

or

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Short-cut method :

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

where $d = X - A$

A = assumed mean.

2. In Case of Discrete Frequency Distribution

Let $X_1, X_2, X_3 \dots\dots, X_n$ be n observations occurring with frequencies $f_1, f_2, \dots, f_n$ respectively. The A.M. can be calculated by applying :

(a) *Direct method* :

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

where N is the total number of observations = Σf.

(b) *Short-cut method* :

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

where $d = X - A$

A = assumed mean

$N = \Sigma f$ = total number of observations.

3. In Case of Continuous Series

A.M. can be calculated by applying the following methods :

(a) *Direct method* :

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

where m = mid-value of each class interval = $\frac{\text{lower limit} + \text{upper limit}}{2}$

$n = \Sigma f$ = total number of observations

f = frequency of each class.

(b) *Short-cut method or step deviation method* :

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

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

A = assumed mean

i = step factor or class size.

EXAMPLES

EXAMPLE 23.1 : The following table gives the marks obtained by 10 students of a class :

Roll no. :	1	2	3	4	5	6	7	8	9	10
Marks :	43	60	37	48	65	48	57	78	31	59

Find the mean marks.

SOLUTION : This is a case of individual observations

$$\bar{X} = \frac{\Sigma X}{n}$$

$$= \frac{43+60+37+48+65+48+57+78+31+59}{10}$$

$$= \frac{526}{10} = 52.6 \text{ marks.}$$

EXAMPLE 23.2 : Compute mean in above example by short-cut method. Take assumed mean as 50.

SOLUTION : **Calculation of Arithmetic Mean**

Roll No.	*Marks X*	*d = X - 50*
1	43	−7
2	60	10
3	37	−13
4	48	−2
5	65	15
6	48	−2
7	57	7
8	78	28
9	31	−19
10	59	9
		$\Sigma d = 26$

Take A = 50

$$\bar{X} = A + \frac{\Sigma d}{n} = 50 + \frac{26}{10} = 50 + 2.6 = 52.6 \text{ marks.}$$

EXAMPLE 23.3 : Compute arithmetic mean from the following data :

290.21, 290.22, 290.16, 290.17, 290.15, 290.17,
290.12, 290.15, 290.13, 290.12

SOLUTION : This is a case of indiviudal observations. We have short-cut method to calculate A.M. ...g assumed mean A = 290.

X	*d = X − 290*
290.21	0.21
290.22	0.22
290.16	0.16
290.17	0.17
290.15	0.15
290.17	0.17
290.12	0.12
290.15	0.15
290.13	0.13
290.12	0.12
	$\Sigma d = 1.60$

$$\overline{X} = A + \frac{\Sigma d}{n} = 290 + \frac{1.6}{10} = 290 + 0.16 = 290.16.$$

EXAMPLE 23.4 : Calculate the arithmetic mean for the following data :

X :	3.5	4.5	5.5	6.5	7.5	8.5	9.5
f :	3	7	22	60	85	32	8

SOLUTION : The data is given in the form of discrete series

So, $$\overline{X} = A + \frac{\Sigma fd}{n}$$

where $d = X - A$.

Take $A = 6.5$

Computation of AM

X	*f*	*d* = *X* – 6.5	*fd*
3.5	3	– 3	– 9
4.5	7	– 2	– 14
5.5	22	– 1	– 22
6.5	60	0	0
7.5	85	1	85
8.5	32	2	64
9.5	8	3	24
	N = 217		Σfd = 128

$$\overline{X} = 6.5 + \frac{128}{217} = 6.5 + 0.589 = 7.089.$$

EXAMPLE 23.5 : Compute mean marks from the data gi en below :

Marks :	5	15	25	35	45	55	65
No. of students :	4	6	10	20	10	6	4

SOLUTION : The data is given in the form of discrete series :

Marks *X*	*No. of students* *f*	*fX*
5	4	20
15	6	90
25	10	250
35	20	700
45	10	450
55	6	330
65	4	260
	N = 60	2100

$$\overline{X} = \frac{\Sigma fX}{n} = \frac{2100}{60} = 35 \text{ marks.}$$

EXAMPLE 23.6 : Compute the mean age from the data given below :

Age (years) :	20-25	25-30	30-35	35-40	40-45	45-50
No. of persons :	170	110	[illegible]	[illegible]	40	35

SOLUTION : The data is given in the form of class intervals :

Calculation of Mean A = 325, i = 5

Age	*Mid-values (m)*	*No. of persons (f)*	$d = \frac{m - 32.5}{5}$	[illegible]
20-25	22.5	170	– 2	[illegible]
25-30	27.5	110	– 1	[illegible]
30-35	32.5	80	0	[illegible]
35-40	37.5	45	1	[illegible]
40-45	42.5	40	2	[illegible]
45-50	47.5	35	3	[illegible]
		N = 480		[illegible]

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

$$= 32.5 + \frac{(-220)}{480} \times 5 = 32.5 - 2.29 = 30.21 \text{ years.}$$

EXAMPLE 23.7 : Compute the mean marks obtained by the students of B.Com. (Pass) [illegible] :

Marks :	0-10	10-20	20-30	30-40	40-50
No. of students :	4	6	10	20	10

SOLUTION : Data is given in the form of continuous series :

Calculation of Mean

Marks	*Mid-values (m)*	*No. of students (f)*	*fm*
0-10	5	4	20
10-20	15	6	90
20-30	25	10	250
30-40	35	20	700
40-50	45	10	450
		Σf = 50	Σfm = 1510

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{1510}{50} = 30.2.$$

EXAMPLE 23.8 : The following table gives the profit earned by various companies from the sale of [illegible]. Compute the average profit earned.

Profit (Rs in crore)	*No. of Companies*
Less than 20	5
Less than 30	22
Less than 40	48
Less than 50	60
Less than 60	83
Less than 70	100

SOLUTION : Since we are given the cumulative frequencies, we first find the simple frequenc-

Calculation of Mean

Profit (Rs crore)	*m.p (m)*	*No. of Companies (f)*	$d = \frac{m-45}{10}$	*fd*
10-20	15	5	– 3	– 15
20-30	25	17	– 2	– 34
30-40	35	26	– 1	– 26
40-50	45	12	0	0
50-60	55	23	1	23
60-70	65	17	2	34
		N = 100		Σfd = – 18

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i = 45 - \frac{18}{100} \times 10 = 45 - 1.8 = 43.2.$$

EXAMPLE 23.9 : Calculate A.M. from the following data :

Marks	*No. of Students*
Above 0	150
Above 10	140
Above 20	100
Above 30	80
Above 40	80
Above 50	70
Above 60	30
Above 70	14

SOLUTION : Since we are given cumulative frequencies, we first find simple frequencies.

Calculation of A.M.

Marks	*Mid-values (m)*	*No. of students (f)*	$d = \frac{m-35}{10}$	*fd*
0-10	5	10	– 3	– 30
10-20	15	40	– 2	– 80
20-30	25	20	– 1	– 20
30-40	**A = 35**	0	0	0
40-50	45	10	1	10
50-60	55	40	2	80
60-70	65	16	3	48
70-80	75	14	4	56
		N = 150		Σfd = 64

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i = 35 + \frac{64}{150} \times 10 = 39.27.$$

EXAMPLE 23.10 : Find the missing frequency from the following data if arithmetic mean is 46.33.

Class Interval :	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency :	4	12	40	–	27	13	9	4

SOLUTION : Let the missing frequency be x.

Calculation of Missing Frequency

Class Interval	*Mid-values (m)*	*Frequency (f)*	*fm*
10-20	15	4	60
20-30	25	12	300
30-40	35	40	1400
40-50	45	x	45x
50-60	55	27	1485
60-70	65	13	845
70-80	75	9	675
80-90	85	4	340
		N = 109 + X	5105 + 45x

$$\bar{X} = \frac{\Sigma fm}{N}$$

$$46.33 = \frac{5105+45x}{109+x}$$

$$\Rightarrow \quad (46.33)(109+x) = 5105 \times 45x$$

$$5049.97 + 46.33x = 5105 + 45x$$

$$1.33x = 55.03$$

$$x = 41.37 \sim 41$$

$\therefore$ Missing frequency is 41.

EXAMPLE 23.11. The mean of 10 observations were found to be 1670. Later on it was discovered that one observation 1950 was wrongly taken as 2000. Find the correct mean corresponding to correct score.

SOLUTION : Given : n = 10, $\bar{X}$ = 1670.

We know that $\quad \bar{X} = \frac{\Sigma X}{n}$

$$1670 = \frac{\Sigma X}{10}$$

$$\Sigma X = 16700$$

$$\text{Correct } \Sigma X = 16700 - \text{wrong item} + \text{correct item}$$

$$= 16700 - 2000 + 1950 = 16650$$

$$\text{Correct } \bar{X} = \frac{\text{Corr } \Sigma X}{n} = \frac{16650}{10} = 1665.$$

EXAMPLE 23.12 : The mean of 100 observations was found to be 40. Later on it was discovered that two items were wrongly taken as 30 and 27 instead of 3 and 72. Find correct mean.

SOLUTION : We know that

$$\bar{X} = \frac{\Sigma X}{}$$

$$40 = \frac{\Sigma X}{100}$$

$$\therefore \quad \Sigma X = 4000$$

$$\text{Correct } \Sigma X = 4000 - \text{wrong item} + \text{correct items}$$
$$= 4000 - 30 - 27 + 3 + 72 = 4018$$

$$\therefore \quad \text{Correct } \overline{X} = \frac{4018}{100} = 40.18.$$

23.6 PROPERTIES OF ARITHMETIC MEAN

1. The sum of the deviations of the items from the arithmetic mean is always zero, *i.e.*, $\Sigma(X - \overline{X}) = 0$.
2. The sum of squares of deviations of the items from the arithmetic mean is minimum

$$\Sigma(X - \overline{X})^2 \text{ is minimum.}$$

3. **Combined Arithmetic Mean.** If we have two related groups such that N_1 and N_2 are the number of observations in first and second groups respectively and $\overline{X}_1$ and $\overline{X}_2$ are their arithmetic means respectively. Then their combined A.M. is given by $\overline{X}_{12}$

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

23.7 MERITS AND DEMERITS OF ARITHMETIC MEAN

Merits of Arithmetic Mean

1. It is simple and easy to understand.
2. It takes into consideration each and every item of the series.
3. It is rigidly defined in the sense that everyone will get the same answer when apply the formula of average.
4. It is capable of further algebraic treatment.
5. It does not depend on the position in the series.
6. It does not fluctuate with sampling.

Demerits of Arithmetic Mean

1. It sometimes introduces error in case of open end classes. In such cases median and mode are used.
2. Mean is a good measure of central tendency mainly in normal distribution. In case of U-shaped distribution, the mean is not likely to serve a useful purpose.
3. It is unduly affected by extreme observations *i.e.*, by very large and very small items.

EXAMPLE 23.13. The mean annual salaries paid to 10 male employees of a factory was Rs 2000 while the mean annual salaries paid to the 20 female employees of the same factory was Rs 1500. Determine the overall average salaries paid to all employees of the factory.

SOLUTION : Let N_1 represents no. of male employees and N_2 represents the no. of female employees.

Then $N_1 = 10, N_2 = 20, \overline{X}_1 = 2000, \overline{X}_2 = 1500$.

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

23.16 RELATION BETWEEN THE AVERAGES

In any distribution, the following relationship among averages exists

$$\text{A.M.} \geq \text{G.M.} \geq \text{H.M.}$$

23.17 MEDIAN

Definition. The median refers to the middle value in a distribution. It is a positional average.

FORMULA

1. Case of Individual Observations—

Let $X_1, X_2, \ldots, X_n$ be n observations.

(a) Arrange the given data in ascending or descending order of their magnitudes.

(b) If n is even then

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

If n is odd then

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

Case of discrete series :

1. Arrange the data in ascending or descending order of magnitude *i.e.*, with respect to X.
2. Find out the cumulative frequencies.
3. Apply the formula : Med = Size of $\left(\frac{n+1}{2}\right)^{th}$ item.
4. Now look at the c.f. column and find that total which is either equal to $\left(\frac{n+1}{2}\right)^{th}$ or next higher to that and determine the value of the variable corresponding to this. This gives the value of median.

Case of continuous series :

1. Calculate cumulative frequencies.
2. Find $\frac{N}{2}$ and look at the c.f. column and find that total which is either equal to $\frac{N}{2}$ or next higher to that and determine the class interval corresponding to this. This is the **median class.**
3. Apply the formula

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

where l = lower limit of median class

f = frequency of median class

C = c.f. of class preceding to the median class

h = class size.

Median in case of unequal class intervals :

When the class intervals are unequal, there is no need to adjust the frequencies to make class intervals equal. The same formula of median in case of continuous series can be applied.

Mathematical Property of Median

The sum of deviations of the item from median ignoring signs is the least

i.e., $\Sigma \mid X - \text{Med} \mid = \text{least}.$

23.18 MERITS OF DEMERITS OF MEDIAN

Merits of Median

1. It is very useful in case of **open-end classes** since it is a positional average. Open end classes are those classes in which lower limit of first class interval and the upper limit of the last class interval are unknown.
2. It is easy to compute in case of unequal class intervals.
3. It is not strongly affected by extreme observations.
4. It can be determined graphically whereas the value of mean cannot be graphically determined.

Demerits of Median

1. It is necessary to arrange the data but other averages do not need any arrangement.
2. It does not take into account each and every item of the series.
3. It is not capable of further algebraic treatment.
4. It is affected by fluctuations in sampling.
5. It is erratic if the number of items is less.

23.19 QUARTILES, DECILES AND PERCENTILES

These are also the positional measures like median. Quartiles are those values of the variate which divides the frequency into four equal parts. Consequently there are only 3 quartiles—Q_1, Q_2, Q_3. Q_1 is called lower quartile, Q_3 is called upper quartile, Q_2 is median.

Deciles divides the total frequency into 10 equal parts and percentiles into 100 equal parts. Corresponding we have 9 deciles $D_1, D_2, \ldots, D_9$ and 99 percentiles $P_1, P_2, P_3, \ldots, P_{99}$.

COMPUTATION OF QUARTILES, DECILES AND PERCENTILES

Case of Individual Observations and Discrete Series :

(a) Arrange the data in ascending or descending order of magnitude.

(b) In case of discrete series calculate cumulative frequencies.

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

where r = 1, 2, 3

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

where r = 1, 2, 3, ..., 9

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

where r = 1, 2, 3, ..., 99

In case of discrete series, look at the c.f. column for the above values or next higher to that. The value of the variable corresponding to this gives the required value.

Case of Continuous Series :

1. Locate the Quartile, decile or percentile class as per requirement by finding

$$Q_r = \left(\frac{rn}{4}\right)^{th} \text{ item}$$

$$D_r = \left(\frac{rn}{10}\right)^{th} \text{ item}$$

$$P_r = \left(\frac{rn}{100}\right)^{th} \text{ item.}$$

2. Now look at the c.f. column for above values or next higher to that to determine the class. Then apply the following formula :

$$Q_r = l + \frac{\frac{rn}{4} - C}{f} \times h, \quad \text{where } r = 1, 2, 3$$

$$D_r = l + \frac{\frac{rn}{10} - C}{f} \times h, \quad \text{where } r = 1, 2, 3, ..., 9$$

$$P_r = l + \frac{\frac{rn}{100} - C}{f} \times h, \quad \text{where } r = 1, 2, 3, ..., 99$$

l = lower limit of the quartile, decile, percentile class as the case may be

C = c.f. of the class preceding to the quartile (decile, percentile) class

f = frequency of quartile class (decile, percentile) class

h = class size.

Q_r = r^{th} quartile

Put r = 1 for finding Q_1

r = 2 for finding Q_2

r = 3 for finding Q_3

D_r = r^{th} decile

P_r = r^{th} percentile.

EXAMPLES

EXAMPLE 23.37. Compute median income from the following data of the income of 5 workers

Income (Rs) : 1200, 1000, 900, 1400, 1500

SOLUTION : Arrange the data in ascending order of magnitude 900, 1000, 1200, 1400, 1500

n = 5 (odd)

$$\therefore \quad \text{Med.} = \text{Size of} \left(\frac{n+1}{2}\right)^{th} \text{ item} = \text{Size of } 3^{rd} \text{ item} = 1200$$

EXAMPLE 23.38. Obtain the median from the following data :

100, 105, 90, 95, 70, 102.

SOLUTION : This is a case of individual observations. Arranging in ascending order 70, 90, 95, 100. 102, 105

n = 6 (even)

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

$$= \frac{\text{Size of } 3^{\text{rd}} \text{ item} + \text{Size of } 4^{\text{th}} \text{ item}}{2} = \frac{95+100}{2} = \frac{195}{2} = 97.5.$$

EXAMPLE 23.39. Calculate median from the following series :

X :	10	11	12	13	14
f :	3	12	18	12	3

SOLUTION : Clearly, it is a discrete series

Calculation of Median

X	*f*	*c.f*
10	3	3
11	12	15
12	18	33
13	12	45
14	3	48
	N = 48	

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

$$= \text{Size of } \left(\frac{48+1}{2}\right)^{\text{th}} \text{ item} = \text{Size of } 24.5^{\text{th}} \text{ item}$$

Looking the value 24.5 or next higher to it in c.f. column, we get median = 12.

EXAMPLE 23.40. Find the median of the following data :

Variable :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	7	12	18	25	16	14	8

SOLUTION : **Calculation of Median**

Variable	*Frequency* *f*	*c.f.*
0-10	7	7
10-20	12	19
20-30	18	37
30-40	25	62
40-50	16	78
50-60	14	92
60-70	8	100
	N = 100	

$$\text{Med.} = \frac{N}{2} = \frac{100}{2} = 50^{\text{th}} \text{ item}$$

Looking this value in c.f. column, we get the median class is 30-40

$$\therefore \qquad \text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

Here $l = 30,\ C = 37,\ f = 25,\ h = 10$

$$\therefore \qquad \text{Med.} = 30 + \frac{50-37}{25} \times 10 = 35.2.$$

EXAMPLE 23.41. Calculate median from the following data :

Mid-value :	15	25	35	45	55	65	75	85
Frequency :	5	9	13	21	20	15	8	3

SOLUTION : Since we are given the mid-values, we first determine the class intervals.

Here, the difference between the two consecutive mid-values is 10.

∴ Class size is 10

∴ First class interval is 10-20 as mid-value is 15.

Calculation of Median

Class Intervals	*Frequency* *f*	*c.f.*
10-20	5	5
20-30	9	14
30-40	13	27
40-50	21	48
50-60	20	68
60-70	15	83
70-80	8	91
80-90	3	94
	N = 94	

$$\text{Med.} = \text{Size of} \left(\frac{n}{2}\right)^{\text{th}} \text{item} = \frac{94}{2} = 47^{\text{th}} \text{ item.}$$

∴ Median class is 40-50

∴ $l = 40,\ f = 21,\ C = 27,\ h = 10$

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

$$= 40 + \frac{47-27}{21} \times 10 = 49.524.$$

EXAMPLE 23.42. Obtain median from the following data :

Marks :	40-45	35-40	30-35	25-30	20-25	15-20	10-15	5-10	0-5
No. of students :	2	3	6	10	52	117	241	195	29

SOLUTION : Arranging the data in ascending order :

Marks	*No. of students* f	*C.f.*
0-5	29	29
5-10	195	224
10-15	241	465
15-20	117	582
20-25	52	634
25-30	10	644
30-35	6	650
35-40	3	653
40-45	2	655
	N = 655	

$$\text{Med.} = \text{Size of } \left(\frac{N}{2}\right)^{th} \text{ item} = \frac{655}{2} = 327.5^{th} \text{ item}$$

$\Rightarrow$ Median class is 10-15

$\therefore\ l = 10,\ h = 5,\ f = 241,\ C = 224$

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h = 10 + \frac{327.5 - 224}{241} \times 5 = 10 + 2.147 = 12.147$$

$\therefore$ Med. = 12.147.

EXAMPLE 23.43. Find the missing frequency from the following distribution if median is 35 and N = 170.

Variable :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	10	20	—	40	—	25	15

[DU. B.Com (H

SOLUTION : Let the missing frequencies be x and y.

Then $10 + 20 + x + 40 + y + 25 + 15 = 170$

$\Rightarrow \quad x + y = 170 - 110$

$\Rightarrow \quad x + y = 60 \quad \text{or} \quad y = 60 - x$

Determination of Missing Frequency

Variable	*Frequency (f)*	*c.f.*
0-10	10	10
10-20	20	30
20-30	x	30 + x
30-40	40	70 + x
40-50	y	70 + x + y
50-60	25	95 + x + y
60-70	15	110 + x + y
	N = 170	

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

Since median is 35.

∴ Median lies is the class 30-40

∴ $l = 30,\ h = 10,\ f = 40,\ C = 30 + x$

$$\therefore \quad 35 = 30 + \frac{\frac{170}{2} - (30 + x)}{40} \times 10$$

$$35 - 30 = \frac{85 - 30 - x}{4}$$

$$20 = 55 - x$$

$$x = 55 - 20 = 35 \qquad \therefore \quad x = 35$$

$$y = 60 - x = 60 - 35 = 25$$

∴ Missing frequencies are 35, 25.

EXAMPLE 23.44. From the following data, compute median :

Marks	*No. of students*
0-10	8
10-20	10
20-40	22
40-60	25
60-80	10
80-100	5

SOLUTION : In this question, the class intervals are unequal but we know that while calculating median there is no need to make class intervals equal by adjusting frequencies.

Calculation of Median

Marks	*No. of students* *f*	*c.f.*
0-10	8	8
10-20	10	18
20-40	22	40
40-60	25	65
60-80	10	75
80-100	5	80
	N = 80	

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

∴ Median class = 20-40

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

Here $l = 20$, $c = 18$ $f = 22$ $h = 20$

$$\therefore \quad \text{Med.} = 20 + \frac{40 - 18}{22} \times 20 = 20 + 20 = 40$$

EXAMPLE 23.45. Obtain median from the following data :

Marks	*No. of students*
Less than 5	4
Less than 10	10
Less than 15	20
Less than 20	30
Less than 25	55
Less than 30	77
Less than 35	95
Less than 40	100

SOLUTION : Since we are given c.f.

$\therefore$ We first find simple frequencies.

Calculation of Median

Marks	*No. of students (f)*	*c.f.*
0-5	4	4
5-10	6	10
10-15	10	20
15-20	10	30
20-25	25	55
25-30	22	77
30-35	18	95
35-40	5	100
	N = 100	

$$\text{Med.} = \text{size of } \left(\frac{N}{2}\right)^{th} \text{ item} = \frac{100}{2} = 50^{th} \text{ item}$$

$\therefore$ Median class is $20-25$

$\therefore$ $l=20, \quad h=5, \quad f=25, \quad c=30$

$$\text{Med.} = l + \frac{\frac{n}{2}-C}{f} \times h = 20 + \frac{50-30}{25} \times 5 = 24.$$

EXAMPLE 23.46. Obtain median from the following data :

Profit (Rs in lakhs)	*No. of companies*
More than 10	100
More than 20	94
More than 30	86
More than 40	74
More than 50	56
More than 60	31
More than 70	15
More than 80	7
More than 90	2

SOLUTION : Since we are given more than c.f., we first find simple frequencies and then less than c.f. Please note that formula for median is based on less than c.f.

Calculation of Median

Profit (Rs in lakhs)	*f*	*c.f.*
10-20	6	6
20-30	8	14
30-40	12	26
40-50	18	44
50-60	25	69
60-70	16	85
70-80	8	93
80-90	5	98
90-100	2	100
	N = 100	

$$\text{Med.} = \text{Size of} \left(\frac{N}{2}\right)^{\text{th}} \text{item} = \frac{100}{2} = 50^{\text{th}} \text{ item}$$

$\therefore$ Median class is $50-60$

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

Here $l = 50$, $h = 10$, $f = 25$, $c = 44$

$$\therefore \quad \text{Med.} = 50 + \frac{50-44}{25} \times 10 = 50 + 2.4 = 52.4$$

EXAMPLE 23.47. Calculate median from the data given below :

Class Interval :	5-9	10-14	15-19	20-24	25-29	30-34	35-39
Frequency :	8	15	18	30	16	12	6

SOLUTION : We first make the class intervals continuous by subtracting 0.5 from the lower limit of each class interval and adding 0.5 to upper limit of each class interval.

Calculation of Median

Class Intervals	*Frequency (f)*	*c.f.*
4.5-9.5	8	8
9.5-145	15	23
14.5-19.5	18	41
19.5-24.5	30	71
24.5-29.5	16	87
29.5-34.5	12	99
34.5-39.5	6	105
	N = 105	

$$\text{Med.} = \text{Size of}\left(\frac{N}{2}\right)^{th} \text{item} = \frac{105}{2} = 52.5^{th} \text{ item}$$

∴ Med. class is 19.5-24.5

$l = 19.5, \quad h = 5, \quad c = 41, \quad f = 30$

$$\therefore \quad \text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h = 19.5 + \frac{52.5 - 41}{30} \times 5 = 19.5 + 1.92 = 21.42$$

EXAMPLE 23.48 : Calculate upper and lower quartile from the following data :

Variable :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	10	20	35	40	25	25	15

SOLUTION : It is a continuous series :

Calculation of Quartiles

Variable	*Frequency (f)*	*Cumulative frequency c.f.*
0-10	10	10
10-20	20	30
20-30	35	65
30-40	40	105
40-50	25	130
50-60	25	155
60-70	15	170
	$N = 170$	

We know that $\quad Q_r = \left(\frac{rn}{4}\right)^{th}$ item

For lower quartile Q_1, put $r = 1$

$$Q_1 = \left(\frac{n}{4}\right)^{th} \text{item} = \frac{170}{4} = 42.5$$

⇒ Lower quartile class is 20-30

Now $$Q_1 = l + \left(\frac{\frac{n}{4} - C}{f}\right) \times h$$

Here $l = 20, \quad h = 10, \quad c = 30, \quad f = 35$

$$\therefore \quad Q_1 = 20 + \frac{42.5 - 30}{35} \times 10 = 20 + 3.57 = 23.57$$

Upper Quartile

For Q_3, putting $r = 3$, we get

$$Q_3 = \text{size of}\left(\frac{3n}{4}\right)^{th} \text{item}$$

$$Q_3 = \frac{3 \times 170}{4} = 127.5$$

$\therefore$ Upper quartile class is 40-50

$\therefore$ $l=40,\quad h=10,\quad f=25,\quad c=105$

$$Q_3 = l + \frac{\frac{3n}{4} - C}{f} \times h = 40 + \frac{127.5-105}{25} \times 10 = 40 + 9 = 49.$$

EXAMPLE 23.49 : Calculate D_5, P_{60} from the following table :

Salary (more than) :	0	500	1000	1500	2000	2500	3000	3500
No. of persons :	250	200	120	80	55	30	15	5

SOLUTION : Since we are given more than c.f., we first calculate simple frequencies and then less than c.f.

Salary	*No. of persons*	*c.f.*
0-500	50	50
500-1000	80	130
1000-1500	40	170
1500-2000	25	195
2000-2500	25	220
2500-3000	15	235
3000-3500	10	245
3500-4000	5	250
	N = 250	

We know that $D_r = \text{Size of } \frac{rn}{10}^{th}$ item

For D_5, put $r = 5$

$$D_5 = \left(\frac{5n}{10}\right)^{th} \text{ item} = \frac{5 \times 250}{10} = 125^{th} \text{ item}$$

$\therefore$ Decile class is 500-1000

Here $l=500,\quad f=80,\quad c=50,\quad h=500$

$$D_5 = l + \frac{\frac{5n}{10} - C}{f} \times h$$

$$= 500 + \frac{125-50}{80} \times 500 = 500 + 468.75 = 968.75$$

Percentiles $P_r = \text{Size of } \frac{rn}{100}^{th}$ item

For P_{60}, put $r = 60$

$$P_{60} = \frac{60 \times 250}{100} = 150^{th} \text{ item}$$

$\therefore$ Percentile class is 1000-1500

$l=1000,\quad h=500,\quad f=40,\quad c=130$

$$P_{60} = l + \frac{\frac{60n}{100} - C}{f} \times h$$

$$= 1000 + \frac{150-130}{40} \times 500 = 1000 + 250 = 1250$$

EXAMPLE 23.50 : Find the upper and lower quartiles from the data given below :

Roll no. :	1	2	3	4	5	6	7
Marks :	20	28	40	12	30	15	50

SOLUTION : Clearly, data is in the form of individual observations. So we first arrange the data in ascending order :

12, 15, 20, 20, 30, 40, 50

$$Q_1 = \text{Size of } \frac{n+1}{4}^{\text{th}} \text{ item} = \frac{7+1}{4} = 2^{\text{nd}} \text{ item}$$

$\therefore$ $Q_1 = 15$

For Q_3,

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

$$= \frac{3(7+1)}{4} \text{ item} = 6^{\text{th}} \text{ item} = 40$$

$\therefore$ $Q_3 = 40.$

EXAMPLE 23.51 : Find Q_1 and Q_3 from the data given below :

X :	10	60	20	50	40	30
f :	4	2	7	7	8	15

SOLUTION : Clearly, it is a discrete series, so we first arrange the data in ascending order with respect to X.

Calculation of Quartiles

X	f	$C.f.$
10	4	4
20	7	11
30	15	26
40	8	34
50	7	41
60	2	43
	N = 43	

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

$$= \left(\frac{43+1}{4}\right)^{\text{th}} \text{ item} = 11^{\text{th}} \text{ item}$$

$\therefore$ $Q_1 = 20$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4}^{\text{th}} \text{ item} = \frac{3(43+1)}{4}^{\text{th}} \text{ item} = 33^{\text{rd}} \text{ item}$$

$\therefore$ $Q_3 = 40$

23.20 MODE

Definition : Mode is that value of the variate which occurs with the maximum frequency. Mode can also be defined as that value about which the items are most closely concentrated. It is the value which has the greatest frequency density in its immediate neighbourhood.

FORMULAE FOR DETERMINING MODE

1. In Case of Individual Observations : Mode can be determined by counting the number of times, the various values repeat themselves and then the value occurring maximum number of times is the modal value.

2. In Discrete Series : In discrete series, mode can be determined by the following two methods :

(a) *By inspection* : In this method, we look at that value of the variable around which the items are most closely concentrated. Generally, if the distribution is fairly regular, we look at the highest frequency and find the value of the variable corresponding to that. Such value is the modal value.

(b) *By grouping* : It consists of a grouping table having six columns :

(i) Column 1 is the column of frequency given in the question. Put a circle on the maximum frequency.

(ii) In column 2, we add the frequencies in group of two's and then put a circle on the highest total.

(iii) In column 3, leave the first frequency, and then group the frequency in two's. Again put a circle on the highest total.

(iv) In column 4, group the frequencies in three's and again put a circle on the highest total.

(v) In column 5, leave the first frequency and group the frequencies in three's and put the circle on the highest total.

(vi) In column 6, leave the first two frequencies and then group the remaining in three's and mark the highest total.

Now we prepare the analysis table. For this, write column number on L.H.S. and the various values of mode on R.H.S. The values against which the highest frequencies are circled in grouping table are entered by putting 1 in the relevant box corresponding to the values they represent. Now the value having maximum no. of 1's is the mode.

3. Mode in Continuous Series :

(i) Determine the modal class either by inspection or by grouping. If the distribution is not fairly regular, use grouping method.

(ii) Apply the formula :

$$M_0 = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

where $\Delta_1 = |f_m - f_1|$

$\Delta_2 = |f_m - f_2|$

l = lower limit of modal class

h = class size

M_0 = mode

f_m = frequency of modal class

f_1 = frequency of class preceding to the modal class

f_2 = frequency of class succeeding to the modal class

4. Mode in Case of Bimodal Series : Sometimes, there are two values which occur with equal frequencies. The distribution is then called **bimodal** and mode is said to be **ill-defined**. In such cases we calculate **empirical** mode given by

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}.$$

5. Mode in Moderately Skewed Distribution : In moderately skewed distribution or asymmetrical distribution, we calculate mode as follows.

$$\text{Mode} = 3\ \text{Med.} - 2\ \text{Mean}.$$

6. Mode when Class Intervals are Unequal : If class intervals are unequal, then there is a need to make them equal by adjusting the frequencies on the assumption that they are equally distributed throughout the class.

23.21 GRAPHICAL DETERMINATION OF MODE

Steps for determining the mode graphically are :

1. Draw a histogram of the given data.
2. The bar with the highest peak is the modal class bar. Draw two lines diagonally in the inside of the modal class bar starting from each upper corner of the bar to the upper corner of the adjacent bars.
3. Draw a perpendicular from the intersection of the two diagonal lines to the X-axis. Then the foot of the perpendicular will be the mode.

23.22 MERITS AND DEMERITS OF MODE

Merits of Mode

1. Mode is that value which represents the whole distribution.
2. Like median, mode is not unduly affected by extreme values.
3. It can be determined graphically.
4. It can be used to describe qualitative phenomenon.

Demerits of Mode

1. Sometimes the mode is ill-defined. The value of mode cannot always be determined.
2. Mode is not capable of further algebraic treatment.
3. It is not rigidly defined. There are several formulae for determining mode, all of which may give different answers.
4. It is not based on each and every item of the series.

Usefulness of Mode : Mode is generally determined in highly skewed or asymmetrical distributions. It is employed when the most typical value of a distribution is desired.

EXAMPLE 23.52. Calculate mode from the following data :

10, 27, 24, 10, 10, 12, 13, 24, 39, 27, 10.

SOLUTION : Data is given in the form of individual observation.

∴ We use inspection method :

Calculation of Mode

Values	*No. of times it occurs*
10	4
12	1
13	1
24	2
27	2
39	1
	Total = 10

Since the observation 10 is occurring maximum no. of times. ∴ Mode is 10.

EXAMPLE 23.53. Calculate mode for the following data :

Marks :	0-20	20-40	40-60	60-80	80-100	100-120	120-140
No. of students :	4	26	22	10	9	6	3

SOLUTION : It is a continuous series. Since the distribution is not fairly regular, we determine the modal class by grouping method.

Grouping Table

Marks	*f* *Col. I*	*Col. II*	*Col. III*	*Col. IV*	*Col. V*	*Col. VI*
0-20	4	30		52		
20-40	26		48		58	
40-60	22	32				41
60-80	10		19	25		
80-100	9	15			18	
100-120	6		9			
120-140	3					

Analysis Table

Col. no.	*0-20*	*20-40*	*40-60*	*60-80*	*80-100*	*100-120*	*120-140*
I		1					
II			1	1			
III		1	1				
IV	1	1	1				
V		1	1	1			
VI			1	1	1		
Total	1	4	5	3	1	0	0

From the analysis table, it is clear that modal class is 40-60

$\therefore \quad l = 40, \quad f_m = 22, \quad f_1 = 26, \quad f_2 = 10, \quad h = 20$

$$\Delta_1 = |f_m - f_1| = |22 - 26| = 4$$

$$\Delta_2 = |f_m - f_2| = |22 - 10| = 12$$

Now
$$\text{Mode} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

$$= 40 + \frac{4}{4+12} \times 20 = 40 + 5 = 45$$

EXAMPLE 23.54. Obtain mode from the following data :

Wages :	100	200	110	120	105	90	95	125
No. of workers :	8	10	15	25	40	20	15	7

SOLUTION : Clearly, it is a discrete series and the frequency distribution is fairly regular.

$\therefore$ By inspection, mode = 105 corresponding to the highest frequency 40.

EXAMPLE 23.55. Calculate mode from the following data :

Income :	15-24	25-34	35-44	45-54	55-64	65-74
No. of workers :	8	10	15	25	40	20

SOLUTION : We first make the class intervals continuous :

Calculation of Mode

Income	*No. of employees*
14.5-24.5	8
24.5-34.5	10
34.5-44.5	15
44.5-54.5	25
54.5-64.5	40
64.5-74.5	20

Since the maximum frequency is 40, the modal class is 54.5-64.5

$\therefore \quad l = 54.5, \quad f_m = 40, \quad f_1 = 25, \quad f_2 = 20, \quad h = 10$

$$\Delta_1 = |f_m - f_1| = 40 - 25 = 15$$

$$\Delta_2 = |f_m - f_2| = 40 - 20 = 20$$

$$\text{Mode} = l + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

$$= 54.5 + \frac{15}{15+20} \times 10$$

$$= 54.5 + 4.28 = 58.78$$

EXAMPLE 23.56. Calculate mode from the following data :

Variable :	0-10	10-20	20-40	40-50	50-70
Frequency :	5	12	40	32	28

SOLUTION : Since the class intervals are unequal, we first make them equal by adjusting the frequencies.

Variable	*Frequency*
0-10	5
10-20	12
20-30	20
30-40	20
40-50	32
50-60	14
60-70	14

By inspection modal class is 40-50

$\therefore \quad l=40, \quad f_m=32, \quad f_1=20, \quad f_2=14, \quad h=10$

$$\Delta_1 = |f_m - f_1| = 32-20 = 12$$

$$\Delta_2 = |f_m - f_2| = 32-14 = 18$$

$$\text{Mode} = l + \frac{\Delta_1}{\Delta_1+\Delta_2} \times h$$

$$= 40 + \frac{12}{12+18} \times 10 = 40+4 = 44$$

EXAMPLE 23.57. In an asymmetrical distribution, mean = 17.9, mode = 26. Find median.

SOLUTION : We know that

$$\text{Mode} = 3\text{ Med.} - 2\text{ Mean}$$

$$26 = 3\text{ Med.} - 2(17.9)$$

$$26 = 3\text{ Med.} - 35.8$$

$$\Rightarrow \quad 3\text{ Med.} = 26 + 35.8$$

$$\text{Med.} = 20.6.$$

EXAMPLE 23.58. The following data gives the daily wages of 122 workers of a factory. Determine the modal wage.

Wages (Rs.) :	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
No. of workers :	4	6	20	32	33	17	8	2

[*B.Com (H) DU*]

SOLUTION : By inspection, it is difficult to determine the modal class.

$\therefore$ We use grouping method :

Grouping Table

Marks	*f* *Col. I*	*Col. II*	*Col. III*	*Col. IV*	*Col. V*	*Col. VI*
100-110	4	10		30		
110-120	6		26		58	
120-130	20	52				85
130-140	32		65	82		
140-150	33	50			58	
150-160	17		25			27
160-170	8	10				
170-180	2					

Analysis Table

Col. \ Wages	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
I					1			
II			1	1				
III				1	1			
IV				1	1	1		
V		1	1	1	1	1	1	
VI			1	1	1			
Total	0	1	3	5	5	2	1	0

From the analysis table, it is clear that it is a bimodal series. So Mode = 3 Med. – 2 Mean.

Calculation of Median and Mean

Wages	m	f	c.f.	$d = \frac{m-135}{10}$	fd
100-110	105	4	4	–3	–12
110-120	115	6	10	–2	–12
120-130	125	20	30	–1	–20
130-140	135 = A	32	62	0	0
140-150	145	33	95	1	33
150-160	155	17	112	2	34
160-170	165	8	120	3	24
170-180	175	2	122	4	28
		N = 122			$\Sigma fd = 55$

Mean $$\overline{X} = A + \frac{\Sigma fd}{N} \times i$$

$A = 135, \quad \Sigma fd = 55, \quad N = 122, \quad i = 10$

$$\therefore \quad \overline{X} = 135 + \frac{55}{122} \times 10 = 139.51$$

Median $$\text{Median} = \text{Size of}\left(\frac{N}{2}\right)^{\text{th}} \text{item} = \frac{122}{2} = 61^{\text{th}} \text{ item}$$

$\therefore$ Median class is 130-140

$$\text{Med.} = l + \frac{\frac{n}{2} - C}{f} \times h$$

$l = 130, \quad C = 30, \quad f = 32, \quad h = 10$

$$\text{Med.} = 130 + \frac{61-30}{32} \times 10 = 139.69$$

Mode $$\text{Mode} = 3\,\text{Med.} - 2\,\text{Mean} = 3(139.69) - 2(139.51) = 140.05.$$

Chapter 24

Measures of Dispersion

24.1 INTRODUCTION

The various measures of central tendency discussed in the previous chapter gives a single figure that represents the entire distribution. But there may be two or more distributions having the same average value but have wide disparities in their formation. This creates a need of studying variation of the observations. Measures of dispersion helps in studying this important characteristic of a distribution.

24.2 MEANING OF DISPERSION

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

—Spiegel

24.3 PROPERTIES OF A GOOD MEASURE OF DISPERSION

Following are the properties of a good measure of dispersion :

1. It should be easy to compute and simple to understand.
2. It should be rigidly defined.
3. It should be based on each and every item of the distribution.
4. It should be capable of further algebraic treatment.
5. It should not fluctuate with sampling.
6. It should not be unduly affected by extreme observations.

24.4 METHODS OF STUDYING VARIATION

Following are the methods of studying variation :

1. Range
2. Interquartile Range and Quartile Deviation
3. Mean Deviation
4. Standard Deviation and Variance.

24.5 TYPES OF MEASURES OF VARIATION

1. Absolute measure of variation
2. Relative measure of variation.

Absolute Measures of Variation

These are those measures of variation or dispersion which are expressed in the same statistical unit in which the original data are given such as rupees, kilograms, tonnes, pounds, etc. *For example,* range and interquartile range are absolute measures of variation.

Relative Measures of Variation

These are the ratio of a measure of absolute dispersion to an appropriate average. It is a pure number. It is also called a coefficient of dispersion. It has no units of measurement, *e.g.*, coefficient of variation, coefficient of quartile deviation.

24.6 RANGE

It is the simplest method of studying dispersion

$$\text{Range} = L - S$$

where L = largest item

S = smallest item

$$\text{Coefficient of Range} = \frac{L-S}{L+S}$$

It is relative measure corresponding to range.

Range in Case of Continuous Series

To determine range, find the difference between the upper limit of the last class interval and the lower limit of lowest class interval. Range can also be determined by finding the difference of the midpoints of the lowest class and highest class.

Merits and Demerits of Range

Merits :

1. It is the simplest method of studying variation.
2. It is easy to compute.
3. It is a quick method and takes less time for computation.

Demerits :

1. It is subjected to fluctuations of sampling.
2. It is affected by extreme observations.
3. It is not based on each and every item of the distribution.
4. It cannot be computed in open-end distributions.

24.7 INTERQUARTILE RANGE (QUARTILE DEVIATION)

It refers to the range between the quartiles :

$$\text{Interquartile range} = Q_3 - Q_1$$

$$\therefore \text{ Semi-Interquartile Range or Quartile range} = \frac{Q_3 - Q_1}{2}$$

Quartile Deviation gives the average amount by which the two quartiles differ from the median. It is an absolute measure of dispersion. The relative measure of dispersion is coefficient of Quartile Deviation given as

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}.$$

Merits and Demerits of Quartile Deviation

Merits :

1. It is used in open-end distributions as it is a positional measure of dispersion.
2. It is not affected by extreme observations and is therefore used in badly skewed distributions.

Demerits of Quartile Deviation

1. It does not depend on each and every item of the distribution.
2. It is not capable of further algebraic treatment.
3. It is subjected to fluctuations in sampling.

24.8 MEAN DEVIATION

It is also called average deviation. It is the mean of the deviations of item from median ignoring the signs.

Case of Individual Observations

$$\text{M.D.} = \frac{\Sigma|D|}{N}$$

where $|D| = |X - \text{Med.}|$

$N =$ no. of observations

Case of Frequency Distribution (Discrete and Continuous Series)

$$\text{M.D.} = \frac{\Sigma f|D|}{N}$$

where $N = \Sigma f =$ total no. of observations

$|D| = |X - \text{Med}|$

$f =$ frequency.

Coefficient of Mean Deviation

It is a relative measure corresponding to mean deviation

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}}.$$

Note : If arithmetic mean is used as an average for taking deviations then in case of individual observations $\text{M.D.} = \frac{\Sigma|D|}{N}$ where $|D| = |X - \text{mean}|$. Same will be in case of frequency distribution. Moreover, coefficient of $\text{M.D.} = \frac{\text{M.D.}}{\text{Mean}}$.

Merits and Demerits of Mean Deviation

Merits :

1. It is simple to understand.
2. It is easy to compute.
3. It is based on each and every item of the series.
4. It is less affected by the value of extreme observations.
5. It is used by National Bureau of Economic Research to forecast business cycle.

Demerits :

1. It is not capable of further algebraic treatment.
2. This method does not always give accurate results.
3. It has a limited use.
4. The formula of M.D. ignores the algebraic signs which makes this method non-algebraic.

24.9 STANDARD DEVIATION

It is defined as the square root of the mean of the squared deviations from the arithmetic mean. It is denoted by the Greek letter σ (sigma). It is an absolute measure of dispersion. A small S.D. means a high degree of stability or uniformity of the observation as well as homogeneity of series.

❑ FORMULA

1. Case of Individual Observations—

(*a*) $$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n}} \quad \text{or} \quad \sqrt{\frac{\Sigma X^2}{n} - (\overline{X})^2}$$

(*b*) $$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

where $d = X - A$ and A = assumed mean

2. Discrete series—

(*a*) ***Actual mean method***

$$\sigma = \sqrt{\frac{\Sigma f(X-\overline{X})^2}{n}}$$

where $n = \Sigma f$

(*b*) ***Assumed mean method :***

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$

where $d = X - A$ and $n = \Sigma f$

and A = assumed mean

(*c*) ***Step deviation method***

$$\sigma = \sqrt{\frac{\Sigma fu^2}{n} - \left(\frac{\Sigma fu}{n}\right)^2} \times h$$

where $u = \dfrac{X-A}{h}$

A = assumed mean

h = class size

3. Case of Continuous Series—

$$\sigma = \sqrt{\frac{\Sigma fu^2}{n} - \left(\frac{\Sigma fu}{n}\right)^2} \times h$$

where $u = \frac{m - A}{h}$

m = mid-value of each class

h = class size

A = assumed mean.

24.10 PROPERTIES OF STANDARD DEVIATION

1. Combined Standard Deviation : Consider the two groups have arithmetic means $\overline{X}_1$ and $\overline{X}_2$ respectively, standard deviations σ_1 and σ_2 respectively, N_1 and N_2 as no. of observations in first and second groups respectively, then their combined standard deviation denoted by σ_{12} 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 $d_1 = |\overline{X}_1 - \overline{X}_{12}|$ and

$d_2 = |\overline{X}_2 - \overline{X}_{12}|$

$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$ is combined arithmetic mean.

The above formula can be extended to find out combined standard deviation of three or more groups.

For example, $$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1 d_1^2 + N_2 d_2^2 + N_3 d_3^2}{N_1 + N_2 + N_3}}$$

where $d_1 = |\overline{X}_1 - \overline{X}_{123}|$

$d_2 = |\overline{X}_2 - \overline{X}_{123}|$

$d_3 = |\overline{X}_3 - \overline{X}_{123}|$

2. The sum of the squares of the deviation of items from the A.M. is least.

3. The standard deviation enables us to determine where the values of a frequency distribution are located.

For a normal or symmetrical distribution, if A.M. is μ and S.D. is σ then

$\mu \pm \sigma$ covers 68.27% of the items

$\mu \pm 2\sigma$ covers 95.4% of the items

$\mu \pm 3\sigma$ covers 99.7% of the items.

4. S.D. of first n natural numbers is $\sqrt{\frac{n^2 - 1}{12}}$.

24.11 MERITS AND DEMERITS OF STANDARD DEVIATION

Merits :

1. It is based on each and every item of the series.
2. It is less affected by fluctuation of sampling.

3. It is capable of further algebraic treatment. It is possible to calculate combined S.D. of two or more groups. It is used in measuring skewness, correlation etc.

Demerits :

1. It is difficult to compute.
2. It gives more weight to extreme items and less to those which are near the mean.

24.12 RELATION BETWEEN MEASURES OF DISPERSION

$$\text{Q.D.} = \frac{2}{3}\sigma \quad \text{and} \quad \text{M.D.} = \frac{4}{5}\sigma$$

24.13 VARIANCE

The square of standard deviation is called variance.

Variance $= \sigma^2$.

24.14 COEFFICIENT OF VARIATION

It is a relative measure of dispersion corresponding to standard deviation. It was given by Karl Pearson. It is used in such problems where we want to compare the variability of two or more series. The series for which coefficient of variation is greater is said to be more variable or conversely less stable, less uniform, less consistent, less homogeneous. It is denoted by C.V.

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

EXAMPLE 24.1 : The following are the sales of mobiles of Reliance in the whole week. Calculate range and its coefficient 200, 208, 190, 210, 320, 120, 180.

SOLUTION : Here L = 320 and S = 120

$$\therefore \quad \text{Range} = L - S = 320 - 120 = 200$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{200}{320+120} = \frac{200}{440} = 0.45$$

EXAMPLE 24.2 : Determine range and its coefficient from the following data :

Wages :	100-110	110-120	120-130	130-140	140-150	150-160
No. of workers :	10	12	15	16	9	8

SOLUTION : Range $= L - S = 160 - 100 = 60$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{60}{160+100} = \frac{60}{260} = 0.23$$

Aliter

$$\begin{aligned}\text{Range} &= \text{Midpoint of highest class} - \text{Midpoint of lowest class}\\ &= 155 - 105\\ &= 50\end{aligned}$$

$$\text{Coefficient of Range} = \frac{50}{155+105} = \frac{50}{260} = 0.192$$

Note : In above two methods, we get different answers but both the answers are correct.

EXAMPLE 24.3 : Find the Interquartile Range and Coefficient of Quartile Deviation from the data given below :

$$200,\ 210,\ 208,\ 160,\ 220,\ 250,\ 300$$

SOLUTION : Arranging the data in ascending order leads to
160, 200, 208, 210, 220, 250, 300

$$n = 7$$

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

$$= \frac{7+1}{4} = 2^{nd} \text{ item} = Q_1 = 200$$

$$Q_3 = \text{Size of } \frac{3(n+1)^{th}}{4} \text{ item} = 6^{th} \text{ item} = 250$$

$$\therefore \quad \text{Interquartile Range} = Q_3 - Q_1 = 250 - 200 = 50$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{50}{250+200} = \frac{50}{450} = 0.11$$

EXAMPLE 24.4 : Compute semi-interquartile range and its coefficient.

X :	10	11	12	13	14
f :	3	12	18	12	3

SOLUTION : Clearly, it is a discrete series. Data is already arranged in ascending order of magnitude :

Calculation of Semi-Interquartile Range

X	*f*	*C.f.*
10	3	3
11	12	15
12	18	33
13	12	45
14	3	48
	N = 48	

$$Q_1 = \text{Size of} \left(\frac{N+1}{4}\right)^{th} \text{ item} = \frac{48+1}{4} = 12.25^{th} \text{ item}$$

$$\therefore \quad Q_1 = 11$$

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

$$= 3\left(\frac{48+1}{4}\right)^{th} \text{ item} = 36.75^{th} \text{ item}$$

$$\therefore \quad Q_3 = 13$$

$$\therefore \quad \text{Semi-Interquartile Range} = \frac{Q_3 - Q_1}{2} = \frac{13-11}{2} = 1$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{13-11}{13+11} = \frac{2}{24} = 0.08$$

EXAMPLE 24.5 : Calculate Quartile Deviation from the data given below :

Class :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency :	5	8	12	15	20	14	12	6

SOLUTION : It is a continuous series.

Calculation of Quartile Deviation

Class	*Frequency (f)*	*c.f.*
0-10	5	5
10-20	8	13
20-30	12	25
30-40	15	40
40-50	20	60
50-60	14	74
60-70	12	86
70-80	6	92
	N = 92	

$$Q_1 = \text{Size of} \left(\frac{N}{4}\right)^{th} \text{item} = \frac{92}{4} = 23^{rd} \text{ item}$$

$\therefore$ Lower quartile class is $20 - 30$

$\therefore$ $l = 20, h = 10, f = 12, c = 13$

$$Q_1 = l + \frac{\frac{N}{4} - c}{f} \times h$$

$$= 20 + \frac{23 - 13}{12} \times 10 = 20 + 8.33 = 28.33$$

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

$$= \frac{3 \times 92}{4} = 69^{th} \text{ item}$$

$\therefore$ Upper quartile class is $50 - 60$

$l = 50, h = 10, f = 14, c = 60$

Now
$$Q_3 = l + \frac{\frac{3N}{4} - c}{f} \times h$$

$$= 50 + \frac{69 - 60}{14} \times 10 = 50 + 6.43 = 56.43$$

$\therefore$
$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{56.43 - 28.33}{2} = \frac{28.1}{2} = 14.05$$

EXAMPLE 24.6 : Find the Interquartile Range from the following :

Marks (more than) :	0	20	40	60	80	100	120
No. of students :	80	76	50	28	18	9	3

SOLUTION :

Calculation of Interquartile Range

Marks	*f*	*C.f.*
0-20	4	4
20-40	26	30
40-60	22	52
60-80	10	62
80-100	9	71
100-120	6	77
120-140	3	80
	N = 80	

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

$$= \frac{80}{4} = 20^{th} \text{ item}$$

$\therefore$ Quartile class is 20-40

$$Q_1 = l + \frac{\frac{N}{4} - c}{f} \times h$$

$l = 20, h = 20, c = 4, f = 26$

$\therefore$

$$Q_1 = 20 + \frac{20-4}{26} \times 20 = 20 + 12.3 = 32.3$$

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

$$= \frac{3 \times 80}{4} = 60^{th} \text{ item}$$

$\therefore$ Quartile class is 60-80

$$Q_3 = l + \frac{\frac{3N}{4} - c}{f} \times h$$

$l = 60,\ f = 10,\ c = 52,\ h = 20$

$$Q_3 = 60 + \frac{60-52}{10} \times 20 = 60 + 16 = 76$$

$\therefore$ I.Q.R. $= Q_3 - Q_1 = 76 - 32.3 = 43.7$

EXAMPLE 24.7 : Calculate the mean deviation and its coefficient from the following data :

100, 150, 80, 90, 160, 200, 140

SOLUTION : Data is in the form of individual observation. Arranging the data in ascending order,

80, 90, 100, 140, 150, 160, 200

n = 7 (odd)

$\therefore$ $\quad \text{Med} = \text{Size of}\left(\frac{n+1}{2}\right)^{th} \text{item} = \left(\frac{7+1}{2}\right)^{th} \text{item} = 4^{th} \text{item}$

$\therefore$ $\quad \text{Med} = 140$

Calculation of Mean Deviation

X	$\lvert D \rvert = \lvert X - 140 \rvert$
80	60
90	50
100	40
140	0
150	10
160	20
200	60
	$\Sigma \lvert D \rvert = 240$

$$\text{M.D.} = \frac{\Sigma|D|}{N} = \frac{240}{7} = 34.28$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D}}{\text{Median}} = \frac{34.28}{140} = 0.245$$

EXAMPLE 24.8 : Calculate the mean deviation from the following data :

Income :	100	150	80	200	250	180
No. of persons :	24	26	16	20	6	30

SOLUTION : It is a discrete series. We first arrange the data in ascending order of magnitude (with respect to income).

Calculation of Mean Deviation

Income X	No. of persons f	C.f.	$\lvert D \rvert = \lvert X - 150 \rvert$	$f \lvert D \rvert$
80	16	16	70	1120
100	24	40	50	1200
150	26	66	0	0
180	30	96	30	900
200	20	116	50	1000
250	6	122	100	600
	N = 122			$\Sigma f \lvert D \rvert = 4820$

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

$$= \left(\frac{122+1}{2}\right)^{th} = 61.5^{th} \text{item}$$

$\Rightarrow$ $\quad \text{Med} = 150$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{4820}{122} = 39.51$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}} = \frac{39.51}{150} = 0.2634$$

EXAMPLE 24.9 : Calculate coefficient of M.D. from the data given below :

Mid-value :	115	125	135	145	155	165	175	185	195
Frequency :	6	25	48	72	116	60	38	22	3

SOLUTION : Since we are given the mid-values, we first find the class intervals. Since the difference between the two consecutive mid-values is 10.

∴ Class size is 10.

∴ First class interval will be 110-120. Subsequent class intervals can be found easily.

Calculation of Mean Deviation

Class Intervals	*Mid-value m*	*Freq. (f)*	*C.f*	$\lvert D\rvert = \lvert m - 153.8\rvert$	$f\lvert D\rvert$
110-120	115	6	6	38.8	232.8
120-130	125	25	31	28.8	720.0
130-140	135	48	79	18.8	902.4
140-150	145	72	151	8.8	633.6
150-160	155	116	267	1.2	139.2
160-170	165	60	327	11.2	672.0
170-180	175	38	365	21.2	805.6
180-190	185	22	387	31.2	686.4
190-200	195	3	390	41.2	123.6
		N = 390			$\Sigma f\lvert D\rvert = 4915.6$

$$\text{Med} = \left(\frac{N}{2}\right)^{th} = \frac{390^{th}}{2} = 195^{th} \text{ item}$$

∴ Median class is 150-160

$$\text{Med} = l + \frac{\frac{n}{2} - c}{f} \times h$$

$$= 150 + \frac{195 - 151}{116} \times 10$$

$$= 150 + 3.79 = 153.79 \sim 153.8$$

$$\text{M.D.} = \frac{\Sigma f\lvert D\rvert}{N} = \frac{4915.6}{390} = 12.6$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}} = \frac{12.6}{153.8} = 0.0819$$

EXAMPLE 24.10 : Calculate the mean deviation from mean for the following data :

X :	20	30	40	50	60	70
f :	8	12	20	10	6	4

SOLUTION : We first find mean of the given data. It is a discrete series. ∴ $\overline{X} = \dfrac{\Sigma fX}{\Sigma f}$.

Calculation of Mean Deviation

X	f	fX	$\lvert D\rvert = \lvert X-41\rvert$	$f\lvert D\rvert$
20	8	160	21	168
30	12	360	11	132
40	20	800	1	20
50	10	500	9	90
60	6	360	19	114
70	4	280	29	116
	N = 60	$\Sigma fX = 2460$		$\Sigma f\lvert D\rvert = 640$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{2460}{60} = 41$$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{640}{60} = 10.66$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Mean}} = \frac{10.66}{41} = 0.26$$

EXAMPLE 24.11 : Calculate standard deviation from the following data :

290.12 290.13 290.15 290.12 290.17 290.15
290.17 290.16 290.22 290.21

SOLUTION : The data is given in the form of individual observations. Take A = 290.

Calculation of S.D.

X	$d = X - 290$	d^2
290.12	0.12	0.0144
290.13	0.13	0.0169
290.15	0.15	0.0225
290.12	0.12	0.0144
290.17	0.17	0.0289
290.15	0.15	0.0225
290.17	0.17	0.0289
290.16	0.16	0.0256
290.22	0.22	0.0484
290.21	0.21	0.0441
N = 10	$\Sigma d = 1.6$	$\Sigma d^2 = 0.2666$

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{0.2666}{10} - \left(\frac{1.6}{10}\right)^2} = 0.033$$

EXAMPLE 24.12 : Calculate Standard deviation for the following data :

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students :	10	40	20	0	10	40	16	14

SOLUTION :

Calculation of S.D.

Marks	*Mid-value, m*	*f*	$u = \frac{m-35}{10}$	*fu*	*fu*^{2}
0-10	5	10	−3	−30	90
10-20	15	40	−2	−80	160
20-30	25	20	−1	−20	20
30-40	$\boxed{35} = A$	0	0	0	0
40-50	45	10	1	10	10
50-60	55	40	2	80	160
60-70	65	16	3	48	144
70-80	75	14	4	56	224
		N = 150		Σfu = 64	Σfu^2 = 808

$$\sigma = \sqrt{\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2} \times h = \sqrt{\frac{808}{150} - \left(\frac{64}{150}\right)^2} \times 10$$

$$= \sqrt{5.387 - (0.426)^2} \times 10 = 2.28 \times 10 = 22.8$$

EXAMPLE 24.13 : Calculate the variance of the following data :

Variable :	20-25	25-30	30-35	35-40	40-45	45-50
Frequency :	170	110	80	45	40	35

SOLUTION :

Calculation of Variance

Variable	*Midpoint (m)*	*f*	$u = \frac{m-32.5}{5}$	*fu*	*fu*^{2}
20-25	22.5	170	−2	−340	680
25-30	27.5	110	−1	−110	110
30-35	$\boxed{32.5}$	80	0	−0	0
35-40	37.5	45	1	45	45
40-45	42.5	40	2	80	160
45-50	47.5	35	3	105	315
		N = 480		Σfu = −220	Σfu^2 = 1310

$$\text{Var} = \sigma^2$$

$$= \left[\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2\right] \times h^2$$

Here $h = 5$, $\Sigma fu = -220$, $\Sigma fu^2 = 1310$, $N = 480$

$$\text{Var} = \left[\frac{1310}{480} - \left(\frac{-220}{480}\right)^2\right] \times 25$$

$$= (2.729 - 0.21) \times 25$$
$$= 2.519 \times 25 = 62.975$$

EXAMPLE 24.14 : Calculate S.D. from the data given below :

X :	5	15	25	35	45	55	65
f :	5	12	30	45	50	37	21

SOLUTION :

Calculation of S.D.

X	f	$d = \frac{m-35}{10}$	fd	fd^2
5	5	−3	−15	45
15	12	−2	−24	48
25	30	−1	−30	30
$A = \boxed{35}$	45	0	0	0
45	50	1	50	50
55	37	2	74	148
65	21	3	63	189
	N = 200		Σfd = 118	Σfd² = 510

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times h$$

Here $h = 10,\ \Sigma fd^2 = 510,\ \Sigma fd = 118,\ N = 200$

$$\sigma = \sqrt{\frac{510}{200} - \left(\frac{118}{200}\right)^2} \times 10$$

$$= \sqrt{2.55 - .3481} \times 10 = 14.839$$

EXAMPLE 24.15 : Find the standard deviation of the first n natural numbers.

SOLUTION : Let the numbers be 1, 2, 3 .... n

Then $$\overline{X} = \frac{1+2+3+...+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - (\overline{X})^2}$$

Now $$\Sigma X^2 = 1^2 + 2^2 + 3^2 + ... + n^2$$

$$= \frac{n(2n+1)(n+1)}{6}$$

∴ $$\sigma = \sqrt{\frac{n(2n+1)(n+1)}{6n} - \left(\frac{n+1}{2}\right)^2}$$

$$= \sqrt{\frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4}}$$

$$= \sqrt{\frac{4n^2 + 6n + 2 - 3n^2 - 6n - 3}{12}}$$

$$= \sqrt{\frac{n^2 - 1}{12}}$$

EXAMPLE 24.16 : The mean and S.D. of a normal distribution are Rs 60 and Rs 5 respectively. Find the interquartile range and mean deviation of the distribution. [*B.Com (Hons) DU 1991*]

SOLUTION : Given $\overline{X} = 60, \quad \sigma = 5$

We have $$\text{M.D.} = \frac{4}{5}\sigma = \frac{4}{5} \times 5 = 4$$

Also $$\text{Q.D.} = \frac{2}{3}\sigma = \frac{2}{3} \times 5 = \frac{10}{3}$$

$\therefore$ $$\frac{Q_3 - Q_1}{2} = \frac{10}{3}$$

$\Rightarrow$ $$Q_3 - Q_1 = \frac{20}{3}$$

$\Rightarrow$ $$\text{Interquartile Range} = Q_3 - Q_1 = \frac{20}{3}.$$

EXAMPLE 24.17 : Find the missing information in the following table :

	GROUPS			Combined
	A	B	C	
Number :	50	–	90	200
S.D. :	6	7	–	7.746
Mean :	113	–	115	116

[*B.Com (Hons) DU 1992*]

SOLUTION : It is given that $50 + N_2 + 90 = 200 \Rightarrow N_2 = 60$

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3}$$

Here $$\overline{X}_{123} = 116, \quad N_1 = 50, \quad N_2 = 60, \quad N_3 = 90$$

$$\overline{X}_1 = 113, \quad \overline{X}_2 = ?, \quad \overline{X}_3 = 115$$

$$116 = \frac{50(113) + N_2\overline{X}_2 + 90(115)}{200}$$

$$116 = \frac{5650 + 60\overline{X}_2 + 10350}{200}$$

$$23200 = 16000 + 60\,\overline{X}_2 \quad \Rightarrow \quad \overline{X}_2 = 120$$

$$d_1 = |\overline{X}_{123} - \overline{X}_1| = |116 - 113| = 3$$

$$d_2 = |\overline{X}_{123} - \overline{X}_2| = |116 - 120| = 4$$

$$d_3 = |\overline{X}_{123} - \overline{X}_3| = |116 - 115| = 1$$

Also
$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

$$7.746 = \sqrt{\frac{50(6)^2 + 60(7)^2 + 90(\sigma_3^2) + 50(3)^2 + 60(4)^2 + 90(1)^2}{200}}$$

$$7.746 = \sqrt{\frac{1800 + 2940 + 90\sigma_3^2 + 450 + 960 + 90}{200}}$$

$$(7.746)^2 = \frac{6240 + 90\sigma_3^2}{200}$$

$$60 = \frac{6240 + 90\sigma_3^2}{200}$$

$$12000 = 6240 + 90\sigma_3^2$$

$$\sigma_3^2 = 64 \Rightarrow \sigma_3 = 8$$

EXAMPLE 24.18 : For a group containing 100 items, the A.M. and S.D. are 8 and $\sqrt{105}$. For 50 observations selected from these 100 observations, mean and S.D. are 10 and 2 respectively. Find mean and S.D. of the remaining 50 observations. *[B.Com (Hons) DU 1994]*

SOLUTION : Let the two subgroups be denoted by 1 and 2

Then $N_1 = 50$ $\quad \overline{X}_1 = 10$ $\quad \sigma_1 = 2$ $\quad \overline{X}_{12} = 8$

$N_2 = 50$ $\quad \overline{X}_2 = ?$ $\quad \sigma_2 = ?$ $\quad \sigma_{12} = \sqrt{10.5}$

By combined A.M. we know that

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

$$8 = \frac{50 \times 10 + 50\overline{X}_2}{100}$$

$\Rightarrow$
$$800 = 500 + 50\,\overline{X}_2 \Rightarrow \overline{X}_2 = 6$$

Further
$$d_1 = |\overline{X}_{12} - \overline{X}_1| = |8 - 10| = 2$$

$$d_2 = |\overline{X}_{12} - \overline{X}_2| = |8 - 6| = 2$$

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$\sqrt{10.5} = \sqrt{\frac{50 \times 4 + 50\sigma_2^2 + 50 \times 4 + 50 \times 4}{100}}$$

$$10.5 = \frac{200 + 50\sigma_2^2 + 200 + 200}{100}$$

$$10.5 \times 100 = 600 + 50\sigma_2^2$$

$$1050-600=50\sigma_2^2$$

$$\sigma_2^2=9 \quad \Rightarrow \quad \sigma_2=3$$

EXAMPLE 24.19 : If S.D. is 18.5 calculate Q.D. and M.D. *[B.Com (Hons) DU 1996]*

SOLUTION : $\sigma=18.5 \qquad \text{Q.D.}=\frac{2}{3}\sigma$

$$=\frac{2}{3}\times 18.5=12.33$$

$$\text{M.D.}=\frac{4}{5}\sigma=\frac{4}{5}\times 18.5=14.8$$

EXAMPLE 24.20 : Mean and S.D. of 100 items are found by a student as 50 and 0.1. If at the time of calculation, two items are wrongly taken as 40 and 50 instead of 60 and 30, find correct mean and S.D. *[B.Com DU 1996]*

SOLUTION : $\overline{X}=50 \qquad N=100 \qquad \sigma=0.1$

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

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

$\Rightarrow$ $$\Sigma\overline{X}=50\times 100=5000$$

$\Rightarrow$ $$\Sigma X=5000-\text{wrong items}+\text{correct items}$$

$$=5000-40-50+60+30$$

$$=5000$$

$$\text{Corr } \overline{X}=\frac{5000}{100}=50$$

Now $$\sigma=\sqrt{\frac{\Sigma X^2}{N}-(\overline{X})^2}$$

$$\sigma^2=\frac{\Sigma X^2}{N}-(\overline{X})^2$$

Substituting the uncorrect values as

$$\sigma=0.1 \quad \overline{X}=50 \qquad N=100$$

$$(0.1)^2=\frac{\Sigma X^2}{100}-(50)^2$$

$$0.01=\frac{\Sigma X^2}{100}-2500$$

$$\Sigma X^2=250001$$

$$\text{Corr } \Sigma X^2=250001-(40)^2-(50)^2+(60)^2+(30)^2$$

$$=250001-1600-2500+3600+900$$

$$=250401$$

$$\text{Corr S.D.}=\sqrt{\frac{\Sigma X^2}{N}-(\overline{X})^2}=\sqrt{\frac{250401}{100}-(50)^2}=2.0025.$$

EXAMPLE 24.21 : The means and S.D. of two brands of light bulbs are given below :

	Brand I	Brand II
Mean :	800 hrs	770 hrs
S.D. :	100 hrs	60 hrs

Calculate a measure of relative dispersion for the two brands and interpret the result. *[B.Com DU 2000]*

SOLUTION : For Brand I

$$C.V. = \frac{\sigma}{\overline{X}} \times 100 = \frac{100}{800} \times 100 = 12.5\%$$

For Brand II

$$C.V. = \frac{60}{770} \times 100 = 7.8\%$$

Since C.V. of Brand II is lesser, therefore, Brand II bulbs are more consistent in performance.

EXAMPLE 24.22 : An analysis of the monthly salaries paid to employees in two companies X and Y, belonging to the same industry, provides the following results.

	Company X	Company Y
No. of employees :	1200	1500
Average Salary :	Rs 12000	Rs 9000
S.D. :	Rs 200	Rs 225

Find out :

1. Combined average monthly salary and combined S.D.
2. The company having greater variability in individual salaries. *[B.Com DU 2001]*

SOLUTION : Let $N_1 = 1200 \quad N_2 = 1500$

$$\overline{X}_1 = 12000 \qquad \overline{X}_2 = 9000$$

$$\sigma_1 = 200 \qquad \sigma_2 = 225$$

$$\text{Combined A.M.} = \overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2} = \frac{1200 \times 12000 + 1500 \times 9000}{1200 + 1500}$$

$$\overline{X}_{12} = \frac{27900000}{2700} = 10333$$

Now
$$d_1 = |\overline{X}_{12} - \overline{X}_1| = |10333 - 12000| = 1667$$

$$d_2 = |\overline{X}_{12} - \overline{X}_2| = |10333 - 9000| = 1333$$

Combined S.D

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{(1200)(200)^2 + (1500)(225)^2 + (1200)(1667)^2 + (1500)(1333)^2}{1200 + 1500}}$$

$$= \sqrt{\frac{48000000 + 75937500 + 3334666800 + 2665333500}{2700}}$$

$$= \sqrt{\frac{6123937800}{2700}} = \sqrt{2268125.11} = 1506.03$$

To comment on greater variability, we calculate coefficient of variation of both companies.

$$C.V_1 = \frac{\sigma}{\overline{X}} \times 100 = \frac{200}{1200} \times 100 = 1.67\%$$

$$C.V_2 = \frac{225}{9000} \times 100 = 2.5\%$$

Since C.V. of company Y is greater

$\therefore$ Company Y has greater variability.

EXAMPLE 24.23 : The mean of two samples of sizes 50 and 100 are 54.4 and 50.3 and their S.D. are 8 and 7 respectively. Obtain the combined S.D. of 150 items. *[B. Com (Hons) DU 2005]*

SOLUTION : Given $\overline{X}_1 = 54.4$, $\overline{X}_2 = 50.3$, $N_1 = 50$, $N_2 = 100$ $\sigma_1 = 8$, $\sigma_2 = 7$

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

$$\overline{X}_{12} = \frac{50(54.4) + 100(50.3)}{50 + 100} = \frac{2720 + 5030}{150} = 51.67$$

$$d_1 = |\overline{X}_{12} - \overline{X}_1| = |51.67 - 54.4| = 2.73$$

$$d_2 = |\overline{X}_{12} - \overline{X}_2| = |51.67 - 50.3| = 1.37$$

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_1d_2^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{50(8)^2 + 100(7)^2 + 50(2.73)^2 + 100(1.37)^2}{50 + 100}}$$

$$= \sqrt{\frac{50 \times 64 + 100 \times 49 + 50(7.4529) + 100(1.8769)}{150}}$$

$$= \sqrt{\frac{3200 + 4900 + 372.645 + 187.69}{150}} = \sqrt{57.736} = 7.598$$

EXAMPLE 24.24 : The mean and variance of a set of 150 observations were found to be 2800 and 729. Later on, it was discovered that one observation 3650 was wrongly taken as 3050. Find the correct mean, variance and coefficient of variation ?

SOLUTION : Given $\overline{X} = 2800$, $\sigma^2 = 729$, $N = 150$

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

$$2800 = \frac{\Sigma X}{150}$$

$$\Rightarrow \quad \Sigma X = 2800 \times 150 = 420000$$

$$\text{Corr } \Sigma X = 420000 - \text{Wrong item} + \text{Corr item}$$
$$= 420000 - 3050 + 3650$$
$$= 420600$$

$$\text{Corr } \overline{X} = \frac{\text{Corr } \Sigma X}{N} = \frac{420600}{150} = 2804$$

Now $$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\overline{X})^2}$$

or
$$\sigma^2 = \frac{\Sigma X^2}{N} - (\overline{X})^2$$
$$729 = \frac{\Sigma X^2}{150} - (2800)^2$$
$$729 = \frac{\Sigma X^2}{150} - 7840000$$
$$\frac{\Sigma X^2}{150} = 729 + 7840000$$
$$\frac{\Sigma X^2}{150} = 7840729$$
$$\Rightarrow \quad \Sigma X^2 = 7840729 \times 150$$
$$\Sigma X^2 = 1176109350$$
$$\text{Corr } \Sigma X^2 = 1176109350 - (3050)^2 + (3650)^2$$
$$= 1176109350 - 9302500 + 13322500 = 1180129350$$
$$\text{Corr } \sigma^2 = \frac{\text{Corr } \Sigma X^2}{N} - (\text{Corr } \overline{X})^2$$
$$= \frac{1180129350}{150} - (2804)^2 = 7867529 - 7862416$$
$$\Rightarrow \quad \sigma^2 = 5113$$
$$\text{Corr variance} = 5113$$
$$\text{Corr S.D.} = \sqrt{5113} = 71.5$$
$$\text{Corr C.V.} = \frac{\text{S.D.}}{\text{Mean}} \times 100 = \frac{71.5}{2804} \times 100 = 2.549\%.$$

EXAMPLE 24.25 : Calculate coefficient of variation for the following data :

Variable :	20-25	25-30	30-35	35-40	40-45
Frequency :	1	22	64	10	3

SOLUTION : **Calculation of C.V.**

Variable	*Midpoint* m	*Frequency* f	$d = \frac{M - 32.5}{5}$	fd	fd^2
20-25	22.5	1	−2	−2	4
25-30	27.5	22	−1	−22	22
30-35	$\boxed{32.5}$	64	0	0	0
35-40	37.5	10	1	10	10
40-45	42.5	3	2	6	12
		N = 100		$\Sigma fd = -8$	$\Sigma fd^2 = 48$

$$X = A + \frac{\Sigma fd}{N} \times i = 32.5 - \frac{8}{100} \times 5 = 32.1$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{48}{100} - \left(\frac{-8}{100}\right)^2} \times 5 = \sqrt{0.48 - 0.0064} \times 5 = 3.44$$

$$C.V. = \frac{S.D.}{\overline{X}} \times 100 = \frac{3.44}{32.1} \times 100 = 10.72\%.$$

EXAMPLE 24.26 : After settlement the average weekly wages in a factory had increased from Rs 8000 to Rs 12000 and the S.D. had increased from Rs 100 to Rs 150. After settlement, the wages has become higher and more uniform. Do you agree ?

[B.Com (Hons) DU 2005]

SOLUTION : Before settlement, $C.V. = \frac{S.D.}{Mean} \times 100 = \frac{100}{8000} \times 100 = 1.25\%$

After settlement, $C.V. = \frac{150}{12000} \times 100 = 1.25\%$

Since there is no change in C.V., there is no improvement in uniformity.

EXAMPLE 24.27 : The following is the record of goals scored by team A in a football season :

No. of Goals Scored : 0, 1, 2, 3, 4
No. of Matches : 1, 9, 7, 5, 3

For team B, the average number of goals scored per match was 2.5 with a standard deviation of 1.25 goals. Find which team may be considered as more consistancy.

[B.Com (Pass) 2006]

SOLUTION : Since we have to comment on consistency. ∴ we find coefficient of variation of two teams.

Calculation of C.V for Team A

No. of goals x	*No. of matches* f	$d = x - A$ $d = x - 2$	fd	fd^2
0	1	− 2	− 2	4
1	9	− 1	− 9	9
2	7	0	0	0
3	5	1	5	5
4	3	2	6	12
	n = 25		Σfd = 0	Σfd² = 30

$$\overline{X} = A + \frac{\Sigma fd}{n} = 2 + \frac{0}{25} = 2$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$

$$= \sqrt{\frac{30}{25} - \left(\frac{0}{25}\right)^2} = 1.09$$

$$\text{C.V. for team A} = \frac{S.D}{Mean} \times 100$$

$$= \frac{1.09}{2} \times 100 = 54.5\%$$

$$\text{C.V. for team B} = \frac{1.25}{2.5} \times 100 = 50\%.$$

Since C.V. of team B is lesser than that of A. ∴ team B is more consistent.

EXAMPLE 24.28 : Share prices of two companies A Ltd. and B Ltd. were recorded as follows :

A Ltd.	*B Ltd.*
12	113
13	114
15	113
14	115
14	117
14	114
13	112
17	114

Which company's share prices are more variable ? [*B.Com (Pass) DU 2005*]

SOLUTION :

A Ltd.			*B. Ltd.*		
Size (X)	*d = X – 14*	d^2	*Size (X)*	*d = X – 114*	d^2
12	–2	4	113	–1	1
13	–1	1	114	0	0
15	1	1	113	–1	1
14	0	0	115	1	1
14	0	0	117	3	9
14	0	0	114	0	0
13	–1	1	112	–2	4
17	+3	9	114	0	0
$\Sigma X = 112$	$\Sigma d = 0$	$\Sigma d^2 = 16$	$\Sigma X = 912$	$\Sigma d = 0$	$\Sigma d^2 = 16$

For A Ltd.

$$\overline{X} = \frac{\Sigma X}{N} = \frac{112}{8} = 14$$

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{16}{8} - \left(\frac{0}{8}\right)^2} = 1.414$$

$$C.V. = \frac{\sigma}{\overline{X}} \times 100 = \frac{1.414}{14} \times 100 = 10.10\%$$

For B Ltd.

$$\overline{X} = \frac{\Sigma X}{n} = \frac{912}{8} = 114$$

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\frac{16}{8} - \left(\frac{0}{8}\right)^2} = 1.414$$

$$C.V. = \frac{\sigma}{\overline{X}} \times 100 = \frac{1.414}{114} \times 100 = 1.24\%.$$

Since C.V. of A is more than B $\therefore$ prices of A are more variable.