



## **BBA - I YEAR**

### ***DJB1C* : BUSINESS MATHEMATICS AND STATISTICS**

#### **SYLLABUS**

##### **UNIT – I:**

Measures of central tendency – Mean, Median, Mode, G.M. and H.M. Dispersion – Range – Q.D. – M.D. – S.D. – C.V.

##### **UNIT – II:**

Simple Correlation and Regression – Index numbers – Time series – Components – Estimating Trend.

##### **UNIT – III:**

Differential Calculus – Sum, Product, Quotient Rules, Maximum and Minimum.

##### **UNIT – IV:**

Integral Calculus – Rules of Integration – Definite integral – Area interpretation.

##### **UNIT – V:**

Matrices – Inversion – Solving system of equations.

##### **References:**

1. Statistics Theory and Practice – R.S.N. Pillai and V. Bagavathi.
2. An introduction to Business Mathematics – V. Sundaresan and S.D. Jeyaseelan.



## INDEX

Unit		Title	Page Number
I	1.1.	Measures of Central Tendency	3
	1.2.	Dispersion	39
II	2.1.	Correlation	59
	2.2.	Regression	68
	2.3.	Index Numbers	79
	2.4.	Analysis of Time Series	91
III	3.1.	Differential Calculus	105
IV	4.1.	Integral Calculus	119
V	5.1.	Matrices	135



## **UNIT - I:**

*Measures of central tendency – Mean, Median, Mode, G.M. and H.M. Dispersion – Range – Q.D. – M.D. – S.D. – C.V.*

### **MEASURES OF CENTRAL TENDENCY (AVERAGES)**

#### **INTRODUCTION**

Condensation of data is necessary in statistical analysis. This is because a large number of big figures are not only confusing but also difficult to analyse. Therefore, in order to reduce the complexity of data and to make them comparable it is necessary that various phenomena, which are being compared, are reduced to a single figure. The first of such measures is averages or measures of central tendency. Measures of central tendency are a typical value of the entire group or data. It describes the characteristics of the entire mass of data. It reduces the complexity of data and makes them to compare. According to Prof. R. A. Fisher, “The inherent inability of the human mind to grasp in it’s entirety a large body of numerical data compels us to seek relatively few constants that will adequately describe the data.” Human mind is incapable of remembering the entire mass of unwieldy data. So a simple figure is used to describe the series which must be a representative number.

#### **Meaning:**

A central tendency is a central or typical value for a probability distribution. It may also be called a center or location of the distribution. Colloquially, measures of central tendency are often called averages. The term central tendency dates from the late 1920s. If a large volume of data is summarized and given in one simple term. Then it is called as the ‘Central Value’ or an ‘average’. In other words an average is a single value that represents group of values.

#### **Definitions:**

Various statisticians have defined the word average differently. Some of the important definitions are given below:



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

According to Croxton and Cowdon, "An average value is a single value with" the range of the data that is used to represent all of the values in the series. Since an average is somewhere within the range of the data, it is also called the measure of central value."

According to Murry R. Speigal, "Average is a value which is typical or representative of a set of data"

It is clear from the above definition that average is called a 'type' as it is a typical value of the entire data and is a measure of central tendency.

### **Objectives or Functions of an average**

Averages occupy a prime place in the theory of statistical methods. That is why Bowley remarked, "Statistics is a science of averages." The following are the objectives of an average:

#### **1. Facilitates Comparison:**

The foremost purpose of average is that it facilitates comparison. For instance, a comparison of the production of jute in Maharashtra a Punjab shows that production of jute in Maharashtra is much more as compared Punjab.

#### **2. Formulation of Policies:**

Averages are of great use in the formulation of various policy measures. For instance, when the Government, finds that there is a fear of low product of sugar, it can formulate various policies to compensate the same.

#### **3. Short Description:**

Averages help to present the raw data in a brief a systematic manner.



#### **4. Representation of Universe:**

Average represents universe. According conclusions can be drawn in respect of the universe as a whole.

#### **Characteristics of a Typical Average**

A measure of central tendency is a typical value around which other figures congregate. Average condenses a frequency distribution in one figure. According to the statisticians Yule and Kendall, an average will be termed good or efficient if possesses the following characteristics:

1. It should be easily understandable.
2. It should be rigidly defined. It means that the definition should be so clear that the interpretation of the definition does not differ from person to person.
3. The average should be such that it can be easily determined.
4. The average of a variable should be based on all the values of the variable. This means that in the formula for average all the values of the variable should be incorporated
5. The value of average should not change significantly along with the change in sample. This means that the values of the averages of different samples of the same size drawn from the same population should have small variations. In other words, an average should possess sampling stability
6. It should be amenable to algebraic treatment
7. The average should be unduly affected by extreme values. i.e, the formula for average should be such that it does not show large due to the presence of one or two very large or very small values of the variable.

#### **Qualities of Good Averages**

The qualities of good averages are as follows:



1. It should be properly defined, preferably by a mathematical formula, so that different individuals working with the same data should get the same answer unless there are mistakes in calculations.
2. It should be simple to understand and easy to calculate.
3. It should be based on all the observations so that if we change the value of any observation, the value of the average should also be changed.
4. It should not be unduly affected by extremely large or extremely small values.
5. It should be capable of algebraic manipulation. By this we mean that if we are given the average heights for different groups, then the average should be such that we can find the combined average of all groups taken together.
6. It should have quality of sampling stability. That is, it should not be affected by the fluctuations of sampling.

For example, if we take ten or twelve samples of twenty students' each and find the average height for each sample, we should get approximately the same average height for each sample.

### **Kinds of Averages**

The following are the important types of averages:

1. Arithmetic mean
  - i) Simple and
  - ii) Weighted
2. Median
3. Mode
4. Geometric mean and
5. Harmonic Mean

### **MEAN**

Mean is one of the types of averages. Mean is further divided into three kinds, which are the arithmetic mean, the geometric mean and the harmonic mean. These kinds are explained as follows;



## i) Arithmetic Mean: Simple Arithmetic Average:

### A. Individual Observation: Direct Method:

The arithmetic mean is most commonly used average. It is generally referred as the average or simply mean. The arithmetic mean or simply mean is defined as the value obtained by dividing the sum of values by their number or quantity. It is denoted as  $\bar{X}$  (read as X-bar). Therefore, the mean for the values  $X_1, X_2, X_3, \dots, X_n$  shall be denoted by  $\bar{X}$ . Following is the mathematical representation for the formula for the arithmetic mean or simply, the mean.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} = \frac{\Sigma X}{N}$$

Where,  $\bar{X}$  = Arithmetic Mean

$\Sigma x$  = Sum of all the values of the variables i.e.,  $X_1 + X_2 + X_3 + \dots + X_n$

$N$  = Number of observations.

**Illustration 1:** Calculate mean from the following data:

<b>Roll Numbers</b>	1	2	3	4	5	6	7	8	9	10
<b>Marks</b>	40	50	55	78	58	60	73	35	43	48

**Solution: Calculation of mean**

<b>Roll Numbers</b>	<b>Marks (x)</b>
1	40
2	50
3	55
4	78
5	58
6	60
7	73



8	35
9	43
10	48
N = 10	$\Sigma X = 540$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{540}{10} = 54 \text{ marks.}$$

### Short cut method:

The arithmetic mean can also be calculated by short cut method. This method reduces the amount of calculation.

### Formula:

$$\bar{X} = A \pm \frac{\Sigma d}{N}$$

Where,  $\bar{X}$  = Arithmetic Mean; A = Assumed mean;

$\Sigma d$  = Sum of the deviations; N = Number of items.

### Illustration 2: (Solving the previous problem)

Roll Numbers	Marks (X)	d = X - A
1	40	-10
2	50	0
3	55	5
4	78	28
5	58	8
6	60	10
7	73	23
8	35	-15





9	43	-7
10	48	-2
<b>N = 10</b>		<b>Σd = 40</b>

Let the assumed mean, A = 50

$$\bar{X} = A \pm \frac{\Sigma d}{N} = 50 + \frac{40}{10} = 54 \text{ marks.}$$

### B. Discrete Series: Direct Method:

To find out the total of items in discrete series, frequency of each value is multiplied with the respective size. The values so obtained are totaled up. This total is then divided by the total number of frequencies to obtain the arithmetic mean. The formula is

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

Where,  $\bar{X}$  = Arithmetic Mean;  $\Sigma fx$  = the sum of products;  $N$  = Total frequency.

**Illustration 3:** Calculate mean from the following data:

<b>Value</b>	1	2	3	4	5	6	7	8	9	10
<b>Frequency</b>	21	30	28	40	26	34	40	9	15	57

**Solution: Calculation of mean**

<b>x</b>	<b>f</b>	<b>fx</b>
1	21	21
2	30	60
3	28	84
4	40	160
5	26	130
6	34	204
7	40	280



8	9	72
9	15	135
10	57	570
	<b><math>\Sigma f = N = 300</math></b>	<b><math>\Sigma fx = 1716</math></b>

$$\bar{X} = \frac{\Sigma fx}{N} = \frac{1716}{300} = 5.72.$$

### Short cut Method

#### Formula:

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

Where,  $\bar{X}$  = Arithmetic Mean; A = Assumed mean

$\Sigma fd$  = Sum of total deviations; N = Total frequency.

#### Illustration: 4 (Solving the previous problem)

X	f	d = X - A	fd
1	21	-4	-84
2	30	-3	-90
3	28	-2	-56
4	40	-1	-40
5	26	0	0
6	34	1	34
7	40	2	80
8	9	3	27
9	15	4	60
10	57	5	285
	<b><math>\Sigma f = N = 40</math></b>		<b><math>\Sigma fd = + 216</math></b>



Let the assumed mean,  $A = 5$

$$\bar{X} = A \pm \frac{\sum fd}{N}$$

$$\bar{X} = 5 + \frac{216}{300}$$

$$= 5.72.$$

### C. Continuous Series

In continuous frequency distribution, the value of each individual frequency distribution is unknown. Therefore an assumption is made to make them precise or on the assumption that the frequency of the class intervals is concentrated at the centre that the midpoint of each class interval has to be found out. In continuous frequency distribution, the mean can be calculated by any of the following methods:

1. Direct Method
2. Short cut method
3. Step Deviation Method

#### 1. Direct Method

The formula is  $\bar{X} = \frac{\sum fm}{N}$

Where,  $\bar{X}$  = Arithmetic Mean;  $\sum fm$  = Sum of the product of  $f$  &  $m$ ;  $N$  = Total frequency.

**Illustration 5:** From the following find out the mean:

<b>Class Interval</b>	0 – 10	10 – 20	20 – 30	30 – 40	40 - 50
<b>Frequency</b>	6	5	8	15	7

**Solution: Calculation of Mean**

<b>Class Interval</b>	<b>Mid Point (m)</b>	<b>Frequency (f)</b>	<b>fm</b>
0 – 10	$\frac{0 + 10}{2} = 5$	6	30



10 – 20	$\frac{10 + 20}{2} = 15$	5	75
20 – 30	$\frac{20 + 30}{2} = 25$	8	200
30 – 40	$\frac{30 + 40}{2} = 35$	15	525
40 - 50	$\frac{40 + 50}{2} = 45$	7	315
	<b><math>\Sigma f = N = 41</math></b>		<b><math>\Sigma fm = 1145</math></b>

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{1145}{41} = 27.93$$

## 2. Short cut method

**Formula:**

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

Where,  $\bar{X}$  = Arithmetic Mean; A = Assumed mean

$\Sigma fd$  = Sum of total deviations; N = Total frequency.

**Illustration: 6** (Solving the previous problem)

Class Interval	Mid Point (m)	d = m - A	Frequency (f)	fd
0 – 10	$\frac{0 + 10}{2} = 5$	5 – 25 = -20	6	-120
10 – 20	$\frac{10 + 20}{2} = 15$	15 – 25 = - 10	5	-50
20 – 30	$\frac{20 + 30}{2} = 25$	25 – 25 = 0	8	0
30 – 40	$\frac{30 + 40}{2} = 35$	35 – 25 = 10	15	150
40 - 50	$\frac{40 + 50}{2} = 45$	45 – 25 = 20	7	140
	<b><math>\Sigma f = N = 41</math></b>			<b><math>\Sigma fm = +120</math></b>

$$d = m - A; \text{ here } A = 25$$



$$\bar{X} = A \pm \frac{\Sigma fd}{N} = 25 + \frac{120}{41} = 25 + 2.93 = 27.93$$

### 3. Step Deviation Method

**Formula:**

$$\bar{X} = A \pm \frac{\Sigma fd'}{N} \times C$$

Where,  $\bar{X}$  = Arithmetic Mean; A = Assumed mean; N

$\Sigma fd'$  = Sum of total deviations; N = Total frequency; C = Common Factor

**Illustration: 7** (Solving the previous problem)

Class Interval	Mid-Point (m)	Frequency (f)	d = m - A	d' = $\frac{m-A}{C}$	fd'
0 - 10	$\frac{0 + 10}{2} = 5$	6	5 - 25 = -20	-2	-12
10 - 20	$\frac{10 + 20}{2} = 15$	5	15 - 25 = -10	-1	-5
20 - 30	$\frac{20 + 30}{2} = 25$	8	25 - 25 = 0	0	0
30 - 40	$\frac{30 + 40}{2} = 35$	15	35 - 25 = 10	1	15
40 - 50	$\frac{40 + 50}{2} = 45$	7	45 - 25 = 20	2	14
	<b><math>\Sigma f = N = 41</math></b>				<b><math>\Sigma fd' = +12</math></b>

Here A = 25; C = 10

$$\bar{X} = A \pm \frac{\Sigma fd'}{N} \times C$$

$$= 25 + \frac{12}{41} \times 10 = 25 + \frac{120}{41}$$

$$= 25 + 2.93 = 27.93$$



## WEIGHTED ARITHMETIC MEAN

When the values are not of equal importance, we assign them certain numerical values to express their relative importance. These numerical values are called weights. If  $X_1, X_2, \dots, X_k$  have weights  $W_1, W_2, \dots, W_k$ , then the **weighted arithmetic mean** or the **weighted mean**, which is denoted as  $\bar{X}_w$ , is calculated by the following formula;

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + \dots + W_kX_k}{W_1 + W_2 + \dots + W_k} = \frac{\Sigma WX}{\Sigma W}$$

Thus the mean of grouped data may be regarded as the weighted mean of the values of the values  $X_1, X_2, \dots, X_k$  whose weights are the respective class frequencies  $f_1, f_2, \dots, f_k$ .

**Illustration 8:** The marks obtained by a student in **English, Urdu and Statistics** were **70, 76, and 82** respectively. Find the appropriate average if weights of **5, 4 and 3** are assigned to these subjects.

We will use the weighted mean, the weights attached to the marks being 5, 4 and 3. Thus,

$$\begin{aligned}\bar{X}_w &= \frac{W_1X_1 + W_2X_2 + \dots + W_kX_k}{W_1 + W_2 + \dots + W_k} = \frac{\Sigma WX}{\Sigma W} \\ &= \frac{5(70) + 4(76) + 3(82)}{5 + 4 + 3} \\ &= \frac{350 + 304 + 246}{12} = \frac{900}{12} = 75 \text{ pounds}\end{aligned}$$

**Illustration 9:** A contractor employs three types of workers, male, female and children. To a male he pays Rs.40 per days. To a female worker Rs.32 per day and to a child worker Rs.15 per day. The number of male, female and children workers employed is 20, 15 and 15 respectively. Find out the weighted Arithmetic means.



**Solution: Calculation of Weighted Arithmetic Mean**

Daily wages in Rs. (X)	No. of workers (W)	WX
40	2	800
32	15	480
15	15	225
	$\Sigma W = 50$	$\Sigma WX = 1505$

$$\text{Weighted Arithmetic Average } \bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{1505}{50} = 30.10$$

**Illustration 10:** A train runs 25 miles at a speed of 30 mph another 50 miles at a speed of 40 mph; then due to repairs of the track travels for 6 minutes at a speed of 10 m.p.h and finally covers the remaining distance of 24 miles at a speed of 24 mph. What is the average speed in miles per hour?

**Solution:**

$$\text{Time for 25 miles at a speed of 30 m.p.h} = \frac{25}{30} \times 60 = 50 \text{ minutes}$$

$$\text{For 50 miles at a speed of 40 m.p.h.} = \frac{50}{40} \times 60 = 75 \text{ minutes}$$

$$\text{For 24 miles at a speed of 24 m.p.h} = \frac{24}{24} \times 60 = 60 \text{ minutes}$$

Take the time as weights

Speed in miles per hour (X)	Time taken W	WX
30	50	1500
40	75	3000
10	6	60
24	60	1440

$$\text{Average speed } \bar{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{6000}{191} = 31.41 \text{ miles per hour.}$$



## MEDIAN

Median is the value of item that goes to divide the series into equal parts. It may be defined as the value of that item which divides the series into equal parts, one half containing values greater than it and the other half containing values less than it. Therefore, the series has to be arranged in ascending or descending order, before finding the median. If the items of a series are arranged in ascending or descending order of magnitude, the item which falls in the middle of it is called median. Hence it is the “middle most” or “most central” value of a set of number.

### Definitions

According to L. R. Connor, “The median is that value of the variable which divides the group into two equal parts, one part comprising all values greater and the other, all values less than median”.

According to YauLun Chou, “The median, as its name indicates, is the value of the middle item in a series, when items are arranged according to magnitude”.

Calculation of Median – Individual Series

**Illustration 1:** Find out the median of the following items

X: 10, 15, 9, 25, 19.

**Solution: Computation of Median**

S. No.	Size of ascending order	Size of descending order
1	9	25
2	10	19
3	15	15
4	19	10
5	25	9

$$\begin{aligned}\text{Median} &= \text{Size of } \frac{(N+1)}{2} \text{th item} \\ &= \text{Size of } \frac{(5+1)}{2} \text{th item} = 3^{\text{rd}} \text{ item} = 15.\end{aligned}$$





**Illustration 2:** Find out the median of the following items

X: 8, 10, 5, 9, 12, 11.

**Solution: Computation of Median**

S. No.	X
1	5
2	8
3	9
4	10
5	11
6	12

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

$$= \text{Size of } \frac{(6+1)}{2} \text{th item} = \text{Size of } 3.5^{\text{th}} \text{ item}$$

$$= \text{Size of } \frac{(\text{3rd item} + \text{4th item})}{2} = \frac{9+10}{2} = 9.5$$

**Calculation of Median - Discrete Series**

**Illustration 3:** Locate median from the following:

Size of shoes	5	5.5	6	6.5	7	7.5	8
Frequency	10	16	28	15	30	40	34

**Solution: Computation of Median**

Size of shoes	f	c.f
5	10	10
5.5	16	26
6	28	54
6.5	15	69
7	30	99
7.5	40	139
8	34	173



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

$$= \text{Size of } \frac{(173+1)}{2} \text{th item}$$

$$= \text{Size of } 87^{\text{th}} \text{ item} = 7$$

### Calculation of Median - Continuous Series

**Illustration 4:** Calculate the median of the following table:

Marks	10 - 25	25 - 40	40 - 55	55 - 70	70 - 85	85 - 100
Frequency	6	20	44	26	3	1

### Solution: Computation of Median

$x$	$f$	$c.f$
10 - 25	6	6
25 - 40	20	26
40 - 55	44	70
55 - 70	26	96
70 - 85	3	99
85 - 100	1	100

$$\text{Median} = L + \frac{\frac{N}{2} - cf}{f} \times i$$

$$\frac{N}{2} = \frac{100}{2} = 50; L = 40; f = 44; cf = 26; i = 15$$

$$\text{Median} = 40 + \frac{50-26}{44} \times 15 = 40 + 8.18$$

$$= 48.18 \text{ marks}$$

### Merits:

1. It is easy to compute and understand.



2. It eliminates the effect of extreme items.
3. The value of median can be located graphically.
4. It is amenable to further algebraic process as it is used in the measurement of dispersion.
5. It can be computed even if the items at the extremes are unknown.

### **Demerits:**

1. For calculating median, it is necessary to arrange the data; other averages do not need any arrangement.
2. Typical representative of the observations cannot be computed if the distribution of item is irregular.
3. It is affected more by fluctuation of sampling than the arithmetic mean.

### **MODE**

Mode is the value which occur the greatest number of frequency in a series. It is derived from the French word 'La mode' meaning the fashion. It is the most fashionable or typical value of a distribution, because it is repeated the highest number of times in the series.

Mode or the modal value is defined as the value of the variable which occur more number of times or most frequently in a distribution.

### **Definition**

According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated."

### **Types of Mode**

**i) Unimodal:** If there is only one mode in series, it is called unimodal.

Eg., 10, 15, 20, 25, 18, 12, 15 (Mode is 15)



ii) **Bi - modal:** If there are two modes in the series, it is called bi - modal.

Eg., 20, 25, 30, 30, 15, 10, 25 (Modes are 25, 30)

iii) **Tri - modal:** If there are three modes in the series, it is called Tri - modal.

Eg., 60, 40, 85, 30, 85, 45, 80, 80, 55, 50, 60 (Modes are 60, 80, 85)

iv) **Multi – modal:** If there are more than three modes in the series it is called multi-modal.

### Merits:

1. It can be easily ascertained without much mathematical calculation.
2. It is not essential to know all the items in a series to compute mode.
1. Open – end classes do not disturb the position of the mode.
2. Its values can be ascertained graphically as well as empirically.
3. It may be very well applied to qualitative as well quantitative data.
4. It is not affected by extreme values as in the average.

### Demerits:

1. The mode becomes less useful as an average which the distribution is bi-modal.
2. It is not suitable for further mathematical treatment.
3. It is stable only when the sample is large.
4. Mode is influenced by magnitude of the class-intervals.

### Mode - Individual Series

**Illustration : 1.** Calculate the mode from the following data of the marks obtain by 10 students.

Serial No.	1	2	3	4	5	6	7	8	9	1
Marks obtained	60	77	74	62	77	77	70	68	65	80

### Solution:

Marks obtained by 10 students 60, 77, 74, 62, 77, 77, 70, 68, 65, and 80.



Here 77 is repeated three times.

∴ The Mode mark is 77.

## DISCRETE SERIES

A grouping Table has six columns

**Column 1:** In column 1 write the actual frequencies and mark the highest frequency.

**Column 2:** Frequencies are grouped in twos, adding frequencies of items 1 and 2; 3 and 4; 5 and 6; and so on.

**Column 3:** Leave the first frequency and then add the remaining in twos.

**Column 4:** Group of frequencies in threes.

**Column 5:** Leave the first frequency and group the remaining in threes.

**Column 6:** Leave the first two frequencies and then group the remaining the threes.

The maximum frequencies in all six columns are marked with a circle and an analysis table is prepared as follows:

1. Put column number on the left – hand side
2. Put the various probable values of mode on the right – hand side.
3. Enter the highest marked frequencies by means of a bar in the relevant box corresponding to the values they represent.

**Illustration: 2.** Calculate the mode from the following:

<b>Size</b>	10	11	12	13	14	15	16	17	18
<b>Frequency</b>	10	12	15	19	20	8	4	3	2



### Solution: Grouping Table

Size	Frequency					
	1	2	3	4	5	6
10	10					
		22				
11	12			37		
			27			
12	15				46	
		34				
13	19					54
			39			
14	20			47		
		28				
15	8				32	
			12			
16	4					15
		7				
17	3			9		
			5			
18	2					

### Analysis Table

Column No.	Size of item containing maximum frequency				
	11	12	13	14	15
1					
2		1	1		
3			1	1	



<b>4</b>			1	1	1
<b>5</b>	1	1	1	1	
<b>6</b>		1	1	1	
	1	3	<b>5</b>	4	1

The mode is 13, as the size of item repeats 5 times. But through inspection, we say the mode is 14, because the size 14 occurs 20 times. But this wrong decision is revealed by analysis table.

### Calculation of Mode - Continuous Series

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where, Z = Mode;

$L_1$  = Lower limit of the modal class;  $f_1$  = Frequency of the modal;

$f_0$  = Frequency of the class preceding the modal class;

$f_2$  = Frequency of the class succeeding the modal class;  $i$  = Class interval

**Illustration: 3.** Calculate the mode from the following:

Size of item	Frequency
0 – 5	20
5 – 10	24
10 – 15	32
15 – 20	28
20 – 25	20
25 – 30	16
30 – 35	34
35 – 40	10
40 – 45	8



### Solution: Grouping Table

Size of item	Frequency					
	1	2	3	4	5	6
0 - 5	20					
		44		76		
5 - 10	24					
			56			
10 - 15	32				84	
		60				80
15 - 20	28					
			48			
20 - 25	20			64		
		36				
25 - 30	16				70	
			50			
30 - 35	34					60
		44				
35 - 40	10			52		
			18			
40 - 45	8					





### Analysis Table

Column No.	Size of item containing maximum frequency					
	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	30 - 35
1						1
2			1	1		
3		1	1			
4	1	1	1			
5		1	1	1		
6			1	1	1	
	1	3	5	3	1	1

$$Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$L_1 = 10; f_1 = 32; f_0 = 24; f_2 = 28; i = 5$$

$$Z = 10 + \frac{32 - 24}{2 \times 32 - 24 - 28} \times 5$$

$$= 10 + \frac{40}{12} = 10 + 3.33$$

$\therefore$  The Mode is = 13.33

### The relationship among mean, median and mode:

The three averages (mean, median and mode) are identical, when the distribution is symmetrical. In an asymmetrical distribution, the values of mean, median and mode are not equal. In a moderately asymmetrical distribution the distance between the mean and median is about one-third of the distance between mean and mode.

$$\text{Mean} - \text{Median} = 1/3 (\text{Mean} - \text{Mode})$$

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

$$\text{Median} = \text{Mode} + 2/3 (\text{Mean} - \text{Mode})$$



$$\text{Mode} = \text{Mean} - 3 (\text{Mean} - \text{Mode})$$

$$\text{Mean} - \text{Mode} = 3 (\text{Mean} - \text{Median})$$

**Illustration 1:** The value of mode and median for a moderately skewed distribution are 64.2 and 68.6 respectively. Find the value of the mean.

**Solution:**

$$\text{Mean} = \text{Mode} + (3/2) (\text{Median} - \text{Mode})$$

$$\text{Mean} = 64.2 + (3/2) (68.6 - 64.2)$$

$$= 64.2 + (3/2) (68.6 - 64.2) = 64.2 + 6.6 = 70.8$$

∴ The value of the mean is 70.8

**Illustration 2:** In a moderately asymmetrical distribution, the values of mode and mean are 32.1 and 35.4 respectively. Find the median value.

**Solution:**

$$\text{Median} = 1/3 (2 \text{ Median} + \text{Mode})$$

$$\text{Median} = 1/3 (2 \times 35.4 + 32.1)$$

$$= 1/3 \times 102.9 = 34.3$$

∴ The median value is 34.3

**Illustration 3:** In a frequency distribution, the values of arithmetic mean and median are 68 and 66 respectively. Estimate the value of the mode.

**Solution:**

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

Given: Mean = 68, Median = 66



$$\text{Mode} = 3(66) - 2(68) = 198 - 136 = 62$$

∴ The value of the Mode is 62.

## GEOMETRIC MEAN

The geometric mean, G, of a set of  $n$  positive values  $X_1, X_2, \dots, X_n$  is the  $N$ th root of the product of  $N$  items. Mathematically the formula for geometric mean will be as follows;

$$\text{G. M} = \sqrt[n]{X_1 X_2 \dots X_n} = (X_1 X_2 \dots X_n)^{1/n}$$

G.M = Geometric Mean;  $n$  = number of items;  $X_1, X_2, X_3, \dots$  = are various values

**Illustration 1:** The geometric mean of the values 2, 4 and 8 is the cubic root of  $2 \times 4 \times 8$  or

$$\sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4$$

In practice, it is difficult to extract higher roots. The geometric mean is, therefore, computed using logarithms. Mathematically, it will be represented as follows;

$$\text{Geometric Mean} = \text{Antilog of } \frac{\log X_1 + \log X_2 + \log X_3 \dots \log X_n}{N} \text{ or G. M.} = \text{Antilog of } \frac{\log X}{N}$$

Here we assume that all the values are positive, otherwise the logarithms will be not defined.

### Geometric Mean – Individual Series

**Illustration 2:** Calculate the geometric mean of the following:

50    72    54    82    93

**Solution:** Calculation of Geometric Mean

X	log of X
50	1.6990
72	1.8573



54	1.7324
82	1.9138
93	1.9685
<b>N = 5</b>	<b><math>\sum \log X = 9.1710</math></b>

$$\text{G.M.} = \sqrt{50 \times 72 \times 54 \times 82 \times 93} \text{ or}$$

$$\text{G. M.} = \text{Antilog of } \frac{\log X}{N}$$

$$\text{G.M} = \text{Antilog of } \frac{9.1710}{5}$$

$$= \text{Antilog of } 1.8342$$

$$= 68.26$$

### Geometric Mean – Discrete Series

$$\text{G.M.} = \text{antilog of } \frac{\sum f \log X}{N}$$

Where, f = frequency value; log x = logarithm of each value; N = Total frequencies

**Illustration 3:** The following table gives the weight of 31 persons in a sample survey. Calculate geometric mean.

Weight (lbs)	130	135	140	145	146	148	149	150	157
No. of persons	3	4	6	6	3	5	2	1	1



### Solution: Calculation of Geometric Mean

Size of item (X)	Frequency (f)	log X	f log X
130	3	2.1139	6.3417
135	4	2.1303	8.5212
140	6	2.1461	12.8766
145	6	2.1614	12.9684
146	3	2.1644	6.4932
148	5	2.1703	10.8515
149	2	2.1732	4.3464
150	1	2.1761	2.1761
157	1	2.1959	2.1959
	<b>N = <math>\sum f = 31</math></b>		<b><math>\sum f \log X = 66.7710</math></b>

$$\text{G.M.} = \text{antilog of } \frac{\sum f \log X}{N}$$

$$\text{G.M.} = \text{antilog of } \frac{66.7710}{31} = \text{antilog of } 2.1539$$

$$\text{G.M. Weight} = 142.5 \text{ lbs}$$

### Geometric Mean – Continuous Series

$$\text{G.M.} = \text{antilog of } \frac{\sum f \log m}{N}$$

Where, f = frequency; m = midvalue; log m = logarithm of each midvalue; N = Total frequencies

**Illustration 4:** Find out the geometric mean:

Yield of wheat (mounds)	No. of farms
7.5 – 10.5	5
10.5 – 13.5	9



13.5 – 16.5	19
16.5 – 19.5	23
19.5 – 22.5	7
22.5 – 25.5	4
25.5 – 28.5	1

### Solution: Calculation of Geometric Mean

Yield of wheat (mounds)	No. of farms (f)	m	log m	f log m
7.5 – 10.5	5	9	0.9542	4.7710
10.5 – 13.5	9	12	1.0792	9.7128
13.5 – 16.5	19	15	1.1761	22.3459
16.5 – 19.5	23	18	1.2553	28.8719
19.5 – 22.5	7	21	1.3222	9.2554
22.5 – 25.5	4	24	1.3802	5.5208
25.5 – 28.5	1	27	1.4314	1.4314
	<b><math>N = \sum f = 68</math></b>			<b><math>\sum f \log m = 81.9092</math></b>

$$\text{G.M.} = \text{antilog of } \frac{\sum f \log m}{N}$$

$$\text{G.M.} = \text{antilog of } \frac{81.9092}{68} = \text{antilog of } 1.2045$$

$$= 16.02 \text{ maunds}$$

### Uses:

1. Geometric mean is highly useful in averaging ratios, percentages and rate of increase between two periods.
2. Geometric mean is important in the construction of index numbers.
3. In economic and social sciences, where we want to give more weight to smaller items and smaller weight to large items, geometric mean is appropriate.



4. It is the only useful average that can be employed to indicate rate of change.

#### Merits:

1. Every item in the distribution is included in the calculation.
2. It can be calculated with mathematical exactness, provided that all the quantities are greater than zero and positive.
3. Large items have less effect on it than the arithmetic average.
4. It is amenable to further algebraic manipulation.

#### Demerits:

1. It is very difficult to calculate.
2. It is impossible to use it when any item is zero or negative.
3. The value of the geometric mean may not correspond with any actual value in the distribution.
4. It cannot be used in the series in which the end values of the classes are left open.

### HARMONIC MEAN

Harmonic Mean, like geometric mean is a measure of central tendency in solving special types of problems. Harmonic Mean is the reciprocal of the arithmetic average of the reciprocal of values of various items in the variable. The reciprocal of a number is that value, which is obtained dividing one by the value.

**For example**, the reciprocal of 5 is 1/5. The reciprocal can be obtained from logarithm tables.

#### Harmonic Mean – Individual Series

$$\text{H.M.} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}} \text{ or } \text{H.M.} = \frac{N}{\sum \frac{1}{x}}$$

$x_1, x_2, x_3, \dots, x_n$ , refer to the various in the observations



**Illustration 5:** The monthly incomes of 10 families in rupees are given below:

<b>Family:</b>	1	2	3	4	5	6	7	8	9	10
<b>Income:</b>	85	70	10	75	500	8	42	250	40	36

**Solution: Calculation of Harmonic Mean**

Family	Income (x)	Reciprocals( $\frac{1}{x}$ )
1	85	$1/85 = 0.0118$
2	70	$1/70 = 0.0143$
3	10	$1/10 = 0.1000$
4	75	$1/75 = 0.0133$
5	500	$1/500 = 0.0020$
6	8	$1/8 = 0.1250$
7	42	$1/42 = 0.0232$
8	250	$1/250 = 0.0040$
9	40	$1/40 = 0.0250$
10	36	$1/36 = 0.0278$
<b>N = 10</b>		$\sum \frac{1}{x} = \mathbf{0.3464}$

$$\text{H.M.} = \frac{N}{\sum \frac{1}{x}} = \frac{10}{0.3464}$$

$$= \text{Rs. } 28.87/-$$

**Harmonic Mean – Discrete Series**

$$\text{H.M.} = \frac{N}{\sum f(\frac{1}{x})}$$





**Illustration 6:** Calculate harmonic mean from the following data.

<b>Size of items:</b>	6	7	8	9	10	11
<b>Frequency:</b>	4	6	9	5	2	8

**Solution:** Calculation of harmonic mean

Size of items (x)	Frequency (f)	Reciprocal ( $\frac{1}{x}$ )	$f \frac{1}{x}$
6	4	0.1667	0.6668
7	6	0.1429	0.8574
8	9	0.1250	1.1250
9	5	0.1111	0.5555
10	2	0.1000	0.2000
11	8	0.0909	0.7272
	<b>N = <math>\sum f = 34</math></b>		<b><math>\sum f(\frac{1}{x}) = 4.1319</math></b>

$$\text{H.M.} = \frac{N}{\sum f(\frac{1}{x})} = \frac{34}{4.1319} = 8.23$$

**Harmonic Mean - Continuous Series**

$$\text{H.M.} = \frac{N}{\sum f(\frac{1}{m})}$$

**Illustration 7:** Calculate H.M. of the following data:

<b>Size:</b>	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
<b>Frequency:</b>	5	8	12	6	4

**Solution:**

**Calculation of Harmonic Mean**

Size (x)	Frequency (f)	Mid Value (m)	$\frac{1}{m}$	$f(\frac{1}{m})$
0 – 10	5	5	$1/5 = 0.2000$	1.0000



10 – 20	8	15	$1/15 = 0.0667$	0.5336
20 – 30	12	25	$1/25 = 0.0400$	0.4800
30 – 40	6	35	$1/35 = 0.0286$	0.1716
40 - 50	4	45	$1/45 = 0.0222$	0.0888
	$N = \sum f = 35$			$\sum f \left( \frac{1}{m} \right) = 2.274$

$$\text{H.M.} = \frac{N}{\sum f \left( \frac{1}{m} \right)} = \frac{35}{2.274} = 15.39$$

### Relationship between Arithmetic mean, Geometric mean and Harmonic mean.

1. If all the items in a distribution have the same, then,

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

**For example,** we take two positive items 5 and 5

$$\text{A.M} = \frac{5+5}{2} = \frac{10}{2} = 5$$

$$\text{G.M.} = \sqrt{5 \times 5} = 5$$

$$\text{H.M} = \frac{2}{\frac{1}{5} + \frac{1}{5}} = \frac{2}{\frac{2}{5}} = \frac{2 \times 5}{2} = 5$$

$$\text{Thus A.M} = \text{G.M.} = \text{H.M}$$

2. But if the size vary, as will generally be the case,

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

**For example,** we take two positive items 4 and 9

$$\text{A.M} = \frac{4+9}{2} = \frac{13}{2} = 6.5$$

$$\text{G.M.} = \sqrt{4 \times 9} = 6$$



$$H.M = \frac{2}{\frac{1}{4} + \frac{1}{9}} = \frac{2}{\frac{13}{36}} = \frac{2 \times 36}{13} = 5.54$$

Thus  $A.M > G.M > H.M$ .

### Merits:

1. It utilizes all values of a variable.
2. It is very important to small values.
3. It is amenable to further algebraic manipulation.
4. It provides consistent results in problems relating to time and rates than similar averages.

### Demerits:

1. It is not very easy to understand.
2. The method of calculation is difficult.
3. The presence of both positive and negative items in a series makes it impossible to compute its value. The same difficulty is felt if one or more items are zero.
4. It is only a summary figure and may not be the actual item in the series.

\*\*\*\*\*

### Theoretical Questions

1. Define and explain the objectives of averages.
2. What are the functions of averages?
3. Explain the characteristics of typical average.
4. Give the qualities of good average.
5. What are the kinds of averages? Explain in detail.
6. Write a note on weighted arithmetic mean. How do you calculate it?
7. Discuss the merits and demerits of median.
8. What is mode? Explain its types.
9. Discuss the merits and demerits of modes.
10. Explain the procedure to compute mode for discrete and continuous series.



11. What is G.M? Explain its uses
12. Discuss the merits and demerits of G.M.
13. Describe the relationship between A.M, G.M & H.M. with suitable examples.

### Practical Problems

1. The monthly income of 10 families of a certain locality is given in rupees as below.  
Calculate the arithmetic average.

Families	A	B	C	D	E	F	G	H	I	J
Income in rupees	85	70	10	75	500	8	42	250	40	36

(Mean = Rs. 111. 60)

2. The coins are tossed 1024 times. The theoretical frequencies of 10 heads to 0 head are given below. Calculate the mean number of heads per tossing.

No. of heads	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	10	45	120	210	252	210	120	45	10	1

(Mean = Rs. 5)

3. Find mean from the following frequency distribution:

Class Interval	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 - 75
Frequency	4	11	19	14	0	2

(Mean = Rs. 40. 2)

4. The following are the marks scored by 7 students; find out the median marks:

Roll Numbers	1	2	3	4	5	6	7
Marks	45	32	18	57	65	28	46

(Median marks = 45)

5. Find out the median from the following:

57	58	61	42	38	65	72	66
----	----	----	----	----	----	----	----

(Median = 59.5)

6. Find the median

$x$	13	14	15	16	17	18	19	20	21	22	23	24	25
$f$	37	162	343	390	256	433	161	355	65	85	49	46	40

(Median = 18)



7. Find the median:

Wages Rs.	60 – 70	50 – 60	40 – 50	30 – 40	20 - 30
No. of labourers	5	10	20	5	3

(Median = 46.75)

8. 10 persons have the following income:

Rs.	850	750	600	825	850	725	600	850	640	530
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

(Mode = 850)

9. Calculate the mode from the following series:

Size	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Frequency	40	48	52	57	60	63	57	55	50	52	41	57	63	52	48	40

(Mode = 9)

10. Find the mode:

Size	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	5	7	12	18	16	10	5

(Mode = 37.5)

11. Calculate mean, median and mode form the following frequency distribution of marks at a test in statistics:

Marks	5	10	15	20	25	30	40	45	50
No. of students	20	43	75	76	72	45	9	8	50

(Mean = 22. 16; Median = 20; mode = 20)

12. Calculate the mean, median and mode for the following data.

Profits per shop	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 - 60
No. of shops	12	18	27	20	17	6

(Mean = 28; Median = 27.4; mode = 25.62)

13. Calculate the geometric mean of the following series:

2000	35	400	15	40	1500	300	6	90	250	20
------	----	-----	----	----	------	-----	---	----	-----	----

(G.M = 55.35)



14. Calculate G.M from the following series:

Weekly wages (in Rs.)	13	18.5	20.5	22	23	24
No. of workers	8	10	14	6	7	3

(G.M = 19.05)

15. Compute the geometric mean of the following series.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	5	7	15	25	8

(G.M = 25. 63)

16. Calculate the H.M of the following series of monthly expenditure of a batch students:

Rs.	125	130	75	10	45	0.5	0.40	500	150	5
-----	-----	-----	----	----	----	-----	------	-----	-----	---

(H.M = 2. 06)

17. From the following distribution, calculate H.M.

Marks	50	40	25	20	10
No. of Students	10	20	50	30	20

(H.M = 20. 97)

18. Find the harmonic mean from the data given below:

Class boundaries	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75
Frequency	4	11	19	14	0	2

(H.M = 37. 04)

19. The monthly income of 8 families in rupees in a certain locality is given below. Calculate the A.M, the G.M and the H.M and confirm the relationship  $A.M < G.M < H.M$ . holds true.

Family:	A	B	C	D	E	F	G	H
Income Rs.:	70	10	500	75	8	250	8	42

(A.M = Rs. 120.37, G.M = Rs. 45. 27, H.M = 19. 635)

\*\*\*\*\*



## DISPERSION

Dispersion is studied to have an idea of the homogeneity or heterogeneity of the distribution. Measures of dispersion are the measures of scatter or spread about an average. Measures of dispersion are called the averages of the second order.

### Definitions:

According to A.L. Bowley, “Dispersion is the measure of the variation of the items”.

According to L.R. Connor, “Dispersion is a measure of extent to which the individual items vary”.

According to Simpson and Kafka, “The measurement of the scatterness of the mass of figures in a series about an average is called measure of variation or dispersion”.

According to Spiegel, “The degree to which numerical data tend to spread about an average value is called the variation dispersion of the data”.

### Importance of Dispersion

Measures of Dispersion are important in:

1. Testing the reliability of the measures of central tendency.
2. Comparing two or more series on the basis of their variability and deciding the consistency of performance.
3. Enabling to control the variability, in quality of products.

### Characteristics of Dispersion

1. It is easy to understand.
2. It is easy to compute.
3. It should be based on all observations and it should not be affected by extreme observations.
4. It should be amenable to further algebraic treatment.



## Methods of Measuring Dispersion

There are various methods of studying variation or dispersion important methods studying dispersion are as follows:

1. Range
2. Inter - quartile range
3. Mean Deviation
4. Standard Deviation
5. Lorenz curve

### 1. Range

Range is the simplest and crudest measure of dispersion. It is a rough measure of dispersion. It is the difference between the highest and the lowest value in the distribution.

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

Where, L = Largest Value; S = Smallest Value.

The Relative measure of range is called as the Co - efficient of Range.

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

**Illustration1:** Find the range of weights of 7 students from the following

27, 30, 35, 36, 38, 40, 43

**Solution:**

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

Here L = 43; S = 27

$$\therefore \text{Range} = 43 - 27 = 16$$





$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{43-27}{43+27} = \frac{16}{70} = 0.23$$

### Practical utility of Range

1. It is used in industries for the statistical quality control of the manufactured product.
2. It is used to study the variations such as stock, shares and other commodities.
3. It facilitates the use of other statistical measures.

### Advantages

1. It is the simplest method
2. It is easy to understand and the easiest to compute.
3. It takes minimum time to calculate and accurate.

### Disadvantages

1. Range is completely dependent on the two extreme values.
2. It is subject to fluctuations of considerable magnitude from sample to sample.
3. Range cannot tell us anything about the character of the distribution.

## 2. Quartile Deviation

Quartile deviation is an absolute measure of dispersion. Co-efficient of quartile deviation is known as relative measure of dispersion.

In the series, four quartiles are there. By eliminating the lowest items (25%) and the highest items (25%) of a series we can obtain a measure of dispersion and can find out the half of the distance between the first and the third quartiles. That is,  $[Q_3 \text{ (third quartiles)} - Q_1 \text{ (first quartiles)}]$ . The inter-quartile range is reduced to the form of the semi – inter quartile range (or) quartile deviation by dividing it by 2.

$$\text{Inter quartile range} = Q_3 - Q_1$$

$$\text{Inter quartile range or Quartile deviation} = \frac{Q_3 - Q_1}{2}$$



$$\text{Coefficient of Quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

### Quartile Deviation – Individual Series

**Illustration 2:** Find out the value of Quartile Deviation and its coefficient from the following data:

Roll No.	1	2	3	4	5	6	7
Marks	20	28	40	30	50	60	52

**Solution: Calculation of Q.D.**

Marks arranged in ascending order:	20	28	30	40	50	52	60
------------------------------------	----	----	----	----	----	----	----

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{th item}$$

$$Q_1 = \text{size of } \frac{7+1}{4} \text{th item}$$

$$= \text{size of } \frac{8}{4} \text{th item}$$

$$= \text{size of } 2^{\text{nd}} \text{ item} = 28$$

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

$$Q_3 = \text{Size of } 3\left(\frac{7+1}{4}\right) \text{th item}$$

$$= \text{Size of } 3\left(\frac{8}{4}\right) \text{th item} = \text{size of } \frac{24}{4} \text{th item}$$

$$= \text{size of } 6^{\text{th}} \text{ item} = 52$$

$$Q. D. = \frac{Q_3 - Q_1}{2} = \frac{52 - 28}{2} = \frac{24}{2} = 12$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{52 - 28}{52 + 28} = \frac{24}{80} = 0.3$$



## Quartile Deviation – Discrete Series

**Illustration 3:** Find out the value of Quartile Deviation and its coefficient from the following data:

Age in years	20	30	40	50	60	70	80
No. of members	3	61	132	153	140	51	3

**Solution:** Calculation of Q.D.

Age in years (x)	No. of members (f)	c.f.
20	3	3
30	61	64
40	132	196
50	153	349
60	140	489
70	51	540
80	3	543

$$Q_1 = \text{Value of } \frac{N+1}{4} \text{th item}$$

$$Q_1 = \text{value of } \frac{543+1}{4} \text{th item} = \text{value of } \frac{544}{4} \text{th item}$$

$$= \text{value of } 136^{\text{th}} \text{ item} = 40 \text{ years}$$

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

$$Q_3 = \text{value of } 3\left(\frac{543+1}{4}\right)^{\text{th}} \text{ item} = \text{value of } 3\left(\frac{544}{4}\right)^{\text{th}} \text{ item}$$

$$= \text{value of } 3(136)^{\text{th}} \text{ item} = \text{value of } 408^{\text{th}} \text{ item} = 60 \text{ years}$$

$$Q. D. = \frac{Q_3 - Q_1}{2} = \frac{60 - 40}{2} = \frac{20}{2} = 10 \text{ years}$$



$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{60 - 40}{60 + 40} = \frac{20}{100} = 0.2$$

### Quartile Deviation – Continuous Series

**Illustration 4:** Find out the value of Quartile Deviation and its coefficient from the following data:

<b>Wages (Rs.)</b>	30 – 32	32 – 34	34 – 36	36 – 38	38 – 40	40 – 42	42 – 44
<b>Labourers</b>	12	18	16	14	12	8	6

**Solution: Calculation of Q.D.**

<b>Wages (x)</b>	<b>Labourers (f)</b>	<b>c.f.</b>
30 – 32	12	12
32 – 34	18	30
34 – 36	16	46
36 – 38	14	60
38 – 40	12	72
40 – 42	8	80
42 – 44	6	86

$$Q_1 = \text{size of } \frac{N}{4} \text{th item} = \text{size of } \frac{86}{4} \text{th item} = 21.5^{\text{th}} \text{ item}$$

$\therefore Q_1$  lies in the group 32 - 34

$$Q_1 = L + \frac{\frac{N}{4} - cf}{f} \times h$$

$$= 32 + \frac{21.5 - 12}{18} \times 2 = 32 + \frac{9.5}{18} \times 2$$

$$Q_1 = 32 + \frac{19}{18} = 32 + 1.06 = 33.06$$

$$Q_1 = \text{size of } \frac{3N}{4} \text{th item}$$



$$= \text{size of } \frac{3 \times 86^{\text{th}}}{4} \text{ item} = 64.5 \text{ th item}$$

$\therefore Q_3$  lies in the group 38 – 40.

$$Q_3 = L + \frac{\frac{3N}{2} - cf}{f} \times i$$

$$= 38 + \frac{64.5 - 60}{12} \times 2 = 32 + \frac{4.5}{12} \times 2$$

$$Q_3 = 32 + \frac{9}{12} = 32 + 0.75 = 38.75$$

$$Q. D. = \frac{Q_3 - Q_1}{2} = \frac{38.75 - 33.06}{2} = \frac{5.69}{2} = 2.85$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{38.75 - 33.06}{38.75 + 33.06} = \frac{5.69}{71.81} = 0.08$$

#### Merits:

1. It is simple to calculate.
2. It is easy to understand.
3. Risk of extreme item variation is eliminated, as it depends upon the central 50 percent items.

#### Demerits

1. Items below  $Q_1$  and above  $Q_3$  are ignored.
2. It is not capable of further mathematical treatment.
3. It is affected much by the fluctuations of sampling.
4. It is not calculated from a computed average, but from a positional average.

**3. Mean Deviation** The mean deviation is also known as the average deviation. It is the average difference between the items in a distribution computed from the mean, median or mode of that series counting all such deviation as positive. Median is preferred to the average because the sum of deviation of items from median is minimum when signs are ignored. But, the arithmetic



mean is more frequently used in calculating the value of average deviation. Hence, it is commonly called Mean deviation.

### Mean Deviation – Individual Series

$$\text{M. D. (mean or median or mode)} = \frac{\sum |D|}{N}$$

$$\text{Coefficient of Mean Deviation: } \frac{\text{Mean Deviation}}{\text{Mean or median or mode}}$$

**Illustration 5:** Calculate mean deviation from mean and median for the following data:

100	150	200	250	360	490	500	600	671
-----	-----	-----	-----	-----	-----	-----	-----	-----

### Solution: Calculation of Mean Deviation

X	D  = X - $\bar{X}$ ; X - 369	D  = X - median; X - 360
100	269	260
150	219	210
200	169	160
250	119	110
360	9	0
490	121	130
500	131	140
600	231	240
671	302	311
$\sum X = 3321$	$\sum  D  = 1570$	$\sum  D  = 1561$

$$\text{Mean } \bar{X} = \frac{\sum X}{N}$$

$$= \frac{3321}{9} = 369$$

$$\text{Median} = \text{Value of } \frac{(N+1)}{2} \text{th item}$$

$$= \text{Value of } \frac{(9+1)}{2} \text{th item}$$

$$= \text{Value of 5th item} = 360$$



$$\begin{aligned} \text{M.D. from mean} &= \frac{\sum |D|}{N} \\ &= \frac{1570}{9} = 174.44 \end{aligned}$$

$$\begin{aligned} \text{M.D. from mean} &= \frac{\sum |D|}{N} \\ &= \frac{1561}{9} = 173.44 \end{aligned}$$

$$\begin{aligned} \text{Coefficient of M.D.} &= \frac{\text{M.D.}}{\bar{X}} \\ &= \frac{174.44}{369} = 0.47 \end{aligned}$$

$$\begin{aligned} \text{Coefficient of M.D.} &= \frac{\text{M.D.}}{\text{Median}} \\ &= \frac{173.44}{369} = 0.48 \end{aligned}$$

### Mean Deviation – Discrete Series

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

**Illustration 6:** Calculate mean deviation from mean from the following data:

$x$	2	4	6	8	10
$f$	1	4	6	4	1

### Solution: Calculation of Mean Deviation

$x$	$f$	$fx$	$ D  = x - \bar{X}$	$f D $
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
	$N = \sum f = 16$	$\sum fx = 96$		$\sum f  D  = 24$

$$\text{Mean } \bar{X} = \frac{\sum fx}{N} = \frac{96}{16} = 6$$

$$\text{M.D. from mean} = \frac{\sum f |D|}{N} = \frac{24}{16} = 1.5$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\bar{X}} = \frac{1.5}{6} = 0.25$$



## Mean Deviation - Continuous Series

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

**Illustration 7:** Calculate mean deviation from mean from the following data:

Class interval	2 - 4	4 - 6	6 - 8	8 - 10
Frequency	3	4	2	1

### Solution: Calculation of Mean Deviation

Class Interval (x)	Mid value (m)	f	fm	D  = m - $\bar{X}$	f  D
2 - 4	3	3	9	2.2	6.6
4 - 6	5	4	20	0.2	0.8
6 - 8	7	2	14	1.8	3.6
8 - 10	9	1	9	3.8	3.8
		<b>N = <math>\sum f = 10</math></b>	<b><math>\sum fm = 52</math></b>		<b><math>\sum f  D  = 14.8</math></b>

$$\text{Mean } \bar{X} = \frac{\sum fm}{N} = \frac{52}{10} = 5.2$$

$$\text{M.D. from mean} = \frac{\sum f |D|}{N} = \frac{14.8}{10} = 1.48$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\bar{X}} = \frac{1.48}{5.2} = 0.148$$

### Merits

1. It is clear and easy to understand.
2. It is based on each and every item of the data.
3. It can be calculated from any measure of central tendency and as such is flexible too.
4. It is not disturbed by the values of extreme items as in the case of range.

### Demerits:

1. It is not suitable for further mathematical processing.
2. It is rarely used in sociological studies.





#### 4. Standard Deviation

Karl Pearson introduced the concept of Standard deviation in 1893. Standard deviation is the square root of the means of the squared deviation from the arithmetic mean. So, it is called as Root - Mean Square Deviation or Mean Error or Mean Square Error. The Standard deviation is denoted by the small Greek letter 'σ' (read as sigma)

#### Standard Deviation – Individual Observation

##### Deviation taken from Actual Mean

$$\sigma = \sqrt{\frac{\sum x^2}{N}} \text{ or } \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}} \text{ or } \sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

**Illustration 8:** Calculate the standard deviation from the following data;

14, 22, 9, 15, 20, 17, 12, 11

**Solution:** Calculation of standard deviation from actual mean

Values (X)	X <sup>2</sup>	X - $\bar{X}$ ; (X - 15)	(X - $\bar{X}$ ) <sup>2</sup>
14	196	-1	1
22	484	7	49
9	81	-6	36
15	225	0	0
20	400	5	25
17	289	2	4
12	144	-3	9
11	121	-4	16
<b>N = 8; <math>\sum X = 120</math></b>	<b><math>\sum X^2 = 1940</math></b>		<b><math>\sum (X - \bar{X})^2 = 140</math></b>

$$\bar{X} = \frac{\sum X}{N} = \frac{120}{8} = 15$$



$$\sigma = \sqrt{\frac{\sum x^2}{N}} \text{ or } \sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}} = \sqrt{\frac{140}{8}} = \sqrt{17.5} = 4.18$$

### Alternatively

We can find out standard deviation by using variables directly, i.e., no deviation is found out.

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\frac{1940}{8} - \left(\frac{120}{8}\right)^2} = \sqrt{242.5 - 225} = \sqrt{17.5} = 4.18$$

### Deviation taken from Assumed Mean

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

Where  $d = X - A$

**Illustration 9:** Calculate the standard deviation from the following data;

30, 43, 45, 55, 68, 69, 75.

**Solution:** Calculation of standard deviation from assumed mean

X	d = X - A = X - 55	d <sup>2</sup>
30	-25	625
43	-12	144
45	-10	100
55	0	0
68	13	169
69	14	196
75	20	400
N = 7	$\sum d = 0$	$\sum d^2 = 1634$

$$\sigma = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$



$$= \sqrt{\frac{1634}{7} - \left(\frac{0}{7}\right)^2} = \sqrt{233.429} = 15.28$$

## Standard Deviation - Discrete Series

### Actual Mean Method

$$\sigma = \sqrt{\frac{\sum fd^2}{N}}$$

**Illustration 10:** Calculate the standard deviation from the following data;

Marks	10	20	30	40	50	60
No. of students	8	12	20	10	7	3

**Solution:** Calculation of standard deviation (from actual mean)

Marks (x)	No. of students (f)	fx	d = x - $\bar{X}$ x - 30.8	d <sup>2</sup>	fd <sup>2</sup>
10	8	80	-20.8	432.64	3461.12
20	12	240	-10.8	116.64	1399.68
30	20	600	-0.8	0.64	12.80
40	10	400	9.2	84.64	846.40
50	7	350	19.2	368.64	2580.48
60	3	180	29.2	852.64	2557.92
	<b>N = <math>\sum f</math> = 60</b>	<b><math>\sum fx</math> = 1850</b>			<b><math>\sum fd^2</math> = 10858.40</b>

$$\text{Mean: } \bar{X} = \frac{\sum fx}{N} = \frac{1850}{60} = 30.8$$

$$\text{Standard Deviation: } \sigma = \sqrt{\frac{\sum fd^2}{N}} = \sqrt{\frac{10858.40}{60}} = 13.45$$

### Assumed Mean Method

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}; \text{ Where } d = X - A$$



**Illustration 11: (Solving the previous problem)**

**Solution: Calculation of standard deviation (from assumed mean)**

$x$	$f$	$d = x - 30$	$fd$	$fd^2$
10	8	-20	-160	3200
20	12	-10	-120	1200
30	20	0	0	0
40	10	10	100	1000
50	7	20	140	2800
60	3	30	90	2700
$N = \sum f = 60$			$\sum fd = 50$	$\sum fd^2 = 10900$

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

$$\sigma = \sqrt{\frac{10900}{60} - \left(\frac{50}{60}\right)^2} = \sqrt{181.67 - 0.69} = \sqrt{180.98} = 13.45$$

**Step Deviation Method**

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

Where  $d' = \frac{x - A}{C}$ ; C = Common Factor

**Illustration 12: (Solving the previous problem)**

**Solution: Calculation of standard deviation (from step deviation)**

$x$	$f$	$d' = \frac{x-30}{10}$	$fd'$	$fd'^2$
10	8	-2	-16	32
20	12	-1	-12	12
30	20	0	0	0



40	10	1	10	10
50	7	2	14	28
60	3	3	9	27
	<b>N = <math>\sum f = 60</math></b>		<b><math>\sum fd' = 5</math></b>	<b><math>\sum fd'^2 = 109</math></b>

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

$$\sigma = \sqrt{\frac{109}{60} - \left(\frac{5}{60}\right)^2} \times 10 = \sqrt{1.817 - 0.0069} \times 10$$

$$= \sqrt{1.81} \times 10 = 1.345 \times 10 = \mathbf{13.45}$$

### Standard Deviation - Continuous Series

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

Where  $d = \frac{m - A}{C}$ ; C = Common Factor

**Illustration13:** Compute the standard deviation from the following data:

Class	0 - 10	10 - 20	20 - 30	30 - 40	40-50
Frequency	5	8	15	16	6

**Solution:**

### Computation of standard deviation

x	m	f	$d = \frac{m-25}{10}$	$d^2$	fd	$fd^2$
0 - 10	5	5	-2	4	-10	20
10 - 20	15	8	-1	1	-8	8
20 - 30	25	15	0	0	0	0
30 - 40	35	16	1	1	16	16
40 - 50	45	6	2	4	12	24
		<b>N = <math>\sum f = 50</math></b>			<b><math>\sum fd = 10</math></b>	<b><math>\sum fd^2 = 68</math></b>



$$\begin{aligned}\sigma &= \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C \\ \sigma &= \sqrt{\frac{68}{50} - \left(\frac{10}{50}\right)^2} \times 10 \\ &= \sqrt{1.36 - (0.2)^2} \times 10 \\ &= \sqrt{1.36 - 0.04} \times 10 \\ &= \sqrt{1.32} \times 10 = 1.1489 \times 10 = \mathbf{11.49}\end{aligned}$$

### Distinguish between Mean deviation and standard Deviation.

	<b>Mean Deviation</b>	<b>Standard Deviation</b>
1.	It is simple to calculate.	It is difficult to calculate.
2.	Algebraic signs are ignored while calculating mean deviation.	Signs are taken into account while calculating standard deviation.
3.	Deviations can be computed either from mean, median or mode.	Deviations can be computed either from mean.
4.	It lacks mathematical properties, because algebraic signs are ignored.	It is mathematically sound, because algebraic signs are taken into account.

### Merits:

1. It is rigidly defined determinate.
2. It is based on all the observations of a series.
3. It is less affected by fluctuations of sampling and hence stable.
4. It is amenable to algebraic treatment and is less affected by fluctuations of sampling most other measures of dispersion.
5. The standard deviation is more appropriate mathematically than the mean deviation, since the negative signs are removed by squaring the deviations rather than by ignoring



### **Demerits:**

1. It lacks wide popularity as it is often difficult to compute, when big numbers are involved, the process of squaring and extracting root becomes tedious.
2. It attaches more weight to extreme items by squaring them.
3. It is difficult to calculate accurately when a grouped frequency distribution has extreme groups with no definite range.

### **Uses:**

1. Standard deviation is the best measure of dispersion.
2. It is widely used in statistics because it possesses most of the characteristics of an ideal measure of dispersion.
3. It is widely used in sampling theory and by biologists.
4. It is used in coefficient of correlation and in the study of symmetrical frequency distribution.

### **Co - efficient of variation (Relative Standard Deviation)**

The Standard deviation is an absolute measure of dispersion. The corresponding relative measure is known as the co - efficient of variation. It is used to compare the variability of two or more than two series. The series for which co-efficient or variation is more is said to be more variable or conversely less consistent, less uniform less table or less homogeneous.

### **Variance:**

Square of standard deviation is called variance.

$$\text{Variance} = \sigma^2; \sigma = \sqrt{\text{Variance}}$$

$$\text{Co - efficient of standard deviation} = \frac{\sigma}{\bar{X}}$$

$$\text{Co - efficient of variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100$$



**Illustration 14:** The following are the runs scored by two batsmen A and B in ten innings:

A	101	27	0	36	82	45	7	13	65	14
B	97	12	40	96	13	8	85	8	56	15

Who is the more consistent batsman?

**Solution:**

Batsman A			Batsman B		
Runs Scored X	$dx = X - \bar{X}$	$dx^2$	Runs Scored Y	$dy = Y - \bar{Y}$	$dy^2$
101	62	3844	97	54	2916
27	-12	144	12	-31	961
0	-39	1521	40	-3	9
36	-3	9	96	53	2809
82	43	1849	13	-30	900
45	6	36	8	-35	1225
7	-32	1024	85	42	1764
13	-26	676	8	-35	1225
65	26	676	56	13	169
14	-25	625	15	-28	784
$\Sigma X = 390$		$\Sigma dx^2 = 10404$	$\Sigma Y = 430$		$\Sigma dy^2 = 12762$

**Batsman A**

**Batsman B**

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

$$\frac{390}{10} = 39$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{430}{10} = 43$$

$$\sigma_X = \sqrt{\frac{\Sigma dx^2}{N}} = \sqrt{\frac{10404}{10}} = 32.26 \quad \sigma_Y = \sqrt{\frac{\Sigma dy^2}{N}} = \sqrt{\frac{12762}{10}} = 35.72$$

$$C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{32.26}{39} \times 100$$

$$C.V. = \frac{\sigma}{\bar{Y}} \times 100 = \frac{35.72}{43} \times 100$$

$$= 82.72\%$$

$$= 83.07\%$$





Batsman A is more consistent in his batting, because the co-efficient of variation of runs is less for him.

### Theoretical Questions

1. Define and explain characteristics of dispersion.
2. Give the importance dispersion.
3. Elucidate the methods of measuring dispersion.
4. What is the practical utility of range?
5. Explain its merits & demerits of range.
6. Distinguish between Mean deviation and standard Deviation.
7. What are the uses of S.D?

### Practical Problems

1. Calculate Range, Q.D, M.D (from mean), S.D and C.V of the marks obtained by 10 students given below:

50	55	57	49	54	61	64	59	59	56
----	----	----	----	----	----	----	----	----	----

(Range = 15, Q.D = 2.75, M. D = 3.6, S.D = 4.43 and C.V = 7.85%)

2. Compute Q.D from the following data:

Height in inches	58	59	60	61	62	63	64	65	66
No. of students	15	20	32	35	33	22	20	10	8

(Q.D = 1.5)

3. Compute quartile deviation from the following data.

<i>x</i> :	4 – 8	8 – 12	12 - 16	16 - 20	20 - 24	24 - 28	28 - 32	32 – 36	36 - 40
<i>f</i> :	6	10	18	30	15	12	10	6	2

(Q.D = 5.21)

4. Calculate mean deviation (from mean) from the following data:

<i>x</i> :	2	4	6	8	10
<i>f</i> :	1	4	6	4	1

(M.D = 1.5)



5. Calculate the mean deviation from the following data.

$x:$	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
$f:$	449	705	507	281	109	52	16	4

(M.D = 5.25)

6. Calculate the S.D of the following:

Size of the item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

(S.D = 1.6)

7. Compute standard deviation of the following data:

Wages(Rs. Per day)	1 - 3	3 - 5	5 - 7	7 - 9	9 - 11
No. of workmen	15	18	27	10	6

(S.D = 2.33)

8. Two cricketers scored the following runs in the several innings. Find who is a better run getter and who more consistent player is?

A:	42	17	83	59	72	76	64	45	40	32
B:	28	70	31	0	59	108	82	14	3	95

(C.V (A) = 38% and C.V (B) = 75.6%)

9. Two brands of types are tested with the following results:

Life ('000 miles)	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50
Brand X	8	15	12	18	13	9
Brand Y	6	20	32	30	12	0

(C.V (X) = 21.82% and C.V (Y) = 16.11%)

\*\*\*\*\*



## **UNIT – II:**

*Simple Correlation and Regression - Index numbers - Time series - Components - Estimating Trend.*

### **CORRELATION ANALYSIS**

Correlation refers to the relationship of two or more variables. For example, there exists some relationship between the height of a mother and the height of a daughter, sales and cost and so on. Hence, it should be noted that the detection and analysis of correlation between two statistical variables requires relationship of some sort which associates the observation in pairs, one of each pair being a value of the two variables. The word ‘relationship’ is of important and indicates that there is some connection between the variables under observation. Thus, the association of any two variates is known as correlation.

#### **Definitions:**

According to YaLun Chou, “Correlation analysis attempts to determine the degree of relationship between variables”.

According to L.R.Connon “If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in the other(s), then they are said to be correlated.”

#### **Uses of correlation**

1. Correlation is very useful in physical and social sciences, business and economics.
2. Correlation analysis is very useful to economists to study the relationship between variables like price and quantity demanded.
3. It helps for businessmen to estimate costs, sales, price and other related variables.
4. Correlation is the basis of the concept of regression and ratio of variation.
5. Correlation analysis helps in calculating the sampling error.



## Types of Correlation

1. Positive and Negative correlation
2. Simple and Multiple correlations
3. Partial and total correlation
4. Linear and Non-linear correlation.

### 1. Positive and Negative Correlation :

The correlation is said to be positive when the values of two variables move in the same direction, so that an increase in the value of one variable is accompanied by an increase in the value of the other variable or a decrease in the value of one variable is followed by a decrease in the value of the other variable.

The correlation is said to be negative when the values of two variables move in opposite direction, so that an increase or decrease in the values of one variable is followed by a decrease or increase in the value of the other.

### 2. Simple and multiple Correlation :

When only two variables are studied then the correlation is said to be simple correlation.

**Example:** The study of price and demand of an article.

When more than two variables are studied simultaneously, the correlation is said to be multiple correlation.

**Example:** the study of price, demand and supply.

### 3. Partial and total Correlation:

Partial correlation coefficient provides a measure of relationship between a dependent variable and a particular independent variable when all other variables involved are kept constant. i.e., when the effect of all other variables are removed.



**Example:** When we study the relationship between the yield of rice per acre and both the amount of rainfall and the amount of fertilizers used. In these relationship if we limit our correlation analysis to yield and rainfall. It becomes a problem relating to simple correlation.

#### **4. Linear and Non-linear Correlation :**

The correlation is said to be linear, if the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable.

The correlation is non-linear, if the amount of change in one variable does not bear a constant ratio to the amount of change in the other related variable.

#### **Methods of studying correlation**

The following correlation methods are used to find out the relationship between two variables.

##### **A. Graphic Method :**

- i.** Scatter diagram (or) Scattergram method.
- ii.** Simple Graph or Correlogram method.

##### **B. Mathematical Method :**

- i.** Karl Pearson's Coefficient of Correlation.
- ii.** Spearman's Rank Correlation of Coefficient
- iii.** Coefficient of Concurrent Deviation
- iv.** Method of least squares.

##### **C. Graphic Method**

#### **i. Scatter diagram method of correlation**

This is the simplest method of finding out whether there is any relationship present between two variables by plotting the values on a chart, known as scatter diagram. In this method, the given data are plotted on a graph paper in the form of dots. X variables are plotted on the horizontal axis and Y variables on the vertical axis. Thus we have the dots and we can know the scatter or concentration of various points.



If the plotted dots fall in a narrow band and the dots are rising from the lower left hand corner to the upper right-hand corner it is called high degree of positive correlation.

If the plotted dots fall in a narrow band from the upper left hand corner to the lower right hand corner it is called a high degree of negative correlation.

If the plotted dots are scattered all over the diagram, there is no correlation between the two variables.

**Merits:**

1. It is easy to plot even by beginner.
2. It is simple to understand.
3. Abnormal values in a sample can be easily detected.
4. Values of some dependent variables can be found out.

**Demerits:**

1. Degree of correlation cannot be predicted.
2. It gives only a rough idea.
3. The method is useful only when number of terms is small.

**ii. Simple Graph method of correlation**

In this method separate curves are drawn for separate series on a graph paper. By examining the direction and closeness of the two curves we can infer whether or not variables are related. If both the curves are moving in the same direction correlation is said to be positive. On the other hand, if the curves are moving in the opposite directions is said to be negative.

**Merits:**

1. It is easy to plot
2. Simple to understand
3. Abnormal values can easily be deducted.



## Demerits

1. This method is useless when number of terms is very big.
2. Degree of correlation cannot be predicted.

## B. Mathematical Method:

### i. Karl Pearson's Coefficient of Correlation.

Karl Pearson, a great biometrician and statistician, introduced a mathematical method for measuring the magnitude of linear relationship between two variables. This method is most widely used in practice. This method is known as Pearsonian Coefficient of Correlation. It is denoted by the symbol ' $r$ '; the formula for calculating Pearsonian  $r$  is:

$$(i) r = \frac{\text{Covariance of } xy}{\sigma_x \times \sigma_y}, (ii) r = \frac{\Sigma xy}{N \sigma_x \times \sigma_y}, (iii) r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$x = (x - \bar{X}), y = (y - \bar{Y})$$

$\sigma_x = \text{Standard deviation of series } x$

$\sigma_y = \text{Standard deviation of series } y$

The value of the coefficient of correlation shall always lie between **+1 and -1**.

When  $r = +1$ , then there is perfect positive correlation between the variables.

When  $r = -1$ , then there is perfect negative correlation between the variables.

When  $r = 0$ , then there is no relationship between the variables.

**Illustration 1:** Calculate Karl Pearson coefficient of correlation from the following data:

<b>X</b>	100	101	102	102	100	99	97	98	96	95
<b>Y</b>	98	99	99	97	95	92	95	94	90	91



**Solution: Calculation of coefficient of correlation**

X	$x = X - \bar{X}$ $= X - 99$	$x^2$	Y	$y = Y - \bar{Y}$ $= Y - 95$	$y^2$	xy
100	1	1	98	3	9	3
101	2	4	99	4	16	8
102	3	9	99	4	16	12
102	3	9	97	2	4	6
100	1	1	95	0	0	0
99	0	0	92	-3	9	0
97	-2	4	95	0	0	0
98	-1	1	94	-1	1	1
96	-3	9	90	-5	25	15
95	-4	16	91	-4	16	16
$\Sigma X = 990$		$\Sigma x^2 = 54$	$\Sigma Y = 950$		$\Sigma y^2 = 96$	$\Sigma xy = 61$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{990}{10} = 99;$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{950}{10} = 95;$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{61}{\sqrt{54 \times 96}} = \frac{61}{\sqrt{5184}} = \frac{61}{72} = +0.85$$

**Illustration 2:** Calculate Karl Pearson coefficient of correlation from the following data:

<b>X:</b>	6	2	10	4	8
<b>Y:</b>	9	11	5	8	7

**Solution: Calculation of coefficient of correlation**

X	$X^2$	Y	$Y^2$	XY
6	36	9	81	54
2	4	11	121	22
10	100	5	25	50
4	16	8	64	32





8	64	7	49	56
$\sum X = 30$	$\sum X^2 = 220$	$\sum Y = 40$	$\sum Y^2 = 340$	$\sum XY = 214$

$$r = \frac{N\sum XY - (\sum X \sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2} \times \sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$= \frac{(5 \times 214) - (30 \times 40)}{\sqrt{5 \times 220 - (30)^2} \times \sqrt{5 \times 340 - (40)^2}}$$

$$= \frac{1070 - 1200}{\sqrt{1100 - 900} \times \sqrt{1700 - 1600}}$$

$$= \frac{-130}{\sqrt{200} \times \sqrt{100}}$$

$$r = -0.92$$

\*\*\*\*\*

## RANK CORRELATION COEFFICIENT

### Spearman's rank correlation co-efficient.

In 1904, a famous British psychologist Charles Edward Spearman found out the method of ascertaining the coefficient of correlation by ranks. This method is based on rank. Rank correlation is applicable only to individual observations. This measure is useful in dealing with qualitative characteristics such as intelligence, beauty, morality, character, etc.,

The formula for Spearman's rank correlation which is denoted by P is;

$$P = 1 - \frac{6\sum D^2}{N(N^2-1)} \text{ or } P = 1 - \frac{6\sum D^2}{(N^3-N)}$$

Where, P = Rank co-efficient of correlation

D = Difference of the two ranks

$\sum D^2$  = Sum of squares of the difference of two ranks



$N$  = Number of paired observations

Like the Karl Pearson's coefficient of correlation, the value of  $P$  lies between  $+1$  and  $-1$ .

### Where ranks are given

**Illustration 3:** Following are the rank obtained by 10 students in two subjects, Statistics and Mathematics. To what extent the knowledge of the students in the two subjects is related?

<b>Statistics</b>	1	2	3	4	5	6	7	8	9	10
<b>Mathematics</b>	2	4	1	5	3	9	7	10	6	8

**Solution:** Calculation of Pearman's rank correlation coefficient

Rank of Statistics ( $x$ )	Rank of Mathematics ( $y$ )	$D = x - y$	$D^2$
1	2	-1	1
2	4	-2	4
3	1	+2	4
4	5	-1	1
5	3	+2	4
6	9	-3	9
7	7	0	0
8	10	-2	4
9	6	+3	9
10	8	+2	4
<b><math>N = 10</math></b>			<b><math>\Sigma D^2 = 40</math></b>

$$P = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$P = 1 - \frac{6 \times 40}{10(10^2-1)}$$



$$= 1 - \frac{240}{10(100 - 1)}$$

$$\mathbf{P} = 1 - \frac{240}{990} = 1 - 0.24 = + \mathbf{0.76}$$

**Where ranks are not given**

**Illustration 4:** A random sample of 5 college students is selected and their grades in Mathematics and Statistics are found to be:

<b>Mathematics</b>	85	60	73	40	90
<b>Statistics</b>	93	75	65	50	80

**Solution: Calculation of Pearman's rank correlation coefficient**

Marks in Mathematics (x)	Rank x	Marks in Statistics (y)	Rank y	D = x - y	D <sup>2</sup>
85	2	93	1	+1	1
60	4	75	3	+1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1
					<b>ΣD<sup>2</sup> = 4</b>

$$\mathbf{P} = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

$$\mathbf{P} = 1 - \frac{6 \times 4}{5(5^2 - 1)} = 1 - \frac{24}{5(25 - 1)} = 1 - \frac{24}{5(24)}$$

$$= 1 - \frac{1}{5} = 1 - 0.2 = + \mathbf{0.8}$$

**Equal or Repeated Ranks**

When two or more items have equal values, it is difficult to give ranks to them. In that case the items are given the average of the ranks they would have received, if they are not tied. A slightly different formula is used when there is more than one item having the same value.



$$P = 1 - 6 \left\{ \frac{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \dots}{N^3 - N} \right\}$$

$m$  = the number of items whose ranks are common

**Illustration 5:** From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

<b>X</b>	48	33	40	9	16	16	65	24	16	57
<b>Y</b>	13	13	24	6	15	4	20	9	6	19

**Solution: Calculation of Pearman's rank correlation coefficient**

<b>X</b>	<b>Rank x</b>	<b>Y</b>	<b>Rank y</b>	<b>D = R(x) - R(y)</b>	<b>D<sup>2</sup></b>
48	8	13	5.5	2.5	6.25
33	6	13	5.5	0.5	0.25
40	7	24	10	-3.0	9.00
9	1	6	2.5	-1.5	2.25
16	3	15	7	4.0	16.00
16	3	4	1	2.0	4.00
65	10	20	9	1.0	1.00
24	5	9	4	1.0	1.00
16	3	6	2.5	0.5	0.25
57	9	19	8	1.0	1.00
					<b><math>\Sigma D^2 = 41</math></b>

$$P = 1 - 6 \left\{ \frac{\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \dots}{N^3 - N} \right\}$$

$$= 1 - 6 \left\{ \frac{41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) \dots}{10^3 - 10} \right\} = 1 - \left\{ \frac{6(41 + 2 + 0.5 + 0.5)}{990} \right\}$$

$$= 1 - \left\{ \frac{264}{990} \right\} = 1 - 0.267 = + \mathbf{0.733}$$



### Merits:

1. It is simple to understand and easier to apply.
2. It can be used to any type of data, qualitative or quantitative.
3. It is the only method that can be used where we are given the ranks and not the actual data.
4. Even where actual data are given, rank method can be applied for ascertaining correlation by assigning the ranks to each data.

### Demerits:

1. This method is not useful to find out correlation in a grouped frequency distribution.
2. For large samples it is not convenient method. If the items exceed 30 the calculations become quite tedious and require a lot of time.
3. It is only an approximately calculated measure as actual values are not used for calculations.

## CONCURRENT DEVIATION METHOD

The method of studying correlation is the simplest of all the methods. The only thing that is required under this method is to find out the direction of change of X variable and Y variable. This is based on the signs of deviations; for each term the change is the value of the variable from its preceding or previous value which may be plus (+) or minus (-).

The formula is:

$$r_{(c)} = \pm \sqrt{\frac{2C-N}{N}}$$

Where,  $r_{(c)}$  = Coefficient of correlation by the concurrent deviation method

C = Number of concurrent deviations ( $C = \sum dx dy$ )

N = Number of pairs of deviation compared



**Illustration 6:** Calculate the coefficient of concurrent deviations from the data given below:

Month	Jan.	Feb.	Mar.	Apr.	May.	June	July	Aug.	Sep.
Supply	160	164	172	182	166	170	178	192	186
Price	292	280	260	234	266	254	230	190	200

**Solution:** Calculation of coefficient of concurrent deviations

Month	Supply x	dx	Price y	dy	dx dy
Jan.	160		292		
Feb.	164	+	280	-	-
Mar.	172	+	260	-	-
Apr.	182	+	234	-	-
May	166	-	266	+	-
June	170	+	254	-	-
July	178	+	230	-	-
Aug.	192	+	190	-	-
Sept.	186	-	200	+	-
N = 8					C = $\sum dx dy = 0$

$$\begin{aligned}
 r_{(c)} &= \pm \sqrt{\frac{2C - N}{N}} \\
 &= \pm \sqrt{\frac{2 \times 0 - 8}{8}} \\
 &= \pm \sqrt{\frac{-8}{8}} = \pm \sqrt{\frac{-1}{1}} \\
 &= \pm \sqrt{-(-1)} \\
 &= \pm \sqrt{(+1)} = -1
 \end{aligned}$$

\*\*\*\*\*



## REGRESSION

The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called Regression. The dictionary meaning of the word regression is 'return' or 'going back'. In 1877, Sir Francis Galton, first introduced the word 'regression'. The tendency to regression or going back was called by Galton as the 'Line of Regression'. The line describing the average relationship between two variables is known as the line of regression. The regression analysis confined to the study of only two variables at a time is termed as simple regression. The regression analysis for studying more than two variables at a time is known as multiple regressions.

### Definitions

According to Blair, "Regression is the measure of the average relationship between two or more variables in terms of the original units of the data".

According to Taro Yamane, "One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related causally, is regression analysis".

### Uses of Regression Analysis

The following are the uses of regression analysis

1. It is useful to estimate the relationship between two variables.
2. It is useful for prediction of unknown value.
3. It is widely used in social sciences like economics, natural and physical sciences.
4. It is useful to forecast the business situations.
5. It is useful to estimate thereof in sampling.
6. It is useful to calculate correlation co-efficient and co-efficient of determination.



## Differentiate Correlation from Regression.

<b>Correlation</b>	<b>Regression</b>
It is the relationship between two or more variable.	It is a mathematical measure showing the average relationship between two variables.
Both x and y are random variables.	Here x is a random variable and y is a fixed variable.
It gives limited information after verifying the relationship between the variables	It is used for the prediction of one value, in relationship to the other given value.
The range of relationship between -1 and +1.	Regression value is an absolute figure.
It studies linear relationship between the variables.	It studies linear and non-linear relationship between the variables.
If the coefficient of correlation is positive, then the two variables are positively correlated and vice versa.	The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable.

## Regression lines

If we take two variables X and Y we have two regression lines:

- i) Regression of X on Y and
- ii) Regression of Y on X

The regression line of X on Y gives the most probable value of X for any given value of Y. The regression of Y on X gives the most probable value of Y for any given value of X. There are two regression lines in the case of two variables.

## Regression Equations

The algebraic expressions of the two regression lines are called regression equations.





### Regression Equation of X on Y:

$$X_c = a + by$$

To determine the values of 'a' and 'b', the following two normal equations are to be solved simultaneously.

$$\sum X = Na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2$$

### Regression Equation of Y on X:

$$Y_c = a + bx$$

To determine the value of 'a' and 'b', the following two normal equations are to be solved simultaneously.

$$\sum Y = Na + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2$$

We can call these equations as normal equations.

**Illustration 1:** Determine the two regression equations of a straight line which best fits the data.

<b>X</b>	10	12	13	16	17	20	25
<b>Y</b>	10	22	24	27	29	33	37

### Solution: Calculation of Regression

<b>X</b>	<b>X<sup>2</sup></b>	<b>Y</b>	<b>Y<sup>2</sup></b>	<b>XY</b>
0	100	10	100	100
12	144	22	484	264
13	169	24	576	312
16	256	27	729	432
17	289	29	841	493



20	400	33	1089	660
25	625	37	1369	925
$\Sigma X = 113$	$\Sigma X^2 = 1983$	$\Sigma Y = 182$	$\Sigma Y^2 = 5188$	$\Sigma XY = 3186$

### Regression Equation of Y on X:

The two normal equations are:

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

Substituting the values,

$$N = 7; \quad \Sigma X = 113; \quad \Sigma X^2 = 1983; \quad \Sigma Y = 182; \quad \Sigma XY = 3186;$$

$$7a + 113b = 182 \quad \dots(1)$$

$$113a + 1983b = 3186 \quad \dots(2)$$

Multiplying (1), by 113,

$$791a + 12769b = 20566 \quad \dots(3)$$

Multiplying (2), by 7,

$$791a + 13881b = 22302 \quad \dots(4)$$

Subtracting (4) from (3)

$$- 1112 b = - 1736$$

$$b = \frac{-1736}{-1112} \Rightarrow b = \mathbf{1.56}$$

Put  $b = 1.56$  in (1) we get

$$7a + 113(1.56)b = 182$$

$$7a + 176.28 = 182 \Rightarrow 7a = 5.72$$



$$a = \frac{5.72}{7} \Rightarrow a = \mathbf{0.82}$$

The equation of straight line is  $Y_c = a + bX$

Put  $a = 0.82$ ,  $b = 1.56$

$\therefore$  The equation of the required straight line is  $Y_c = 0.82 + 1.56 X$

This is called regression of y on x

### Regression Equation of X on Y:

The two normal equations are:

$$\sum X = Na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2$$

Substituting the values,

$$N = 7; \quad \sum X = 113; \quad \sum Y^2 = 5188; \quad \sum Y = 182; \quad \sum XY = 3186;$$

$$7a + 182b = 113 \quad \dots(1)$$

$$182a + 5188b = 3186 \quad \dots(2)$$

Multiplying (1), by 182,

$$1274a + 33124b = 20566 \quad \dots(3)$$

Multiplying (2), by 7,

$$1274a + 36316b = 22302 \quad \dots(4)$$

Subtracting (4) from (3)

$$3192b = 1736$$

$$b = \frac{1736}{3192} \Rightarrow b = \mathbf{0.54}$$



Put  $b = 0.54$  in (1) we get

$$7a + 182(0.54)b = 113$$

$$7a + 98.28 = 113 \Rightarrow 7a = 14.72$$

$$a = \frac{14.72}{7} \Rightarrow a = 2.1$$

The equation of straight line is  $X_c = a + bY$

Put  $a = 2.1$ ,  $b = 0.54$

$\therefore$  The equation of the required straight line is  $X_c = 2.1 + 0.54 Y$

This is called regression of  $x$  on  $y$

### **Deviation taken from Actual Means:**

Regression equation of  $X$  on  $Y$ :

$$X - \bar{X} = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

Where,  $X$  = the value of  $x$  to be estimated for the given  $y$  value.  $\bar{X}$  = Mean value of  $X$  variable.  $Y$  = the value of  $y$  given in the problem;  $\bar{Y}$  = Mean value of  $y$  variables.

$$r \frac{\sigma_x}{\sigma_y} = \frac{\sum xy}{\sum y^2} = \text{Regression co-efficient of } X \text{ on } Y.$$

$$x = X - \bar{X}; y = Y - \bar{Y}$$

### **Regression equation of $Y$ on $X$ :**

$$Y - \bar{Y} = r \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

$$r \frac{\sigma_y}{\sigma_x} = \frac{\sum xy}{\sum x^2} = \text{Regression co-efficient of } Y \text{ on } X.$$



**Illustration 2:** Find regression lines from the following data:

X	3	5	6	8	9	11
Y	2	3	4	6	5	10

And also estimate Y when X is 15.

**Solution: Calculation of Regression Equations (by actual mean)**

X	$x = X - \bar{X}$	$x^2$	Y	$y = Y - \bar{Y}$	$y^2$	xy
3	-4	16	2	-3	9	12
5	-2	4	3	-2	4	4
6	-1	1	4	-1	1	1
8	1	1	6	1	1	1
9	2	4	5	0	0	0
11	4	16	10	5	25	20
$\Sigma X = 42$	$\Sigma x = 0$	$\Sigma x^2 = 42$	$\Sigma Y = 30$	$\Sigma y = 0$	$\Sigma y^2 = 40$	$\Sigma xy = 38$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{42}{6} = 7$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{30}{6} = 5$$

**Regression equation of X on Y:**

$$X - \bar{X} = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

$$r \frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{38}{40} = 0.95$$

$$X - 7 = 0.95(Y - 5)$$

$$X - 7 = 0.95Y - 4.75$$

$$X = 0.95Y + 2.25$$

**Regression equation of Y on X:**

$$Y - \bar{Y} = r \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

$$r \frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{38}{42} = 0.90$$

$$Y - 5 = 0.90(X - 7)$$

$$Y - 5 = 0.90X - 6.30$$

$$Y = 0.90X - 1.30$$

When X is 15, Y will be,  $Y = 0.90 \times 15 - 1.30$

$$= 13.5 - 1.30$$



$$Y = 14.8$$

**Deviation taken from the Assumed Mean:**

**Regression equation of X on Y:**

$$X - \bar{X} = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$$

$$\text{Where, } r \frac{\sigma_x}{\sigma_y} = \frac{N \sum dx dy - (\sum dx)(\sum dy)}{N \sum dy^2 - (\sum dy)^2}$$

$$dx = X - A; dy = Y - A; (A = \text{assumed mean})$$

**Regression equation of Y on X :**

$$Y - \bar{Y} = r \frac{\sigma_y}{\sigma_x} (X - \bar{X})$$

$$r \frac{\sigma_x}{\sigma_y} = \frac{N \sum dx dy - (\sum dx)(\sum dy)}{N \sum dx^2 - (\sum dx)^2}$$

**Illustration: 3.** Find regression lines from the following data:

<b>X</b>	40	38	35	42	30
<b>Y</b>	30	35	40	36	29

Also calculate Karl Pearson's coefficient of correlation.

**Solution: Calculation of Regression Equations (by assumed mean)**

<b>X</b>	<b>dx = X - A</b>	<b>dx<sup>2</sup></b>	<b>Y</b>	<b>dy = Y - A</b>	<b>dy<sup>2</sup></b>	<b>dx.dy</b>
40	5	25	30	0	0	0
38	3	9	35	5	25	15
35	0	0	40	10	100	0
42	7	49	36	6	36	42
30	-5	25	29	-1	1	5
	<b>∑ dx = 10</b>	<b>∑ dx<sup>2</sup> = 108</b>		<b>∑ dy = 1020</b>	<b>∑ dy<sup>2</sup> = 162</b>	<b>dx.dy = 62</b>

$$\bar{X} = A \pm \frac{\sum dx}{N} = 35 \pm \frac{10}{5} = 37$$

$$\bar{Y} = A \pm \frac{\sum dy}{N} = 30 \pm \frac{20}{5} = 34$$



Regression equation of X on Y

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$

$$b_{xy} = \frac{N \sum dx dy - (\sum dx)(\sum dy)}{N \sum dy^2 - (\sum dy)^2}$$

$$= \frac{5 \times 62 - (10)(20)}{5 \times 162 - (20)^2}$$

$$= \frac{310 - 200}{810 - 400} = \frac{110}{410}$$

$$b_{xy} = 0.27$$

$$X - 37 = 0.27(Y - 34)$$

$$X - 37 = 0.27Y - 9.18$$

$$X = 0.27Y + 27.82$$

Regression equation of Y on X

$$Y - \bar{Y} = b_{yx} (X - \bar{X})$$

$$b_{yx} = \frac{N \sum dx dy - (\sum dx)(\sum dy)}{N \sum dx^2 - (\sum dx)^2}$$

$$= \frac{5 \times 62 - (10)(20)}{5 \times 108 - (10)^2}$$

$$= \frac{310 - 200}{540 - 100} = \frac{110}{440}$$

$$b_{yx} = 0.25$$

$$Y - 34 = 0.25(X - 37)$$

$$Y - 34 = 0.25X - 9.25$$

$$Y = 0.25X + 24.75$$

$$r = \sqrt{b_{xy} \times b_{yx}}$$

$$= \sqrt{2.2 \times 0.37} = \sqrt{0.814} = 0.9$$

**Illustration 4:** Given the following data, calculate the expected value of Y when X = 12.

	X	Y
Arithmetic Mean ( $\bar{X}$ )	7.6	14.8
Standard Deviation ( $\sigma$ )	3.6	2.5
Coefficient of correlation ( $r$ ) = 0.99		

**Solution:**

**Regression of Y on X**

$$Y - \bar{Y} = r \frac{\sigma_x}{\sigma_y} (X - \bar{X})$$

$$Y - 14.8 = 0.99 \times \frac{2.5}{3.6} (X - 7.6)$$

$$Y - 14.8 = 0.688 (X - 7.6)$$

$$Y - 14.8 = 0.688 X - 5.23$$

$$Y = 0.688 X - 5.23 + 14.8$$



$$Y = 0.688 X + 9.57$$

$$\text{When } X = 12 \Rightarrow Y = 0.688 (12) + 9.57 = 17.826$$

Hence the expected value of Y is 17.83.

\*\*\*\*\*

### Theoretical Questions

1. Define and explain the uses of correlation.
2. Define and explain the merits & demerits of correlation.
3. Explain the types of correlation.
4. Describe the methods of correlation with relative merits and demerits.
5. Define and explain the uses of Regression.
6. Define and explain the merits & demerits of Regression.
7. Explain the types of Regression.
8. Differences between Correlation and Regression.

### Practical Problems

1. Calculate Pearson's coefficient of correlation between advertisement cost and sales from the following data:

Advertisement Cost('000) Rs.	39	65	62	90	82	75	25	98	36	78
Sales ('00,000)	47	53	58	86	62	68	60	91	51	84

$$(r = + 0.78)$$

2. Compute the coefficient of correlation of the following score of A and B.

A:	5	10	5	11	12	4	3	2	7	1
B:	1	6	2	8	5	1	4	6	5	2

$$(r = + 0.58)$$

3. Ten competitors in a voice contest are ranked by 3 judges in the following orders:

I Judge	1	6	5	10	3	2	4	9	7	8
II Judge	3	5	8	4	7	10	2	1	6	9
III Judge	6	4	9	8	1	2	3	10	5	7





Use the rank correlation to gauge which pair of judges have the nearest approach to common likings in voice.

$$(I \text{ \& II} = - 0.212; II \text{ \& III} = - 0.297; I \text{ \& III} = + 0.636)$$

4. Calculate the rank correlation coefficient for the following table of marks of students in two subjects:

Major I	80	64	54	49	48	35	32	29	20	18	15	10
Major II	36	38	39	41	27	43	45	52	51	42	40	52

$$(r = - 0.685)$$

5. The following table gives the score obtained by 11 students in Mathematics and Statistics. Find the rank correlation coefficient.

Mathematics	40	46	54	60	70	80	82	85	85	90	95
Statistics	45	45	50	43	40	75	55	72	65	42	70

$$(r = + 0.36)$$

6. Calculate the coefficient of concurrent deviation from the data given below:

Year	1976	1977	1978	1979	1980	1981	1982	1983	1984
Supply	160	164	172	182	166	170	178	192	186
Price	292	290	260	234	266	254	230	130	200

$$(r = -1)$$

7. Obtain the equations of lines of regression between the indices.

X:	78	77	85	88	87	82	81	77	76	83	97	93
Y:	84	82	82	85	89	90	88	92	83	89	98	99

$$(X = 0.79 Y + 13.82; Y = 0.59 X + 39.05)$$

8. Calculate the two regression equations of X on Y and Y on X from the data given below, taking deviations from actual means of X and Y.

Price (Rs.)	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs. 20.

$$(X = -0.12 Y + 17.92; Y = -0.25 X + 44.25; Y = 39.25)$$

9. The correlation coefficient between two variable X and Y is  $r = 0.6$ . If  $\sigma_x = 1.5$ ,  $\sigma_y = 2.0$ ,  $\bar{X} = 10$  and  $\bar{Y} = 20$ , find the regression lines of Y on X and X on Y.

$$(X = 0.45 Y + 1; Y = 0.8 X + 12)$$



10. By using the following data. Find out the two lines regression and from them compute the Karl Pearson's Coefficient of Correlation:

$$\sum X = 250; \sum Y = 300; \sum XY = 7900; \sum X^2 = 6500; \sum Y^2 = 10,000; N = 10.$$

$$(r = -0.8)$$

11. Given the following data, estimate the marks in mathematics obtained by a student who has scored 60 marks in statistics.

Mean marks of mathematics	80
Mean marks of statistics	50
S.D marks in mathematics	15
S.D marks in statistics	10
Coefficient of correlation	0.4

$$(X = 0.6 Y + 50; X = 86.)$$

\*\*\*\*\*

## INDEX NUMBERS

An index number is a number which is used to measure the level of a certain phenomenon as compared to the level of the same phenomenon at some standard period. It is a statistical device for comparing the general level of magnitude of a group of related variables in two or more situations. It is a number which indicates the change in magnitude.

### Definitions

According to Croxton and Cowden, "Index numbers are devices for measuring differences in the magnitude of a group of related variables".

According to Spiegel, "An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic location or other characteristics".

According to Maslow, "It is a numerical value characterizing the change in complex economic phenomena over a period of time or space".



## Characteristics of Index numbers

1. Index numbers are specialized averages.
2. Index numbers measure the net change in a group of related variables.
3. Index numbers measure changes not capable of direct measurement.
4. Index numbers are for comparison.
5. Index numbers are expressed in terms of percentages so as to show the extent of change. However, percentage sign (%) is never used.

## Uses of Index Numbers

1. Index numbers are most widely used statistical devices.
2. Index numbers are used to measure the relative changes.
3. Index numbers widely used in the evaluation of business and economic conditions.
4. Index numbers are useful for better standard of living comparison.
5. Index numbers help in framing suitable policies
6. Index numbers are useful to know trends and tendencies.
7. Index numbers measure the purchasing power of money.
8. Index numbers are the wage adjuster.
9. Index numbers are useful in deflation.
10. Index numbers are used as guides.

## Types of Index Numbers

There are various types of index numbers. Some of which are follows.

### 1. Price Index:

It is used for measuring the value of money. It is an index number which compares the prices for a group of commodities at a certain time or at a place with prices of a base period.

There are two types of price index namely,

- i. Wholesale Price index numbers and
- ii. Retail Price index numbers.



The changes in the general price level of a country are called wholesale price index.

The changes in the retail prices of commodities such as, consumption goods, bank deposits, bonds, etc., are called retail price index.

## 2. Quantity Index:

The changes in the volume of goods produced or consumed are studied in quantity index numbers.

## 3. Value Index:

The total value of a certain period is compared with the total value of the base period is called value index numbers.

### Methods of construction of index numbers

The various methods of construction of index numbers are as follows:

#### I. Unweighted Index Numbers.

##### 1. Simple Aggregative method

This is the simplest method of constructing the index numbers. The prices of the different commodities of the current year are added and the total is divided by the sum of the prices of base year commodity and multiplied by 100:

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

**Illustration 1:** Construct an index number for 2016 taking 2015 as base from the following data.

Commodity	Price in 2015 (Rs.)	Price in 2016 (Rs.)
A	50	60
B	40	80
C	70	110
D	90	70
E	50	40



### Solution: Calculation of index number

Commodity	Price in 2015 (Rs.)	Price in 2016 (Rs.)
A	50	60
B	40	80
C	70	110
D	90	70
E	50	40
	$\Sigma P_0 = 300$	$\Sigma P_1 = 360$

$$\text{Price Index, } P_{01} = \frac{\Sigma P_1}{\Sigma P_0} \times 100 = \frac{360}{300} \times 100 = 120$$

This means that as compared to 2015, in 2016 there is a net increase in the prices of commodities to the extent of 20%.

### 2. Simple Average of Price relative method:

In this method, the price relative of each item is calculated separately and then averaged.

$$\text{Arithmetic Mean of Price Relatives: } P_{01} = \frac{\Sigma P}{N}$$

$$\text{Geometric Mean of Price Relatives: } P_{01} = \text{antilog} \frac{\Sigma \log P}{N}$$

$$\text{Where, } P = \frac{P_1}{P_0} \times 100; N = \text{Number of items.}$$

**Illustration 2:** Compute a price index for the following by using both arithmetic mean and geometric mean:

Commodity	A	B	C	D	E	F
Price in 2014 (Rs.)	20	30	10	25	40	50
Price in 2015 (Rs.)	25	30	15	35	45	55



### Solution: Calculation for Price Index

Commodity	Price in 2014 $P_0$	Price in 2015 $P_1$	Price Relative $P = \frac{P_1}{P_0} \times 100$	Log P
A	20	25	125	2.0960
B	30	30	100	2.0000
C	10	15	150	2.1761
D	25	35	140	2.1461
E	40	45	112.5	2.0511
F	50	55	110	2.0414
<b>N = 6</b>			<b><math>\Sigma P = 737.5</math></b>	<b><math>\Sigma \log P = 12.5116</math></b>

$$\text{Arithmetic Mean of Price Relatives: } P_{01} = \frac{\Sigma P}{N} = \frac{737.5}{6} = 122.92$$

$$\text{Geometric Mean of Price Relatives: } P_{01} = \text{Antilog } \frac{\Sigma \log P}{N}$$

$$= \text{Antilog } \frac{12.5116}{6}$$

$$= \text{Antilog } 2.0853$$

$$= 121.7.$$

## II. Weighted Index Number

**Weighted Aggregative Index Numbers:** Under this method weights are assigned to various items and a number of formulae have been used. The following are some of the methods, generally used.

**A. Laspeyre's method:** In this method, the base year quantities are taken as weights:

$$P_{01} (La) = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

**B. Paasche's method:** In this method, current year quantities are taken as weights:

$$P_{01} (Pa) = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times 100$$



**C. BowleyDorfish method:** This is an index number got by the arithmetic mean of Laspeyre's and Paasche's methods. This method takes into account both the current and the base periods.

$$P_{01(B)} = \frac{L+P}{2}$$

Where, L = Laspeyre's method; P = Paasche's method

**D. Fisher's ideal Method:** Fisher's price index number is given by the geometric mean of Laspeyre's and Paasche's formula;

$$P_{01(F)} = \sqrt{L \times P}$$

**E. Marshal Edge worth Method:** Under this method, the arithmetic mean of base year and current year quantities are taken as weights ( $w = \frac{q_0 + q_1}{2}$ )

$$P_{01(Ma)} = \frac{\sum p_1 q_0}{\sum p_0 q_0} + \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

**F. Kelly's method:** Kelly's method number uses quantities of some period (which is neither the base period nor the current period) as weights. This weight is kept constant for all periods. If we denote it by q, then Kelly's method number is given by

$$P_{01(K)} = \frac{\sum P_1 q}{\sum P_0 q} \times 100$$

**G. Walsch's method:** Under this method, the geometric mean of the base year and current year quantities as weights.

$$P_{01(Wa)} = \frac{\sum P_1 \sqrt{q_0 q_1}}{\sum P_0 \sqrt{q_0 q_1}} \times 100$$

**Illustration 3:** Calculate index number from the following data through a) Laspeyre's method b) Paasche's method c) Bowley's method d) Fisher's ideal formula method and e) Marshall Edge worth method:

	Base Year		Current Year	
	Kilo	Rate (Rs.)	Kilo	Rate (Rs.)
<b>Bread</b>	10	3	8	3.25
<b>Meat</b>	20	15	15	20
<b>Tea</b>	2	25	3	23



### Solution: Calculation of Index Numbers

	Base Year		Current Year		$P_1Q_0$	$P_0Q_0$	$P_1Q_1$	$P_0Q_1$
	Kilo	Rate	Kilo	Rate				
	$Q_0$	$P_0$	$Q_1$	$P_1$				
Bread	10	3	8	3.25	32.50	30	26	24
Meat	20	15	15	20	400.00	300	300	225
Tea	2	25	3	23	46.00	50	69	75
<b>Total</b>					$\sum P_1Q_0 = 478.50$	$\sum P_0Q_0 = 380$	$\sum P_1Q_1 = 395$	$\sum P_0Q_1 = 324$

a) Laspeyre's method:  $P_{01(La)} = \frac{\sum P_1Q_0}{\sum P_0Q_0} \times 100 = \frac{478.50}{380} \times 100 = 125.9$

b) Paasche's method:  $P_{01(Pa)} = \frac{\sum P_1Q_1}{\sum P_0Q_1} \times 100 = \frac{395}{324} \times 100 = 121.9$

c) BowleyDorfis method:  $P_{01(B)} = \frac{L+P}{2} = \frac{125.9+121.9}{2} = 123.9$

d) Fisher's ideal Method:  $P_{01(F)} = \sqrt{L \times P} = \sqrt{125.9 \times 121.9}$

$= \sqrt{15347.21} = 123.9$

e) Marshal Edge worth Method:  $P_{01(Ma)} = \frac{\sum p_1q_0}{\sum p_0q_0} + \frac{\sum p_1q_1}{\sum p_0q_1} \times 100$

$$= \frac{478.5}{380} + \frac{395}{324} \times 100 = \frac{873.5}{704} \times 100 = 1.24 \times 100 = 124.$$

### 2. Weighted average of price relative.

$$P_{01} = \frac{\sum PV}{\sum V}; P_{01} = \text{Antilog} \frac{\sum V \log P}{\sum V}$$





**Illustration 4:** Compute a price index for the following by using both arithmetic mean and geometric mean:

Item	Price in 1976	Price in 1977	Quantity in 1976
Wheat	2	2.50	40 kg
Sugar	3	3.25	20 kg
Milk	1.50	1.75	10 lit.

**Solution: Construction of Index**

Item	$p_0$	$p_1$	$q_0$	$V = P_0q_0$	$P = p_1/p_0 \times 100$	PV	Log P	Log PV
Wheat	2	2.50	40	80	125	10000	2.0969	167.7520
Sugar	3	3.25	20	60	108.33	6499.8	2.0346	122.0760
Milk	1.50	1.75	10	15	116.67	1750.05	2.0669	31.0065
<b>Total</b>				$\Sigma V = 155$		$\Sigma PV = 18249.85$		$\Sigma \text{Log PV} = 320.8345$

$$P_{01} = \frac{\Sigma PV}{\Sigma V} = \frac{18249.85}{155} = 117.74$$

$$P_{01} = \text{Antilog} \frac{\Sigma V \log P}{\Sigma V} = \text{Antilog} \frac{320.8345}{155}$$

$$= \text{Antilog} 2.0699 = 111.4$$

**Test of consistency of Index Numbers**

### 1. Time Reversal Test

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}; P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$P_{01} \times P_{10} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{1} = 1$$

It shows that the Time Reversal Test is satisfied

### 2. Factor Reversal Test

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}; Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$



$$P_{01} \times Q_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}}$$

$$= \sqrt{\frac{\sum p_1 q_1}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_0}} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

**Illustration 5:** Compute Index Number, using Fishers Ideal Formula and show that it satisfies time – reversal test and factor – reversal test.

Commodity	Quantity	Base Year Price	Quantity	Current Year Price
A	12	10	15	12
B	15	7	20	5
C	24	5	20	9
D	5	16	5	14

**Solution: Computation of Index Number**

Commodity	q <sub>0</sub>	p <sub>0</sub>	q <sub>1</sub>	p <sub>1</sub>	p <sub>1</sub> q <sub>0</sub>	p <sub>0</sub> q <sub>0</sub>	p <sub>1</sub> q <sub>1</sub>	p <sub>0</sub> q <sub>1</sub>
A	12	10	15	12	144	120	180	150
B	15	7	20	5	75	105	100	140
C	24	5	20	9	216	120	180	100
D	5	16	5	14	70	80	70	80
					∑ p <sub>1</sub> q <sub>0</sub> = 505	∑ p <sub>0</sub> q <sub>0</sub> = 425	∑ p <sub>1</sub> q <sub>1</sub> = 530	∑ p <sub>0</sub> q <sub>1</sub> = 470

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100 = \sqrt{\frac{505}{425} \times \frac{530}{470}} \times 100 = \sqrt{1.188 \times 1.128} \times 100 = 115.8$$

**Test of consistency of Index Numbers**

### 1. Time Reversal Test

Time Reversal Test is satisfied when  $P_{01} \times P_{10} = \sqrt{1} = 1$

$$P_{01} \times P_{10} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum p_0 q_1}{\sum p_1 q_1} \times \frac{\sum p_0 q_0}{\sum p_1 q_0}} = \sqrt{1} = 1$$



$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{470}{530} \times \frac{425}{505}}$$

$$P_{01} \times P_{10} = \sqrt{\frac{505}{425} \times \frac{530}{470} \times \frac{470}{530} \times \frac{425}{505}} = \sqrt{1} = 1$$

## 2. Factor Reversal Test

Factor Reversal Test is satisfied when  $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100; Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma p_0 q_0}{\Sigma q_0 p_1}} \times 100$$

$$P_{01} \times Q_{01} = \sqrt{\frac{505}{425} \times \frac{530}{470} \times \frac{470}{425} \times \frac{530}{505}} = \frac{530}{425} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

∴ Hence the given data satisfies the time – reversal test and factor – reversal test.

## Single commodity case

### Link Relatives:

In this method, the relative for each year is found out from the prices of the preceding year. Here the base changes from year to year. The index numbers by using this method are called link relative index numbers or link relatives.

### Chain Base method:

Chain index is one in which the figures for each year are first expressed as percentages of the preceding year. These percentages are then chained together by successive multiplication.

The following formula is used to find out chain index:

$$\text{Chain Index} = \frac{\text{Link relative of the current year} \times \text{previous year chain index}}{100}$$

$$\text{Link relative} = \frac{\text{Current year price}}{\text{previous year price}} \times 100$$



### Advantages of chain base index

1. It is a great advantage to economists and businessmen.
2. Without any recalculation we can omit old items and introduce new items.
3. Weights can be adjusted as frequently as possible.
4. It is free from seasonal variations.

### Fixed Base index

In this method, no link relative method is used. Weights cannot be adjusted frequently. The base year does not change and it is suitable for long period only.

### How would you convert the Chain Base index to fixed base index?

The following steps are to be followed to convert the chain base index to fixed base index

1. First year is fixed as the chain base index and the same will be taken as the base which is 100.
2. For other years to find out the index numbers the following formula is used.

$$\text{Current year FBI} = \frac{\text{Current year C.B.I}}{\text{Previous year's FBI}} \times 100$$

### Base shifting

The change of reference base period is called shifting the base. For comparison, it is necessary to shift the base from one period to another period. We can express the series in terms of a recent period. It is called base shifting

$$\text{Index number} = \frac{\text{Current year's old index number}}{\text{New base year's old index number}} \times 100$$

### Differentiate chain base and fixed base.

	Chain Base	Fixed Base
1.	The base year changes.	The base year does not change.



2.	Here the link relative method is used.	No such link relative method is used.
3.	Introduction and deletion of items are easy to calculate without recalculation of the entire series.	Any change in the commodities, will involve the entire index number to be recast.
4.	It cannot be computed if data for any one year is missing.	There is no such problem.
5.	It is suitable for short period only.	It is suitable for long periods only.
6.	Index number wrong if an error is committed in the calculation of any link index number.	This is not so, the error is confined to the index of that year only.

### **Consumer price index (or) Cost of living index.**

Consumer price index is designed to measure the change in the cost of living of workers due to change in the retail price. A change in the price level affects the cost of living of the people. People consume different types of commodities. So there is need to construct consumer's price index. Consumer price index can be used in different places for many purposes.

### **Methods of construction of consumer price index.**

- a) Aggregate Expenditure method
- b) Family Budget method.

a) Aggregate expenditure method:  $P_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$

This method is based upon the Laspeyre's method.

b) Family Budget method:  $P_{01} = \frac{\sum PV}{\sum V}$

Where;  $P = \frac{P_1}{P_0} \times 100$ ;  $V = p_0 q_0$  (value weight)



## **Uses of cost of living index**

The following are the main uses of cost of live index:

1. It is useful or fixing the wages.
2. It is useful to know the purchasing power of money.
3. By using the cost of living index the Government decides about the price and other variables.
4. Dearness Allowances and Bonus are regulated with the help of this index.
5. It is useful for the analysis of price situation.

## **Limitations of index numbers**

1. If the chosen base year is not a normal one, all the purpose is lost.
2. Every index number has its own purpose. No index number can serve all the purpose.
3. These are only approximate indications of the relative level.

\*\*\*\*\*

## **ANALYSIS OF TIME SERIES**

An arrangement of statistical data in accordance with time of occurrence or in chronological order is called a time series. Thus when we observe numerical data at regular intervals of time the set of observation is known as time series. The regular intervals may be an hour, a day, a week a month, a year, a decade etc.,

### **Definitions**

According to Wessel and Welelt, “When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series”.

According to Morris Hamburg, “A time series is a set of observations arranged in chronological order”.



According to Cecil H. Maers, “A time series may be defined as a sequence or repeated measurements of a variable made periodically through time”.

According to Patterson, “A time series consist of statistical data which are collected, recorded or observed over successive increments”.

According to Ya-Luh-Chou, “A time series may be defined as a collection of magnitudes belonging to different time periods, of some variable or composite of variables, such as production of steel, per capital income, gross national product, price of tobacco or index of industrial production”.

### **Uses of time series**

Time series analysis is useful indifferent fields like economics, science, research work, etc because of the following reasons.

1. It helps in understanding the past behavior.
2. It helps in planning and forecasting the future operations.
3. It facilitates comparison between data of one period with another period.
4. It is useful not only to economists but also to the businessman.
5. It helps in evaluating current accomplishments.
6. It is useful for forecasting the trade cycles.

### **Types of variations or components or elements of time series.**

There are four basic types of variations and these are called the components or elements of time series.

(i) Secular trend (ii) Seasonal variation (iii) Cyclical variations (iv) Irregular variation.



### **i) Secular Trend:**

A secular or long-term trend refers to the movement of a series reflecting continuous growth or decline over a long period of time. There are many types of trend. Some trend rise upward and some trend fall downward.

The following are the types of trend.

- i. Linear (or) Straight Line Trend.
- ii. Non-Linear (or) Curvilinear Trend

If the values of time series are plotted on a graph and if it forms a straight line then it is called Linear (or) Straight Line Trend.

If the values of time series are plotted on a graph and if it forms a curve then it is called Non-linear (or) Curvilinear Trend.

### **ii) Seasonal variation:**

Seasonal variations are those periodic movements in business activities within the year, recurring periodically year after year. Generally, seasons variations appear at weekly, monthly or quarterly intervals.

#### **Causes:**

Seasonal variation may occur due to climate, weather conditions, customs, habits and traditions.

#### **Uses:**

Seasonal variation analysis is used to formulate correct policy decisions and of planning future operations. It is useful to businessmen, producers, agriculturist etc., to study effects of seasonal variations and to isolate them from the trend.





### **Cyclical variation:**

According to Linecoin L. Chou, “Up and down movements are different from seasonal fluctuations, in that they extend over longer period of time - usually two or more years”.

Business time series is influenced by the wave-like change of prosperity and depression. This up and down movement is known as cyclical variation. Cyclical variation analysis is helpful to businessmen for stabilizing the level of business activities and also useful to the economist for formulating suitable policies.

### **Irregular variation:**

Irregular variations refer to such variations in business activities which do not repeat in a definite pattern. They are also called ‘erratic’ accidental or random variations which are generally non-recurring and unpredictable. Irregular variations of time series are either random or caused by some sporadic forces such as war, flood, revolution, etc. It is usually a short-term one, but it will affect all the components of time series.

### **Methods of estimating the trend components**

There are four methods of estimating trend.

#### **1. The free hand Graphic method.**

In this method we must plot the original data on the graph. Draw a smooth curve carefully which will show the direction of the trend.

#### **2. Semi - Averages**

In this method the original data are divided into two equal parts and averages are calculated for both the parts. These averages are called Semi –Averages. Trend line is drawn with the help of this average.

#### **3. Moving average method :**

In this method, the average value for a number of years or months or weeks is taken into an account and placing it at the centre of the time span (period of moving average) and it is the normal or trend value for the middle period.

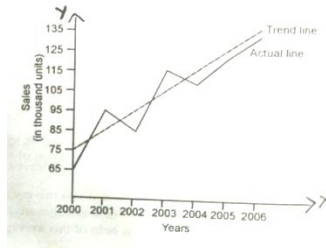


#### 4. Method of Least squares :

It is a mathematical as well as an analytical method. Under this method a straight line trend can be fitted to the given time series of data.

**Illustration 1:** Fit a trend line for the following data by the free - hand method.

<b>Year :</b>	2000	2001	2002	2003	2004	2005	2006
<b>Sales in units: (in '000s)</b>	65	95	85	115	110	120	130



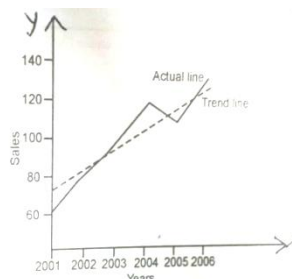
**Illustration 2:** Draw a trend line by the method of semi averages

<b>Year :</b>	2001	2002	2003	2004	2005	2006
<b>Sales in units : (in '000s)</b>	60	77	82	120	116	130

**Solution: Calculation of trend line using semi average method**

Year	Sales in units : (in '000s)
2001	60
2002	77 $219/3 = 73$
2003	82
2004	120
2005	116 $366/3 = 122$
2006	130

**Note:** If the even number of periods is given we can divide two equal parts. If the period is given in odd number of years, the value of the middle year is omitted.





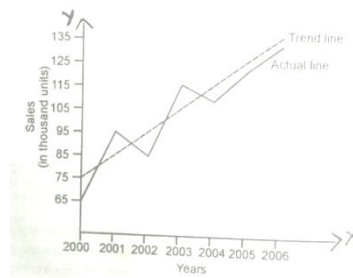
**Illustration 3:** Draw a trend line by the method of semi-averages

<b>Year :</b>	2000	2001	2002	2003	2004	2005	2006
<b>Sales in units : (in '000s)</b>	100	110	120	115	130	135	140

**Solution:**

### Calculation of trend line using semi average method

Year	Sales in units : (in '000s)
2000	100
2001	330/3 = 110
2002	120
2003	115 Left out the middle item
2004	130
2005	405/3 = 135
2006	140



### 3. Moving average method

In this method the average value for a number of years or months or weeks is taken into account. It is placed at the centre of the time - span. It is the normal or trend value for the middle period.

#### Calculation of moving averages:

The formula for calculating 3 yearly moving averages is

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}$$

The formula for 4 yearly moving averages is



$$\frac{a+b+c+d}{4}, \frac{b+c+d+e}{4}, \frac{c+d+e+f}{4}$$

The formula for 5 yearly moving averages is

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}$$

### Steps to calculate 3 yearly moving averages:

1. Compute the value of first three years and place the three year total against the middle year.
2. Leave the first year's value and compute the value of next three years and place the total against the middle year ie., against 3<sup>rd</sup> year.
3. This process should be continued until the last year's value is taken for calculation.
4. The three yearly total values are divided by 3 and the value placed in the next column. This value is called the Trend value of moving average.

### Steps for calculating of 4yearly moving average (or) Even period of moving average:

1. Computer the value of first four years and place the total value in between second and the third year.
2. Leave the first year's value and compute the value of next four years and place the total in between third and fourth year.
3. This process must be continued until the last year's value is taken for calculation.
4. Computer the first two four year totals and place the value against the middle of two four years totals.
5. Leave the first four years total and compute the next two four year's totals and place it against the middle of next four year.
6. This method must be continued until all the four year totals.
7. The two four yearly totals is divided by 8 and put the value in the next column. This value is called Trend value.



**Illustration 4:** Find the 3-yearly moving average from the following time series data.

<b>Year :</b>	1998	1999	2000	2001	2002	2003	2004	2005
<b>Sales in units : (in tones)</b>	30.1	45.4	39.3	41.4	42.2	46.4	46.6	49.2

**Solution:**

**Calculation of 3 yearly moving average method**

<b>Year</b>	<b>Sales (in tones)</b>	<b>3 yearly moving total</b>	<b>3 yearly moving value</b>
1998	30.1	-	-
1999	45.4	114.8	38.27
2000	39.3	126.1	42.03
2001	41.4	122.9	40.97
2002	42.2	130.0	43.33
2003	46.4	135.2	45.07
2004	46.6	142.2	47.40
2005	49.2	-	-

**Illustration 5:** Calculate the trend value by using three yearly moving averages of the following data.

<b>Year :</b>	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
<b>Production (in '000 tons)</b>	21	22	23	25	24	22	25	26	27	26

**Solution:** Calculation of trend value by using 3 yearly moving averages

<b>Year</b>	<b>Production (in '000 tons)</b>	<b>3 yearly moving total</b>	<b>3 yearly moving Average (Trend value)</b>
1990	21	-	-
1991	22	66	22.00
1992	23	70	23.33
1993	25	72	24.00
1994	24	71	23.67
1995	22	71	23.67
1996	25	73	24.33
1997	26	78	26.00
1998	27	79	26.33
1999	26	-	-



**Illustration 6:** Calculate the 5 yearly moving averages from the following data:

<b>Year :</b>	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
<b>No. of Students:</b>	705	685	703	687	705	689	715	685	725	730

**Solution: Calculation of 5 yearly moving average method**

Year	No. of students	5 yearly moving total	Moving Value
1997	705		
1998	683		
1999	703	3485	697.0
2000	687	3469	693.8
2001	705	3499	699.8
2002	689	3481	696.2
2003	715	3519	703.8
2004	685	3544	708.8
2005	725		
2006	730		

**Illustration 7:** Calculate the four – yearly moving average for the following data

<b>Year :</b>	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
<b>Production : (in '000 tons)</b>	464	515	518	467	502	540	557	571	586	612

**Solution: Calculation of 4 yearly moving average method**

Year	Production	4 Yearly Moving	Combined Total	Moving Average
1994	464			
1995	515			
		1964		
1996	518		3966	$3966/8 = 495.75$
		2002		
1997	467		4029	$4029/8 = 503.63$
		2027		
1998	502		4093	$4093/8 = 511.63$
		2066		
1999	540		4236	$4236/8 = 529.50$
		2170		
2000	557		4424	$4424/8 = 553.00$
		2254		
2001	571		4580	$4580/8 = 572.50$
		2326		
2002	586			
2003	612			



#### 4. Methods of least square

A Straight line trend can be fitted from the time series of given data. The trend line is called the line of best fit.

The sum of deviations of the actual values of Y and the trend value ( $Y_c$ ) is zero. And the sum of squares of deviations of actual value of Y and the trend value ( $Y_c$ ) is the least.

$$\text{i.e. } (Y - Y_c) = 0 \text{ and } (Y - Y_c)^2 = \text{least value}$$

This method is called the method of least square (or) the line of best fit. The method of least square can be used in the liner and non-linear trend

$$Y_c = a + bx;$$

$Y_c$  = required trend value; X = unit of time; a and b are constants

The constants a and b are calculated from the following two normal equations.

$$\Sigma Y = Na + b\Sigma X \quad (1)$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2 \quad (2)$$

Where N = Number of years or months

When  $x = 0$  (middle year taken as origin)

$$\Sigma X = 0$$

$$b\Sigma X = 0 \text{ and } a\Sigma X = 0$$

$$\text{From (1), } \Sigma Y = Na + 0 \rightarrow a = \frac{\Sigma Y}{N} \text{ or } \bar{Y}$$

$$\text{From (2), } \Sigma XY = 0 + b\Sigma X^2 \rightarrow b = \frac{\Sigma XY}{\Sigma X^2}$$

$$a = \frac{\Sigma Y}{N} \text{ and } b = \frac{\Sigma XY}{\Sigma X^2}$$



**Illustration 8:** Compute the trend from the following by the method of least squares:

<b>Year :</b>	2000	2001	2002	2003	2004
<b>Population (in Lakhs )</b>	830	920	710	900	1690

**Solution: Computation of trend values**

<b>Year</b>	<b>Production (in lakhs) (Y)</b>	<b>Deviation from 2002 (X)</b>	<b>XY</b>	<b>X<sup>2</sup></b>
2000	830	2000-2002=-2	-1660	4
2001	920	2001-2002=-1	-920	1
2002	710	2002-2002=0	0	0
2003	900	2003-2002=1	900	1
2004	1690	2004-2002=2	3380	4
	<b><math>\Sigma y = 5050</math></b>	<b><math>\Sigma x = 0</math></b>	<b><math>\Sigma xy = 1700</math></b>	<b><math>\Sigma x^2 = 10</math></b>

$$\text{Since } \Sigma x = 0, \quad a = \frac{\Sigma y}{N} = 5050/5 = 1010$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = 1700 / 10 = 170$$

$$Y_c = a + bx = 1010 + 170 x$$

$$Y_c = 1010 + 170 x$$

$$\text{When } X = -2, \quad Y_{2000} = 1010 + 170(-2) = 1010 - 340 = 670$$

$$\text{When } X = -1, \quad Y_{2001} = 1010 + 170(-1) = 1010 - 170 = 840$$

$$\text{When } X = 0, \quad Y_{2002} = 1010 + 170(0) = 1010 - 0 = 1010$$

$$\text{When } x = 1, \quad Y_{2003} = 1010 + 170(1) = 1010 + 170 = 1180$$

$$\text{When } x = 2, \quad Y_{2004} = 1010 + 170(2) = 1010 + 340 = 1350$$





**Illustration 9:** Fit a straight line trend by the method of least squares from the following data and estimate the earnings for the year 2008:

<b>Year :</b>	1999	2000	2001	2002	2003	2004	2005	2006
<b>Earnings (in Lakhs )</b>	38	40	65	72	69	60	87	95

**Solution: Computation of trend values**

<b>Year</b>	<b>Earnings (Rs. in lakhs) (Y)</b>	<b>Deviation from 2002.5 (X)</b>	<b>XY</b>	<b>X<sup>2</sup></b>
1999	38	1999 - 2002.5 = -3.5	-133.0	12.25
2000	40	2000 - 2002.5 = -2.5	-100.0	6.25
2001	65	2001 - 2002.5 = -1.5	-97.5	2.25
2002	72	2002 - 2002.5 = -0.5	-36.0	0.25
2003	69	2003 - 2002.5 = 0.5	34.5	0.25
2004	60	2004 - 2002.5 = 1.5	90.0	2.25
2005	87	2005 - 2002.5 = 2.5	217.5	6.25
2006	95	2006 - 2002.5 = 3.5	332.5	12.25
<b>N = 8</b>	<b>Σy = 526</b>	<b>Σx = 0</b>	<b>Σxy = 308</b>	<b>Σx<sup>2</sup> = 42</b>

$$Y_c = a + bx$$

$$a = \frac{\Sigma y}{N} = 526/8 = 65.75$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = 308 / 42 = 7.33$$

$$\text{That is } Y_c = 65.75 + 7.33 x$$

In the year 2008, the value of x is 5.5 that is (2008 – 2002.5)

∴ The earnings for the year 2008 will be

$$Y_{2008} = 65.75 + 7.33(5.5) = 65.75 + 40.31 = 106.06$$

\*\*\*\*\*



## Theoretical Questions

1. Explain the characteristics of index numbers.
2. What are the uses of index numbers?
3. Discuss the types of index numbers in detail.
4. Describe the various methods of construction of index numbers.
5. Define and explain the advantages of chain base index.
6. How would you convert the Chain Base index to fixed base index?
7. Differentiate chain base and fixed base.
8. Explain the methods of construction of consumer price index.
9. What are uses of cost of living index?
10. Give the limitations of index numbers.
11. Define and explain the types of time series.
12. Describe the components of time series.
13. Explain the uses and causes of time series.
14. What are the steps to calculate 3 yearly moving averages?
15. Explain the steps for calculating of 4yearly moving average.

## Practical Problems

1. Construct an index number for 2016 taking 2015 as base from the following data.

Commodity	Price in 1984 (Rs.)	Price in 1985 (Rs.)
A	90	95
B	40	60
C	90	110
D	30	35

( $P_{01} = 120$ )



2. Compute a price index for the following by average of price relative method.

Commodity	A	B	C	D	E	F
Price in 2005 (Rs.)	10	25	30	20	40	50
Price in 2006 (Rs.)	15	30	40	30	50	55

$$(P_{01} = 131.39)$$

3. Compute Index Number, using Fishers Ideal Formula and show that it satisfies time – reversal test and factor – reversal test.

Commodity	Quantity	Base Year Price	Quantity	Current Year Price
A	50	6	56	10
B	100	2	120	2
C	60	4	60	6
D	30	10	24	12
E	40	8	36	12

$$(P_{01} = 139.8)$$

4. Calculate a) Paasche's index, b) Laspeyre's index and c) Fisher's index numbers for the following data:

Commodity	$P_0$	$q_0$	$P_1$	$q_1$
A	12	20	15	25
B	10	8	16	10
C	15	2	12	1
D	60	1	65	1
E	3	2	10	1

$$(\text{Ans. a) } 130.13; \text{ b) } 129.09; \text{ and c) } 129.60)$$

5. Compute the trend from the following by the method of least squares:

Year :	1976	1977	1978	1979	1980	1981
Population (in Lakhs )	24	25	29	26	22	24

Estimate the likely production for 1984.

$$(\text{Ans. } 23.119)$$



6. The following table gives the price per egg on the first of each month.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1988	18	19	20	21	23	24	26	26	28	25	23	29
1989	18	21	21	21	22	24	25	26	27	24	23	21
1990	19	17	20	22	23	25	25	26	26	24	22	20

Calculate the seasonal variations for each of the months Jan. to Dec.

(Ans. 80.9; 83.9; 89.7; 94.2; 100; 107.5; 111.9; 114.9; 119.3; 107.5; 100.9; 89.8)

\*\*\*\*\*



### **UNIT - III:**

*Differential Calculus - Sum, Product, Quotient Rules, Maximum and Minimum.*

## **DIFFERENTIAL CALCULUS**

### **INTRODUCTION**

Sir Isaac Newton of England and Gottfried Wilhelm Leibnitz of Germany invented calculus in the 17<sup>th</sup> century, independently.

Leibnitz's calculus originated from his attempts to solve some problems in geometry. Newton's calculus originated from his attempts to solve some problems in physics and astronomy.

### **CALCULUS**

#### **Origin**

'Calculus' is a Latin word which means a 'pebble' or a 'small stone'. In ancient times, pebbles were used as for calculations.

#### **Meaning**

Calculus is the mathematics of motion and change. When increasing or decreasing quantities are made the subject of mathematical investigation, it frequently becomes necessary to estimate their rate of growth or decay.

### **BASIC OPERATIONS OF CALCULUS**

The basic operations of calculus are divided into two main parts. They are:

I. Differential Calculus and II. Integral calculus

### **DIFFERENTIAL CALCULUS**

Differential calculus deals with the rate of change of one quantity with respect to another.



## USES OF DIFFERENTIAL CALCULUS

- i) It is used in calculating the rate of change of velocity of a vehicle with respect to time.
- ii) It is used in calculating the rate of change of growth of population with respect to time.
- iii) Differential calculus is also a method by which maxima and minima of functions are obtained.
- iv) Marginal analysis is one of the important applications of calculus in business and economics.
- v) The problems of maximizing profit and minimizing cost under various assumptions can be solved.

## OBJECTIVES OF THE DIFFERENTIAL CALCULUS

The main objective of the differential calculus is to describe an instrument for the measurement of such rates to frame rules for its formation and use.

## RULES OF DIFFERENTIATION

### Constant Rules

The derivative of a constant is zero

If  $y = c$  (constant)

$$\text{Then, } \frac{dy}{dx} = \frac{dc}{dx} = 0$$

### Power Function Rule

The derivative of a power function:  $x^n$  is  $nx^{n-1}$

If  $y = x^n$

$$\text{Then, } \frac{dy}{dx} = nx^{n-1}$$



**For Example,**

$$Y = x^6 \text{ then } \frac{dy}{dx} = \frac{d}{dx}(x^6) = 6x^{6-1} = 6x^5$$

$$Y = x \text{ then } \frac{dy}{dx} = \frac{d}{dx}(x) = 1x^{1-1} = 1x^0 = 1$$

**Sum Rule:**

The derivative of a sum of two functions is the sum of the derivatives of the two functions.

$$\text{If } y = u + v$$

$$\text{Then, } \frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}$$

**Illustration 1:** Differentiate:  $y = 3x^2 + 5x^4$

**Solution:**

$$\frac{dy}{dx} = \frac{d}{dx}(3x^2) + \frac{d}{dx}(5x^4)$$

$$= 6x + 20x^3$$

**Difference Rule:**

The derivative of difference of two functions is the difference of the derivatives of the two functions.

$$\text{If } y = u - v$$

$$\text{Then, } \frac{dy}{dx} = \frac{du}{dx} - \frac{dv}{dx}$$



**Illustration 3:** Differentiate:  $y = 2x^2 - 5x^3 + 5$

**Solution:**

$$\begin{aligned}\frac{dy}{dx} &= \frac{d}{dx}(2x^2) - \frac{d}{dx}(5x^3) + \frac{d}{dx}(5) \\ &= 4x - 15x^2\end{aligned}$$

**Product Rule:**

The derivative of the product of two functions is equal to the first function times the derivative of the second function plus the second function times the derivative of the first function.

$$y = uv$$

$$\frac{dy}{dx} = u \frac{dv}{dx} + v \frac{du}{dx}$$

**Illustration 4:** Differentiate:  $y = (4x + 8)(4x^3)$

**Solution:**

$$u = 4x + 8; v = 4x^3$$

$$\frac{dy}{dx} = \frac{d(uv)}{dx} = (4x + 8) \frac{d}{dx}(4x^3) + (4x^3) \frac{d}{dx}(4x + 8)$$

$$= (4x + 8) 12x^2 + (4x^3)(4)$$

$$= 48x^3 + 96x^2 + 16x^3$$

$$= 64x^3 + 96x^2$$

**Quotient Rule:**

The derivative of the quotient of two functions is given by the following formula.





If  $u$  and  $v$  are differential function and if  $v(x) \neq 0$ ,

$$\text{Then, } y = \frac{u}{v}$$

$$\frac{dy}{dx} = \frac{\frac{vdu}{dx} - \frac{udv}{dx}}{v^2}$$

**Illustration: 5.** Differentiate:  $y = \frac{2x+1}{x}$

**Solution:**

$$\frac{dy}{dx} = \frac{vdu - udv}{v^2}$$

$$y = \frac{u}{v}; u = 2x + 1; v = x$$

$$\frac{dy}{dx} = \frac{d\left(\frac{u}{v}\right)}{dx} = \frac{\frac{xd(2)}{dx} - (2x+1) \cdot \frac{d(1)}{dx}}{x^2}$$

$$= \frac{x \cdot 2 - (2x+1) \cdot 1}{x^2} = \frac{2x - 2x - 1}{x^2} = \frac{-1}{x^2}$$

**Function of a Function Rule:**

If we have a function  $y = f(x)$  where  $x$  is in turn a function of another variable,  $u$ , then the derivative of  $y$  with respect to  $u$  is equal to the derivative of  $y$  with respect to  $x$  times in the derivative of  $x$  with respect to  $u$ .

$$\text{If } y = f(x) \text{ and } x = g(u)$$

$$\text{Then } \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

**Illustration 6:** Differentiate  $y = (x^2 + 8x - 8)^{16}$  with respect to  $x$ .

**Solution:**

$$\text{Let } y = (x^2 + 8x - 8)^{16} \text{ and } u = x^2 + 8x - 8$$



Then  $y = u^{16}$  and  $u = (x^2 + 8x - 8) \frac{dy}{dx} = \frac{d}{dx}(u^{16}) \cdot \frac{d}{dx}(x^2 + 8x - 8) = 16u^{15} (2x + 8) = 16(x^2 + 8x - 8)^{15} \cdot (2x + 8)$

### Function of a Function Rule or Chain Rule of Differentiation

y	$\frac{dy}{dx}$
f(x)	$\frac{dy}{du} \cdot \frac{du}{dx}$

**Illustration 7:** Find  $\frac{d}{dx} e^{ax}$

**Solution:**

$$\text{Let } u = ax \qquad y = e^u$$

$$\frac{du}{dx} = a, \qquad \frac{dy}{du} = e^u = e^{ax}$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$$\therefore \frac{dy}{dx} = ae^{ax}$$

$$\text{So if } y = e^{3x}, \frac{dy}{dx} = 3e^{3x}$$

$$y = e^{-5x}, \frac{dy}{dx} = -5e^{-5x}$$

$$y = e^{-x}, \qquad \frac{dy}{dx} = -e^{-x}$$

$$y = 1/e^{4x} = e^{-4x}, \qquad \frac{dy}{dx} = -4e^{-4x}$$

$$= -4/e^{4x}$$



## Some other Important Rules

Sometimes we have to differentiate logarithmic and exponential functions. Their derivatives are:

### Logarithmic function

$$\text{If } y = \log x \text{ then } \frac{dy}{dx} = \frac{1}{x}$$

### Exponential function

$$\text{If } y = e^x \text{ then } \frac{dy}{dx} = e^x$$

### Higher Order Derivatives:

$\frac{d}{dx}(\frac{dy}{dx})$  is denoted by  $\frac{d^2y}{dx^2}$  read as: “ $d$  – squared  $y$  by  $dx$  squared”. This is called the second order differential coefficient of  $y$  with respect to  $x$ . Various symbols used are  $y^2$ ,  $y''$ ,  $y''(x)$ ,  $f^2(x)$ ,  $D^2y$ .

Similarly  $\frac{d^3y}{dx^3}$  is the third order differential coefficient of  $y$  with respect  $x$  and so on..

**Illustration 8.** Consider the function  $y = x^5$

**Solution:**

$$\frac{dy}{dx} = 5x^4$$

$$\frac{d^2y}{dx^2} = 20x^3$$

$$\frac{d^3y}{dx^3} = 60x^2$$

$$\frac{d^4y}{dx^4} = 120x$$



## Simple applications to Economics

### Elasticity of the function:

The elasticity of a function  $y = f(x)$  at the point  $x$  is the rate of proportional change in  $y$  per unit proportional change in  $x$ .

$$\frac{Ey}{Ex} = \frac{x}{y} \cdot \frac{dy}{dx}$$

**Illustration 9:** Let  $y = 3x - 6$ . Find i) elasticity and ii) the elasticity if  $x = 8$ .

**Solution:**

$$\frac{Ey}{Ex} = \frac{x}{y} \cdot \frac{dy}{dx} = \frac{x}{3x-6} \cdot \frac{d(3x-6)}{dx} = \frac{x}{3x-6} (3) = \frac{3x}{3(x-2)} = \frac{x}{x-2}$$

$$\text{When } x = 8, \frac{Ey}{Ex} = \frac{8}{8-2} = \frac{8}{6} = \frac{4}{3}$$

### Elasticity of Demand

The elasticity of demand denoted by  $\eta$  is defined by  $\eta = -\frac{p}{q} \cdot \frac{dq}{dp}$

**Illustration 10:**  $q = 32 - 4p - p^2$ . Find  $\eta$  when  $p = 3$ .

**Solution:**

$$\eta = -\frac{p}{q} \cdot \frac{dq}{dp}$$

$$= -\frac{3}{32 - 4(3) - (3)^2} \cdot \frac{d(32 - 4p - p^2)}{dp}$$

$$= -\frac{3}{11} (-4 - 2p) = -\frac{3}{11} (-4 - 2(3)) = \frac{30}{11}$$

$$\eta = \frac{30}{11} \text{ or } 2.727.$$



**Marginal Revenue:**

It is defined as  $\frac{dR}{dq}$

$$\frac{dR}{dq} = p(1 - \frac{1}{\eta})$$

**Illustration 11:** If  $p = 3$  and  $\eta = \frac{30}{11}$ . Find  $\frac{dR}{dq}$ .

**Solution:**

$$\frac{dR}{dq} = p(1 - \frac{1}{\eta}) = 3(1 - \frac{11}{30}) = 3(\frac{19}{30}) = \frac{19}{10} = 1.9$$

**Illustration 12:** The total cost, C, of making  $x$  units of a product is,  $C = 0.00003x^3 - 0.045x^2 + 8x + 25000$ . Find the marginal cost at 1000 units out-put.

**Solution:**

$$C = 0.00003x^3 - 0.045x^2 + 8x + 25,000 \dots\dots\dots(1)$$

Marginal Cost at the level  $x$  is  $\frac{dC}{dx}$

Differentiating both sides of with respect to 'x'

$$\frac{dC}{dx} = 0.00009 x^2 - 0.09 x + 8$$

Marginal Cost at  $x = 1000$  is  $\frac{dC}{dx}$  at  $x = 1000$ .

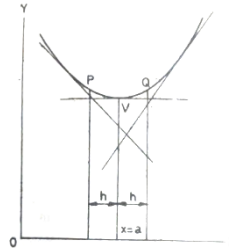
$$\text{When } x = 1000 \Rightarrow \frac{dC}{dx} = 0.00009 (1000)^2 - 0.09(1000) + 8 = 90 - 90 + 8 = 8$$

$\therefore$  Marginal Cost at  $x = 1000$  is 8.



## CRITERIA FOR MINIMA

Consider the curve  $y=f(x)$  given in the below figure. From the figure we see that V is the lowest point of the curve and so the  $y$  - coordinate of V is the minimum value that  $y$  can take. From Figure we see that the ordinates at the points on either side of V are all greater than the ordinate of V.



### Definition

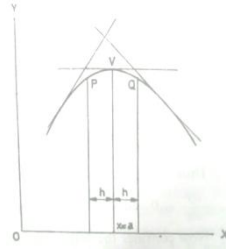
A function  $y = f(x)$  is said to have a **minimum** at  $x = a$  if  $f(a + h) - f(a) \geq 0$ , for sufficiently small  $h$ , positive or negative. So, we find that the function  $y = f(x)$  decreases as  $x$  increases upto V and then increases as  $x$  increases after crossing V.

It is clear that the value of  $\frac{dy}{dx}$  will be negative upto V, becomes zero at V and will be positive after crossing V. Therefore its differential coefficient  $\frac{d^2y}{dx^2}$  should be positive at V.

Thus we have,  $\frac{dy}{dx} = y' = 0$  and  $\frac{d^2y}{dx^2}$  is positive at a point where  $y$  is minimum.

## CRITERIA FOR MAXIMA

Consider the curve  $y = f(x)$  given in the below Fig. From the figure we see that V is the highest point of the curve and so the  $y$  co - ordinate of V is the maximum value that  $y$  can take. The ordinates of the points on either side of V are all less than the ordinate of V.



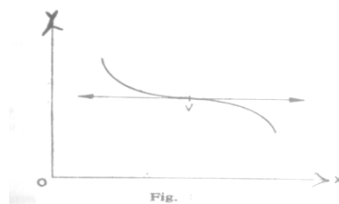
**Definition** A function  $y = f(x)$  is said to have a **maximum** at  $x = a$  if  $f(a + h) - f(a) \leq 0$  for sufficiently small  $h$  positive or negative. So, we find that the function  $y = f(x)$  increases as  $x$  increases upto  $V$ , and then decreases as  $x$  increases after crossing  $V$ . It is clear from figure that the value of  $\frac{dy}{dx}$  will be positive up to  $V$ , becomes zero at  $V$  and will be negative after crossing  $V$ .

Therefore its differential coefficient  $\frac{d^2y}{dx^2}$  should be negative at  $V$ .

Thus we have,  $\frac{dy}{dx} = y' = 0$  and  $\frac{d^2y}{dx^2}$  is negative at a point where  $y$  is maximum.

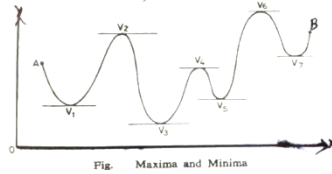
### Points of inflexion

Consider the below figure. Here at the point  $V$  the tangent is parallel to the  $x$ -axis and therefore  $\frac{dy}{dx} = 0$ . But  $V$  gives neither a maximum point nor a minimum point. Actually the tangent crosses the curve here. At such a point  $\frac{d^2y}{dx^2} = 0$  and such points are called **points of inflexion**.



### Extreme Values:

The maximum and minimum values of a function are called the **extreme values** of the function.



### Step by Step Procedure to find Maxima and Minima of a Function of one Variable

1. Find  $\frac{dy}{dx}$  for the given function  $y = f(x)$
2. Find the value or values of  $x$  which make  $\frac{dy}{dx}$  zero. Let these be  $x_1, x_2, x_3, \dots$
3. Find  $\frac{d^2y}{dx^2}$ .
4. Find the sign of  $\frac{d^2y}{dx^2}$  at  $x = x_1, x_2, \dots$ . And hence decide which of these  $x_1, x_2, \dots$  maximize or minimize the function.
5. Find the maximum and minimum values of the function by substituting for  $x$  in  $y = f(x)$  by choosing the suitable value of  $x$  from  $x_1, x_2, x_3, \dots$

Sometimes, by the very nature of the problem, it will be clear whether the value of  $x$  which makes  $\frac{dy}{dx}$  zero maximizes  $y$  or minimizes  $y$ . In such cases there is no need for the second derivative test.

We tabulate the result as:

	Maximum	Minimum
Necessary condition	$\frac{dy}{dx} = 0$	$\frac{dy}{dx} = 0$
Sufficient condition	$\frac{dy}{dx} = 0; \frac{d^2y}{dx^2} < 0,$	$\frac{dy}{dx} = 0; \frac{d^2y}{dx^2} > 0$

**Illustration: 13.** Examine the cost function,  $y = 40 - 4x + x^2$  for maximum or minimum.

**Solution:**

$$y = 40 - 4x + x^2 \quad \dots\dots(1)$$

$$\frac{dy}{dx} = -4 + 2x \quad \dots\dots(2)$$





$$\frac{dy}{dx} = 0 \quad \text{if } -4 + 2x = 0 \Rightarrow x = 2$$

Differentiating (2), with respect to  $x$ ,

$$\frac{d^2y}{dx^2} = 2 > 0 \text{ always}$$

$\therefore x = 2$  makes  $y$ , a minimum and this minimum value of  $y$  is got by putting  $x = 2$  in (1).

$$y = 40 - 4 \times 2 + 2^2 = 40 - 8 + 4 = 36.$$

**Illustration 14:** Examine the function,  $y = 2x^2 - x^3 + 5$  for maximum and minimum.

**Solution:**

$$y = 2x^2 - x^3 + 5 \quad \dots\dots(1)$$

$$\frac{dy}{dx} = 4x - 3x^2 \quad \dots\dots(2)$$

$$\frac{dy}{dx} = 0 \text{ gives, } 4x - 3x^2 = 0 \Rightarrow x = 0 \text{ or } \frac{4}{3}.$$

Differentiating (2), with respect to  $x$ ,

$$\frac{d^2y}{dx^2} = 4 - 6x$$

$$\text{At } x = 0, \frac{d^2y}{dx^2} = 4 - 0 = 4 > 0$$

$\therefore x = 0$  makes  $y$ , a minimum and this minimum value of  $y$  is got by putting  $x = 0$  in (1) is 5.

$$\text{At } x = \frac{4}{3}, \frac{d^2y}{dx^2} = 4 - 6\left(\frac{4}{3}\right) = -4 < 0$$

$\therefore x = \frac{4}{3}$  makes  $y$ , a maximum and this maximum value of  $y$  is got by putting  $x = \frac{4}{3}$  in (1) is  $\frac{167}{27}$ .



**Illustration 15:** Examine for extrema, if any, for the function  $y = x^5 - 5x^3 + 6$ . ( $x$  is real).

**Solution:**

$$y = x^5 - 5x^3 + 6$$

$$\frac{dy}{dx} = 5x^4 + 15x^2 \quad \dots\dots(2)$$

$$\frac{dy}{dx} = 0 \text{ gives, } 5x^4 + 15x^2 = 0 \Rightarrow x = 0.$$

Since  $x$  is real  $x^2 + 3 \neq 0$

$$\frac{d^2y}{dx^2} = 20x^3 + 30x$$

$$\text{When } x = 0, \frac{d^2y}{dx^2} = 0$$

$$\frac{d^2y}{dx^2} = x(20x^2 + 30) \text{ and is less than zero if } x \text{ is less than zero.}$$

$$\text{Also } \frac{d^2y}{dx^2} \text{ is } 0 > \text{ if } x > 0$$

$$\frac{d^2y}{dx^2} \text{ changes sign when it passes through } x = 0$$

Hence  $x = 0$  is a **point of inflexion**.

\*\*\*\*\*

**Practical Problems:**

1. Differentiate the following with respect to 'x':

a.  $-x^2$       b.  $2e^x$       c.  $x^5 \cdot e^x$       d.  $e^x \log x$       e.  $\frac{2x+1}{3x-2}$

2. Find  $\frac{d^2y}{dx^2}$  and  $\frac{d^3y}{dx^3}$  if  $y = 5x^2 - 7x$ . (Ans. 30)



3. The total cost  $C$  of making  $x$  units of product is  $C = 50 + 3x + \sqrt{x}$ . Find the marginal cost at 100 units output. (Ans. 3.05)
4. Find the minimum of the function  $3x^4 - 8x^3 - 90x^2$ . (Ans. -1375 at  $x = -5$ , -351 at  $x = 3$ )
5. Find the maximum of the function  $3x^4 - 10x^3 + 3x^2 + 12x + K$ . (Ans.  $8 + K$ )
6. Find the maximum or minimum for the function  $y = x^2 - 6x$ . [Ans. (3, -9), minimum]
7. Find the maximum and minimum of  $2x^3 - 3x^2 - 36x + 10$ .  
( $x = 3$  min  $y = -71$  &  $x = -2$  max  $y = -54$ )
8. Find the point of inflexion, of the function  $y = 3x^3 - 6x^2$   
(No point of inflexion)
9. Find the point of inflexion, of the function  $y = x^4 - 6x^3$  (Ans. 0, 0)
10. Find the point of inflexion, of the function  $y = ce^{-\frac{x^3}{2}}$  ( $1, \frac{c}{\sqrt{e}}$ ), ( $-1, \frac{c}{\sqrt{e}}$ )



## **UNIT - IV:**

*Integral Calculus - Rules of Integration - Definite integral - Area interpretation.*

### **INTEGRAL CALCULUS**

#### **Integral:**

The technique of finding the functional relationship is called “Integration”. It is the reverse or inverse of differentiation.

#### **Integral calculus:**

Integral calculus deals with finding a function when its rate of change is known.

#### **Constant:**

A constant is a quantity which has only one value throughout a particular investigation. The constant can be classified into two classes. They are: i) Absolute constant and ii) Arbitrary constants

#### **Absolute constant:**

It is constant whose value never changes.

#### **Arbitrary constants:**

It is a constant whose value may change from problem to problem but remains fixed in a single problem.

For example : a, b, c, .... etc.

#### **Definition:**

A function  $F(x)$  is called as anti-derivative or integral of a function  $f(x)$  on an interval  $I$  if,

$$F'(x) = f(x) \text{ for every value of } x \text{ in } I$$



If the differential coefficient of  $F(x)$  with respect to  $x$  is  $f(x)$ , then the integral of  $f(x)$  with respect to  $x$  is  $F(x)$ , in symbols.

$$\text{If } \frac{dF(x)}{dx} = f(x)$$

$$\text{Then, } \int f(x)dx = F(x)$$

$\int$  is used to denote summation. The differential  $dx$  tells us that we have to integrate with respect to  $x$ . The variable  $x$  in  $dx$  is called variable of integration or integrator. The function  $f(x)$  is called integrand. The process of finding the integral is called integration.

### Constant of Integration

Consider the following examples.

$$\text{i.) } \int 5 dx = 5x + c$$

$$\text{ii.) } \int 2x dx = x^2 + c$$

$c$  in the present case is called the constant of integration. Since  $c$  can take any constant value, the integral is called indefinite integral. If  $c$  assumes particular values then the integral becomes definite integral.

### Some Standard Results

$$\text{a.) } \int x^n dx = \frac{x^{n+1}}{n+1} + c \quad (n \neq -1)$$

$$\text{b.) } \int \frac{1}{x} dx = \log x + c$$

$$\text{c.) } \int e^x dx = e^x + c$$

$$\text{d.) } \int e^{-x} dx = -e^{-x} + c$$

$$\text{e.) } \int e^{ax} dx = \frac{e^{ax}}{a} + c$$

$$\text{f.) } \int dx = x + c$$



$$g.) \int a^x dx = \frac{a^x}{\log a} + c$$

**Illustration: 1:** Evaluate the following

$$a.) \int x dx = \int \frac{x^{1+1}}{1+1} + c = \int \frac{x^2}{2} + c$$

$$b.) \int x^2 dx = \int \frac{x^{2+1}}{2+1} + c = \frac{x^3}{3} + c$$

$$c.) \int \frac{1}{x^3} dx = \int x^{-3} dx = \int \frac{x^{-3+1}}{-3+1} + c = \int \frac{x^{-2}}{-2} + c = \frac{-1}{2x^2} + c$$

$$d.) \int \sqrt{x} dx = \int x^{1/2} dx = \frac{x^{\frac{1}{2}+1}}{\frac{1}{2}+1} + c = \frac{x^{\frac{3}{2}}}{\frac{3}{2}} + c = \frac{2x^{3/2}}{3} + c$$

$$e.) \int x^{-4} dx = \frac{x^{-4+1}}{-4+1} + c = \frac{x^{-3}}{-3} + c = \frac{-1}{3x^3} + c$$

$$f.) \int 4x^3 dx = 4 \int x^3 dx = 4 \left[ \frac{x^{3+1}}{3+1} \right] + c = \frac{4x^4}{4} + c = x^4 + c$$

$$g.) \int \frac{1}{x+5} dx = \log(x+5) + c$$

$$h.) \int e^{2x} dx = \frac{e^{2x}}{2} + c$$

**Illustration 2:** Let  $x$  represent the investment in rupees and  $g(x) = \frac{x}{20}$  be the marginal out-put of paddy. Find the total out-put function denoted by  $f(x)$  when no paddy can be produced without any investment. What will be the out-put function if the land is in such a bad condition that only after spending Rs. 1000 for clearing it, one can start farming?

$$g(x) = \frac{x}{20}$$

$$f(x) = \int \frac{x}{20} dx = \frac{x^2}{40} + c$$

$$f(x) = \frac{x^2}{40} + c \quad \dots (1)$$

(i) Given that when  $x = 0$ ,  $f(x) = 0$ . Using this in (1) we have

$$0 = 0 + c \quad c = 0$$

Hence 
$$f(x) = \frac{x^2}{40}$$



(ii) Given that when  $x = 1000, f(x) = 0$

Using this in (1) we have

$$0 = \frac{1000^2}{40} + c$$

$$c = -\frac{1000^2}{40} = -25000$$

$$f(x) = \frac{x^2}{40} - 25,000$$

Integration is inherently more difficult than differentiation. As we have done in differentiation we will have lists of standard forms and certain rules of integration using which we shall be in a position to integrate functions which normally arise in business situations.

### Standard Forms List I

$$(1) \int x^n dx = \frac{x^{n+1}}{n+1} + c$$

Where  $n$  is any constant other than  $-1$

$$\begin{aligned} \text{For, } \frac{d}{dx} \left( \frac{x^{n+1}}{n+1} + c \right) &= \frac{1}{n+1} \frac{d}{dx} x^{n+1} \\ &= \frac{1}{n+1} (n+1) x^{n+1-1} = x^n \end{aligned}$$

(2) If  $n = -1,$

$$\begin{aligned} \int x^n dx &= \int x^{-1} dx = \int \frac{1}{x} dx \\ &= \log x + c \end{aligned}$$

$$\text{since } \frac{d}{dx} (\log x + c) = \frac{1}{x}$$



$$(3) \quad \int e^{kx} dx = \frac{e^{kx}}{k} + c \text{ where } k \text{ is any constant}$$

$$\begin{aligned} \text{For, } \frac{d}{dx} \left( \frac{e^{kx}}{k} + c \right) &= \frac{1}{k} \frac{d}{dx} e^{kx} \\ &= \frac{1}{k} \times k e^{kx} = e^x \end{aligned}$$

$$\therefore \int e^{kx} dx = \frac{e^{kx}}{k} + c$$

$$(4) \quad \int a^{kx} dx = \frac{a^{kx}}{k \log a} + c$$

Where, 'k' and 'a' are constants

$$\begin{aligned} \text{For, } \frac{d}{dx} \left( \frac{a^{kx}}{k \log a} \right) &= \frac{1}{k \log a} \frac{d}{dx} a^{kx} \\ &= \frac{1}{k \log a} a^{kx} k \log a = a^{kx} \end{aligned}$$

$$\therefore \int a^{kx} dx = \frac{a^{kx}}{k \log a} + c$$

Particular useful cases of these standard forms:

$$\begin{aligned} \int dx &= \int 1 dx = \int x^0 dx \\ &= \frac{x^{0+1}}{0+1} + c = x + c \end{aligned}$$

This is a particular case of the standard form (1) when,  $n = 0$ . Even direct method of evaluation is very simple once we observe that

$$(a) \int dx = \int 1. dx \text{ and } \frac{d}{dx}(x) = 1$$

$$(b) \int e^x dx = e^x + c$$





particular case of the standard form (3) when  $k = 1$

$$(c) \int a^x dx = \frac{a^x}{\log a} + c \text{ is also a particular case of (4)}$$

when  $k = 1$

### Some Rules of Integration

$$(1) \int \{f_1(x) + f_2(x)\} dx = \int f_1(x) dx + \int f_2(x) dx$$

This can be extended to a sum of any finite number of functions. This rule states that integral of a sum of any finite number of functions is equal to the sum of the integrals of the individual functions.

Similarly we have

$$\begin{aligned} \int \{f_1(x) - f_2(x)\} dx \\ = \int f_1(x) dx - \int f_2(x) dx \end{aligned}$$

and

$$\begin{aligned} \int \{f_1(x) + f_2(x) + f_3(x)\} dx \\ = \int f_1(x) dx + \int f_2(x) dx + \int f_3(x) dx \end{aligned}$$

and so on.

$$(2) \int kf(x) dx = k \int f(x) dx \text{ Where } k \text{ is any constant.}$$

(Proofs can be supplied by using the corresponding rules we have in differentiation)

**Illustration 3:** Evaluate  $\int x^{1/2} dx$

**Solution:**

$$\text{We know that } \int x^n dx = \frac{x^{n+1}}{n+1} + c \text{ if } n \neq -1$$



$$\begin{aligned}\int x^{1/2} dx &= \frac{x^{1/2+1}}{\frac{1}{2}+1} + c \\ &= \frac{x^{3/2}}{\frac{3}{2}} + c = \frac{2}{3} x^{3/2} + c\end{aligned}$$

**Illustration 4:** Evaluate  $\int (x^3 + 4x^2 - 5x - 6) dx$

**Solution:**

$$\begin{aligned}\int (x^3 + 4x^2 - 5x - 6) dx &= \int x^3 dx + \int 4x^2 dx - \int 5x dx - \int 6 dx \\ &= \frac{x^{3+1}}{3+1} + 4 \int x^2 dx - 5 \int x dx - 6 \int dx + c \\ &= \frac{x^4}{4} + 4 \frac{x^{2+1}}{2+1} - 5 \frac{x^{1+1}}{1+1} - 6x + c \\ &= \frac{x^4}{4} + \frac{4}{3} x^3 - 5 \frac{x^2}{2} - 6x + c\end{aligned}$$

**Illustration 5:** Evaluate  $\int \left( x^{3/2} - \frac{1}{x^2} + \frac{5}{x} - 7 \right) dx$

**Solution:**

$$\begin{aligned}\int \left( x^{3/2} - \frac{1}{x^2} + \frac{5}{x} - 7 \right) dx &= \int x^{3/2} dx - \int x^{-2} dx + \int \frac{5}{x} dx - \int 7 dx \\ &= \frac{x^{3/2+1}}{\frac{3}{2}+1} - \int x^2 dx + 5 \int \frac{1}{x} dx - 7 \int dx + c \\ &= \frac{x^{5/2}}{\frac{5}{2}} - \frac{x^{-2+1}}{-2+1} + 5 \log x - 7x + c \\ &= \frac{2}{5} x^{5/2} + x^{-1} + 5 \log x - 7x + c\end{aligned}$$



**Illustration 6:** Evaluate  $\int (3^x + e^x + x^{-7/2}) dx$

**Solution:**

$$\begin{aligned}\int (3^x + e^x + x^{-7/2}) dx &= \int 3^x dx + \int e^x dx + \int x^{-7/2} dx \\ &= \frac{3^x}{\log 3} + e^x + \frac{x^{-7/2+1}}{-\frac{7}{2}+1} + c = \frac{3^x}{\log 3} + e^x - \frac{2}{5} x^{-5/2} + c\end{aligned}$$

**Illustration 7:** Evaluate  $\int \left( \frac{4x^7 + 3x^3 - 5x^2}{x^4} \right) dx$

**Solution:**

$$\begin{aligned}\int \left( \frac{4x^7 + 3x^3 - 5x^2}{x^4} \right) dx &= \int \frac{4x^7}{x^4} dx + \int \frac{3x^3}{x^4} dx - \int \frac{5x^2}{x^4} dx \\ &= \int 4x^3 dx + \int \frac{3}{x} dx - \int \frac{5}{x^2} dx \\ &= 4 \frac{x^4}{4} + 3 \log x - 5 \int x^{-2} dx + c \\ &= x^4 + 3 \log x - 5 \frac{x^{-2+1}}{-2+1} + c \\ &= x^4 + 3 \log x + 5x^{-1} + c\end{aligned}$$

**Illustration 8:** Evaluate  $\int \left( x + \frac{1}{x} \right)^2 dx$

**Solution:**

$$\begin{aligned}\int \left( x + \frac{1}{x} \right)^2 dx &= \int \left( x^2 + 2 + \frac{1}{x^2} \right) dx \\ &= \int x^2 dx + \int 2 dx + \int x^{-2} dx \\ &= \frac{x^{2+1}}{2+1} + 2x + \frac{x^{-2+1}}{-2+1} + c = \frac{x^3}{3} + 2x - x^{-1} + c\end{aligned}$$



**Illustration 9:** Evaluate  $\int (e^{2x} - 5x^{-1}) dx$

**Solution:**

$$\int (e^{2x} - 5x^{-1}) dx = \int e^{2x} dx - \int 5x^{-1} dx = \frac{e^{2x}}{2} - 5 \log x + c$$

### Definite Integral

Consider the integral  $\int 3x^2 dx$

We know that  $\int 3x^2 dx = x^3 + c$

When  $x = 4$ , the value of the integral is  $4^3 + c = 64 + c$

When  $x = 2$ , the value of the integral is  $2^3 + c = 8 + c$

(The value of the integral  $\int 3x^2 dx$  when  $x = 4$ ) - (The value of the integral  $\int 3x^2 dx$  when  $x = 2$ )

$$= 64 + c - (8 + c) = 64 - 8 = 56$$

We say that we have integrated the function  $3x^2$  with respect to  $x$  from  $x = 2$  to  $x = 4$  and we denote the whole thing as

$$\int_2^4 3x^2 dx = 56$$

$$\int_2^4 3x^2 dx = \text{value of } \int 3x^2 dx \text{ at } x = 4$$

$$= \text{value of } \int 3x^2 dx \text{ at } x = 2$$



2 and 4 are called limits of integration. 2 is called the lower limit of the integral and 4 is called the upper limit of the integral and  $\int_2^4 3x^2 dx$  is called the “**Definite Integral**” of  $3x^2$  integrated from  $x = 2$  to  $x = 4$ . Notice that in the value of the definite integral, C, the arbitrary constant of integration disappears. That is, the integral has no arbitrariness. It has got a definite value and hence the name definite integral.

### Definition

If  $\int f(x) dx = g(x) + c$ , then  $\int_a^b f(x) dx = g(b) - g(a)$  is called the **Definite Integral** of  $f(x)$  with respect to  $x$  integrated from  $x = a$  to  $x = b$ , ‘a’ and ‘b’ are called the lower and upper limits of integration respectively.

We denote  $g(b) - g(a)$  by  $[g(x)]_a^b$

$$\therefore \int_a^b f(x) dx = [g(x)]_a^b$$

It is easy to verify the following facts

$$(i) \int_a^b f(x) dx = - \int_b^a f(x) dx$$

$$(ii) \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx, a < c < b$$

$$= \int_a^{c_1} f(x) dx + \int_{c_1}^{c_2} f(x) dx + \dots + \int_{c_{n-1}}^b f(x) dx$$

Where  $a < c_1 < c_2 \dots \dots < c_{n-1} < b$

Proofs follow at once from the definition of the definite integral.

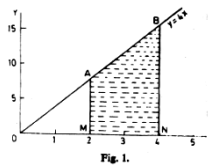


## Meaning of a Definite Integral

$$\text{Consider } \int_2^4 4x dx$$

$$\int_2^4 4x dx = [2x^2]_2^4 = 32 - 8 = 24$$

Has the number 24 any significance? Consider the graph of the integrand *i.e.*,  $4x$ . Let  $y = 4x$ . We know that the graph of  $y = 4x$  is a straight line. Consider the portion of the graph lying between  $x = 2$  (the lower limit of the integral) and  $x = 4$  (the upper limit of the integral).



Consider the area included by the part of this line bounded by the ordinates at  $x = 2$  (the lower limit of the definite integral) and at  $x = 4$  (the upper limit of the definite integral) and the part of the  $x$  - axis lying between these ordinates as shown in Fig. 1.

The area bounded by the stated lines is the trapezium AMNBA. We know that the area of AMNB

$$\begin{aligned} &= \frac{1}{2}(AM + BN)MN \\ &= \frac{1}{2}(8 + 16) \times 2 = 24 \end{aligned}$$

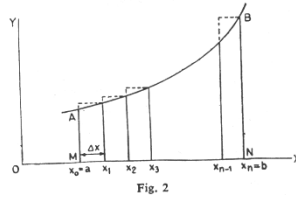
Thus we notice that  $\int_2^4 4x dx = 24$ , is the area bounded by  $y = 4x$ , the ordinates at  $x = 2$ ,  $x = 4$ , and the part of the  $x$  - axis lying between them. This is not a coincidence. This is in fact generally true. The definite integral  $\int_a^b f(x) dx$  is the area bounded by the curve  $y = f(x)$ , the ordinates at  $x = a$ ,  $x = b$  and the  $x$  - axis cut off by these ordinates.



## Area under a Curve

Let us find the area below the curve  $y = f(x)$ , bounded by the  $x$  – axis and the lines  $x = a, x = b$ .

Divide the length  $MN(= b - a)$  into  $n$  equal parts each of length  $\Delta x$ .



Then 
$$\Delta x = \frac{b-a}{n}$$

The points of subdivision are

$$x_0 = a$$

$$x_1 = a + \Delta x$$

$$x_2 = a + 2\Delta x$$

$$x_3 = a + 3\Delta x$$

$$x_n = a + n\Delta x$$

Let the area of AMNB be  $A$ .

The area AMNB to be found out can be approximated by the sum of the areas of  $n$  rectangles as follows.

$$A = f(x_1)\Delta x + f(x_2)\Delta x + f(x_3)\Delta x + \dots + f(x_n).\Delta x \text{ nearly}$$

*i. e.,*

$$A = \sum_{i=1}^n f(x_i) \Delta x \text{ nearly}$$

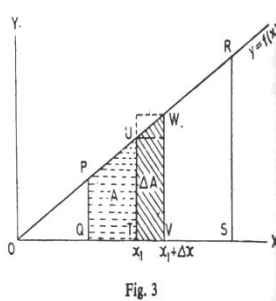


We will have better approximation if we have finer subdivisions i.e., more rectangles with still smaller width and thus,

$$A = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x$$

### Area by Integration: The Definite Integral

The limit of the sum of areas can be evaluated by integration. Let us find the area bounded by the curve  $y = f(x)$ , the lines  $x = a, x = b$  and the  $x$  – axis as shown in the figure.



(We take the curve to be a straight line for simplicity sake) We can consider the area  $PQRS$  as that traced out by moving  $PQ$  parallel to itself in such a way that  $P$  moves along the curve  $y = f(x)$  and  $Q$  moves along the  $x$  – axis starting from the position  $PQ$  marked in the figure to the position  $RS$ . Let  $TU$  be the position of  $PQ$  at some general stage. Let  $x_1$  be the  $x$  –coordinate of  $T$  and therefore  $TU = f(x_1)$ . Let the area  $PQTU$  be  $A$  and let  $VW$  be another position of  $PQ$  such that it is very close to  $TU$ . Let  $TV = \Delta x$  and the area  $UTVW$  be  $\Delta A$ . It is clear that the area is a function of  $x_1$  for if we vary  $x_1$  the area  $A$  will vary in general and  $\Delta A$  is nothing but the increase in area  $A$  as  $x_1$  increases by  $\Delta x$  to  $x_1 + \Delta x$ . Apparently  $\Delta A$  is larger than the rectangle which is determined by the base  $\Delta x$  times the height  $f(x_1)$  and smaller than the rectangle determined by  $\Delta x$  times  $f(x_1 + \Delta x)$ .

i.e., 
$$f(x_1)\Delta x < \Delta A < f(x_1 + \Delta x)\Delta x$$

Dividing this inequality throughout by  $\Delta x$  and noting that  $\Delta x$  is positive we have

$$f(x_1) < \frac{\Delta A}{\Delta x} < f(x_1 + \Delta x) \quad \dots(1)$$





Since we consider continuous function only

$$f(x_1 + \Delta x) \rightarrow f(x_1) \quad \text{as } \Delta x \rightarrow 0$$

and also

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta A}{\Delta x} = \frac{dA}{dx}$$

Taking limit as  $\Delta x \rightarrow 0$  of (1)

$$f(x_1) < \lim_{\Delta x \rightarrow 0} \frac{\Delta A}{\Delta x} < f(x_1)$$

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta A}{\Delta x} = f(x_1)$$

i.e.,

$$\frac{dA}{dx} = f(x_1)$$

Since  $x_1$  is a general value of  $x$  which can be anywhere between  $a$  and  $b$  we can write

$$\frac{dA}{dx} = f(x)$$

$$dA = f(x)dx$$

$$A = \int f(x)dx + c$$

Let

$$\int f(x)dx = g(x)$$

Then

$$A = g(x) + c \quad \dots(2)$$

From the diagram we see that  $A$  represents the area under the curve  $y = f(x)$  and between the lines  $x = a$  and  $x = x_1$ .

When  $x_1 = a$  (its initial value),  $A = 0$  trivially.

(2) becomes

$$0 = g(a) + c$$

$$c = -g(a) \quad \dots(3)$$



Using (3) in (2) we have

$$A = g(x) - g(a) = \int_a^x f(x)dx \quad \dots (4)$$

When  $x = b$  we get the whole area  $PQRS$ . From (2) we have

$$A = g(b) + c = g(b) - g(a) \quad \text{using (3).}$$

$$\therefore \text{Reqd. area } PQRS = A = \int_a^b f(x)dx.$$

Combining the results in this and the previous articles we have the following important result which illustrates the fact that integration is ‘after all’ an ‘infinite summation’

$$A = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i) \Delta x = \int_a^b f(x)dx$$

**Illustration 10:** Evaluate  $\int_0^2 e^{2x} dx$

**Solution:**

$$\begin{aligned} \int_0^2 e^{2x} dx &= \left( \frac{e^{2x}}{2} \right)_0^2 = \frac{e^4}{2} - \frac{e^0}{2} \\ &= \frac{e^4}{2} - \frac{1}{2} = \frac{1}{2}(e^4 - 1) \end{aligned}$$



**Illustration 11:** Evaluate  $\int_1^5 \left(x - \frac{2}{x}\right) dx$

**Solution:**

$$\begin{aligned}\int_1^5 \left(x - \frac{2}{x}\right) dx &= \left(\frac{x^2}{2} - 2 \log_e x\right)_1^5 \\ &= \left(\frac{25}{2} - 2 \log_e 5\right) - \left(\frac{1}{2} - 2 \log 1\right) \\ &= \frac{25}{2} - 2 \log_e 5 - \frac{1}{2} \\ &= \frac{24}{2} - 2 \log_e 5 \\ &= 12 - 2 \log_e 5\end{aligned}$$

**Practical Problems:**

1. Evaluate the following integrals

- |  |  |
|--|--|
| (i) $\int \left(x + \frac{1}{x}\right) dx$ | (ii) $\int (x^{3/2} + 2x^2 - 4) dx$                            |
| (iii) $\int \frac{2x-3}{6x^2} dx$          | (iv) $\int \left(2x + \frac{3}{x}\right)^2 dx$                 |
| (v) $\int \frac{2x^3-5x^2+7}{x} dx$        | (vi) $\int \left(4e^x + \frac{5}{x} - \frac{1}{x^2}\right) dx$ |

2. Find the approximate values of the area bounded by the line  $y = 4x$  and the ordinates at  $x = 2, x = 4$  and the  $x$ -axis considering the sum of the rectangles by taking

- (a)  $\Delta x = 1$ , ordinates at  $x = 3, 4$
- (b)  $\Delta x = 0.5$ , ordinates at  $x = 2.5, 3.0, 3.5, 4$
- (c)  $\Delta x = 0.25$ , ordinates at  $x = 2.25, 2.5, 2.75, 3.00, 3.25, 3.5, 3.75, 4$ .

Compare these approximations with the actual area got by integration.



3. Evaluate the following:

i.  $\int_0^2 e^{-0.5x} dx$     ii.  $\int_2^4 6^x dx$     iii.  $\int_0^{25} \frac{1}{\sqrt{x}} dx$     iv.  $\int_0^{1/4} (x^{1/2} - 1) dx$

4. Find the area included between the curve  $y = 9 - x^2$ , the  $x$  - axis and the lines  $x=0$  and  $x = 3$ .

5. Find the area under  $y = 5^x$  over the interval from  $x = 0$  to  $x = 2$ .

6. Find the area under  $xy = 10$  over the interval from  $x = 2$  to  $x = 3$ .



## UNIT - V:

*Matrices - Inversion - Solving system of equations.*

### MATRICES

#### Definition

A matrix is a rectangular array consists of elements arranged in rows and columns enclosed by a pair of brackets.

#### Order of Matrix

A rectangular arrangement of elements in 'm' rows and n columns is called a matrix of order m x n.

#### General format of a matrix

The general format of a matrix having 'm' rows and 'n' columns is

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & & \cdot & & \\ & & \cdot & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \quad m \times n$$

#### Notations:

Every matrix is denoting by capital letters and its elements by corresponding small letters.

#### Types of Matrices:

The following are the types of matrix

##### 1. Row matrix:

A matrix that consists of only one row is called a row matrix.



**For example:**  $B = [1\ 2\ 3\ 4]$

## 2. Column matrix:

A matrix consists of only one column is called a column matrix.

**For example:**  $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$

## 3. Square matrix:

A matrix, in which the number of rows is equal to the number of columns, is called square matrix.

**For example:**  $X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

## 4. Zero matrix or Null matrix:

A matrix in which all the elements are zero is called a null matrix.

**For example:**  $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

## 5. Diagonal Matrix:

A square matrix in which all the elements are zero, except the leading diagonal elements.

**For example:**  $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$

## 6. Scalar Matrix:

A diagonal matrix in which all the diagonal elements are equal is called a scalar matrix.



**For example:**  $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

### 7. Identity / Unit Matrix:

A square matrix is said to be unit matrix if all the elements are zero except diagonal elements and diagonal elements are equal to 1. It is denoted by  $I$ .

**For example:**  $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

### 8. Triangular Matrix:

A square matrix in which all the elements below the leading diagonal elements are zero, it is called Upper Triangular Matrix.

**For example:**  $A = \begin{bmatrix} 13 & 5 \\ 0 & 5 & 4 \\ 0 & 0 & 9 \end{bmatrix}$

### 9. Lower Triangular Matrix:

A square matrix in which all the elements above the leading diagonal elements are zero is called a lower triangular matrix.

**For example:**  $A = \begin{bmatrix} -1 & 0 & 0 \\ 8 & 1 & 4 \\ 3 & -4 & 2 \end{bmatrix}$

### 10. Singular and Nonsingular Matrices:

A square matrix  $A$  is said to be singular if its determinant is zero.  $|A| = 0$

A square matrix  $A$  is said to be non-singular if its determinant is not equal to zero.  $|A| \neq 0$



### 11. Transpose Matrix:

The matrix obtained by interchanging rows and column of the matrix A is called the transpose of Matrix A. It is denoted by  $A^T$  (read as A transpose).

$$\text{If } A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\text{Then } A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

### 12. Symmetric Matrix:

A square matrix is called a symmetric matrix. If  $a_{ij} = a_{ji}$  for all i and j.

$$\text{For example: } A = \begin{bmatrix} 1 & -5 & 3 \\ -5 & -2 & 4 \\ 3 & 4 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} x & a & b \\ a & y & c \\ b & c & z \end{bmatrix}$$

### 13. Skew-symmetric Matrix:

A square matrix is said to be skew symmetric if  $a_{ij} = -a_{ji}$  for all i and j and  $a_{ii} = 0$ .

$$\text{For example: } X = \begin{bmatrix} 0 & -3 & 2 \\ 3 & 0 & 4 \\ -3 & 4 & 2 \end{bmatrix}$$

### 14. Equal Matrix:

Two matrices are said to be equal if and only if they are comparable and each element of one is equal to the corresponding element of the other.

$$\text{For example: If } A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & -5 & 6 \end{bmatrix}$$

### 15. Orthogonal Matrix:

A square matrix A is said to be orthogonal if  $AA^T = A^T A = I$ .





## 16. Sub – Matrix:

Let A be a given matrix. By deleting a few rows and columns we get a new matrix called the sub matrix of A and A is also a sub matrix of itself.

**For example:** If  $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$

The sub – matrices

are  $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, [1 \ 2 \ 3], [4 \ 5 \ 6], \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}, [1], [2], [3], [4], [5], [6], [1 \ 2], [1 \ 3], [2 \ 3], [4 \ 6], [5 \ 6], [1], [2], [3], [4], [5], [6].$

## 17. Scalar Multiplication:

Multiplication of every element in a matrix by a constant is called scalar multiplication.

**For example:**  $A = \begin{bmatrix} 2 & 5 \\ 6 & 8 \end{bmatrix}$  Then  $2A = 2 \begin{bmatrix} 2 & 5 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 12 & 16 \end{bmatrix}$

## Properties of a Matrix

- i) If A and B are two matrices of order 'm' and 'n' then  $A + B = B + A$ .
- ii)  $(A + B) + C = A + (B + C)$
- iii)  $K(A + B) = KA + KB$ .
- iv) Existence of additive identity, that is,  $A + 0 = 0 + A = A$
- v) Existence of additive inverse, that is,  $A + (-A) = (-A) + A = 0$
- vi) Cancellation law  $A + C = B + C \Rightarrow A = B$
- vii) Multiplication is distributive over addition, that is  $A(B + C) = AB + AC$ .
- viii) Matrix multiplication is associative that is,  $(AB)C = A(BC)$



ix) If matrices A and O are of the order  $m \times n$ , then  $AO = O = OA$ .

x) Multiplication of a matrix by unit matrix is that matrix itself  $AI = IA = A$ .

### Product of Matrices

Two matrices A and B can be multiplied if and only if the number of columns in A is equal to the number of rows in B. The product matrix is denoted by AB. In these case we say matrices A and B are compatible for multiplication and the product matrix AB has the same number of rows as A and the same number of columns as B. Thus if A is a  $m \times n$  matrix and B is a  $n \times p$  matrix then AB is a  $m \times p$  matrix.

### Multiplication of Matrices

Two matrices A and B can be multiplied if and only if the number of column in A is equal to the number of rows in B. The resultant matrix AB will have the same number of rows as A and the same number of columns as B.

$$A = [ \quad ]_{m \times n} \quad B = [ \quad ]_{n \times p}$$

$$\therefore AB = [ \quad ]_{m \times p}$$

**Illustration 1:** If  $A = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 5 \end{bmatrix}$  and  $B = \begin{bmatrix} 4 & 1 \\ 2 & 0 \\ 5 & 3 \end{bmatrix}$  find AB

#### Solution:

Here, columns of A = 3 and rows B is also 3. Hence matrix multiplication is possible.

$$\begin{aligned} AB &= \begin{bmatrix} 3 & 4 & 2 \\ 1 & 2 & 5 \end{bmatrix}_{2 \times 3} \begin{bmatrix} 4 & 1 \\ 2 & 0 \\ 5 & 3 \end{bmatrix}_{3 \times 2} \\ &= \begin{bmatrix} (3 \times 4) + (4 \times 2) + (2 \times 5) & (3 \times 1) + (4 \times 0) + (2 \times 3) \\ (1 \times 4) + (2 \times 2) + (5 \times 5) & (1 \times 1) + (2 \times 0) + (5 \times 3) \end{bmatrix} \end{aligned}$$



$$= \begin{bmatrix} (12 + 8 + 10) & (3 + 0 + 6) \\ (4 + 4 + 25) & (1 + 0 + 15) \end{bmatrix} = \begin{bmatrix} 30 & 9 \\ 33 & 16 \end{bmatrix}_{2 \times 2}$$

**Illustration 2:** If  $A = [5 \ 2 \ 8]$  and  $B = \begin{bmatrix} 4 \\ 2 \\ 7 \end{bmatrix}$  find  $AB$ .

**Solution:**

Here, columns of  $A = 3$  and rows of  $B = 3$ . Hence matrix multiplication is possible

$$\begin{aligned} AB &= [5 \ 2 \ 8]_{1 \times 3} \begin{bmatrix} 4 \\ 2 \\ 7 \end{bmatrix}_{3 \times 1} \\ &= [(5 \times 4) + (2 \times 2) + (8 \times 7)] = [(20 + 4 + 56)] = [80] \end{aligned}$$

**Illustration 3:** If  $A = \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}$   $B = \begin{bmatrix} 5 & -2 \\ -1 & 6 \end{bmatrix}$  then verify whether  $AB = BA$ .

**Solution:**

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ -1 & 6 \end{bmatrix} = \begin{bmatrix} (2 \times 5) + (3 \times (-1)) & (2 \times (-2)) + (3 \times 6) \\ ((-1) \times 5) + (4 \times (-1)) & ((-1) \times (-2)) + (4 \times 6) \end{bmatrix} \\ &= \begin{bmatrix} 10 - 3 & -4 + 18 \\ -5 - 4 & 2 + 24 \end{bmatrix} = \begin{bmatrix} 7 & 14 \\ -9 & 26 \end{bmatrix} \quad \dots (1) \end{aligned}$$

$$\begin{aligned} BA &= \begin{bmatrix} 5 & -2 \\ -1 & 6 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} (5 \times 2) + ((-2) \times (-1)) & (5 \times 3) + ((-2) \times 4) \\ ((-1) \times 2) + (6 \times (-1)) & ((-1) \times 3) + (6 \times 4) \end{bmatrix} \\ &= \begin{bmatrix} 10 + 2 & 15 - 8 \\ -2 - 6 & -3 + 24 \end{bmatrix} = \begin{bmatrix} 12 & 7 \\ -8 & 21 \end{bmatrix} \quad \dots (2) \end{aligned}$$

From (1) and (2) we get  $AB \neq BA$

Thus it verified.



**Illustration 4 :** If  $A = \begin{bmatrix} 17 & 5 \\ 14 & 6 \end{bmatrix}$  and  $B = \begin{bmatrix} 1 & 0 \\ 2 & -2 \end{bmatrix}$  then find the value of  $x$  if  $5x + 2B = A$ .

**Solution:**

$$5x + 2B = A$$

$$5x = A - 2B$$

$$2B = 2 \begin{bmatrix} 1 & 0 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 4 & -4 \end{bmatrix}$$

$$5x = A - 2B = \begin{bmatrix} 17 & 5 \\ 14 & 6 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 4 & -4 \end{bmatrix} = \begin{bmatrix} 17-2 & 5-0 \\ 14-4 & 6-(-4) \end{bmatrix} = \begin{bmatrix} 15 & 5 \\ 10 & 10 \end{bmatrix}$$

$$\therefore x = \frac{1}{5} \begin{bmatrix} 15 & 5 \\ 10 & 10 \end{bmatrix} = \begin{bmatrix} \frac{15}{5} & \frac{5}{5} \\ \frac{10}{5} & \frac{10}{5} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

$$\therefore \text{The value of } x = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

**Illustration 5 :** Show that  $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$  satisfies the equation  $A^2 - 4A - 5I = 0$ .

**Solution:**

$$A^2 = AA = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 8 & 8 \\ 8 & 9 & 8 \\ 8 & 8 & 9 \end{bmatrix}$$

$$4A = 4 \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 8 & 8 \\ 8 & 4 & 8 \\ 8 & 8 & 4 \end{bmatrix}$$

$$5I = 5 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$



$$A^2 - 4A = \begin{bmatrix} 9 & 8 & 8 \\ 8 & 9 & 8 \\ 8 & 8 & 9 \end{bmatrix} - \begin{bmatrix} 4 & 8 & 8 \\ 8 & 4 & 8 \\ 8 & 8 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$A^2 - 4A - 5I = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$$

$\therefore A^2 - 4A - 5I = 0$ . Thus it showed that A satisfied the equation.

## Inverse of a Matrix

### Finding the Inverse by Adjoint Matrix Method

$$A^{-1} = \frac{1}{|A|} \text{adj}A$$

An Important Fact.

$A^{-1}$  exists if and only if  $|A| \neq 0$ .

**Illustration 6:** Find the inverse of  $\begin{bmatrix} 2 & 2 \\ 3 & 5 \end{bmatrix}$

**Solution:**

$$A^{-1} = \frac{1}{|A|} \text{adj}A$$

$$|A| = \begin{vmatrix} 2 & 2 \\ 3 & 5 \end{vmatrix} = 2 \times 5 - 2 \times 3 = 10 - 6 = 4 \neq 0$$

$\therefore A^{-1}$  exists.

$\text{Adj} A = (\text{Cofactor of } A)'$

Cofactor of A:

$$C_{11} = (-1)^{1+1} 5 = 5$$



$$C_{12} = (-1)^{1+2} 3 = -3$$

$$C_{21} = (-1)^{2+1} 2 = -2$$

$$C_{22} = (-1)^{2+2} 2 = 2$$

$$\therefore \text{Cofactor of } A \begin{bmatrix} 5 & -3 \\ -2 & 2 \end{bmatrix}$$

$$\text{Adj } A = (\text{Cofactor of } A)' = \begin{bmatrix} 5 & -2 \\ -3 & 2 \end{bmatrix}$$

$$A^{-1} = \frac{1}{|A|} \text{adj}A$$

$$A^{-1} = \frac{1}{4} \begin{bmatrix} 5 & -2 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & \frac{-2}{4} \\ \frac{-3}{4} & \frac{2}{4} \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & \frac{-1}{2} \\ \frac{-3}{4} & \frac{1}{2} \end{bmatrix}$$

**Illustration 7:** Find the inverse of  $\begin{bmatrix} 4 & 0 & 2 \\ 2 & 10 & 2 \\ 3 & 9 & 1 \end{bmatrix}$

**Solution:**

$$A^{-1} = \frac{1}{|A|} \text{adj}A$$

$$|A| = \begin{vmatrix} 4 & 0 & 2 \\ 2 & 10 & 2 \\ 3 & 9 & 1 \end{vmatrix} = 4 \begin{vmatrix} 10 & 2 \\ 9 & 1 \end{vmatrix} - 0 \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} + 2 \begin{vmatrix} 2 & 10 \\ 3 & 9 \end{vmatrix}$$

$$= 4[(10 \times 1) - (9 \times 2)] - 0 + 2[(2 \times 9) - (3 \times 10)]$$

$$= 4(10 - 18) + 2(18 - 30) = 4(-8) + 2(-12) = -32 - 24 = -56 \neq 0$$

$\therefore A^{-1}$  exists.

$$\text{Adj } A = (\text{Cofactor of } A)'$$



Cofactor of A:

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 10 & 2 \\ 9 & 1 \end{vmatrix} = 10 - 18 = -8$$

$$C_{12} = (-1)^{1+2} \begin{vmatrix} 2 & 2 \\ 3 & 1 \end{vmatrix} = -(2 - 6) = +4$$

$$C_{13} = (-1)^{1+3} \begin{vmatrix} 2 & 10 \\ 3 & 9 \end{vmatrix} = 18 - 30 = -12$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 0 & 2 \\ 9 & 1 \end{vmatrix} = -(0 - 18) = +18$$

$$C_{22} = (-1)^{2+2} \begin{vmatrix} 4 & 2 \\ 3 & 1 \end{vmatrix} = 4 - 6 = -2$$

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 4 & 0 \\ 3 & 9 \end{vmatrix} = -(36 - 0) = -36$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} 0 & 2 \\ 10 & 2 \end{vmatrix} = 0 - 20 = -20$$

$$C_{32} = (-1)^{3+2} \begin{vmatrix} 4 & 2 \\ 2 & 2 \end{vmatrix} = -(8 - 4) = -4$$

$$C_{33} = (-1)^{3+3} \begin{vmatrix} 4 & 0 \\ 2 & 10 \end{vmatrix} = 40 - 0 = +40$$

$$\text{Cofactor of A} = \begin{bmatrix} -8 & 4 & -12 \\ 18 & -2 & -36 \\ -20 & -4 & 40 \end{bmatrix}$$

$$\text{Adj A} = (\text{Cofactor of A})' = \begin{bmatrix} -8 & 18 & 20 \\ 4 & -2 & -4 \\ -12 & -36 & 40 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{-56} \begin{bmatrix} -8 & 18 & 20 \\ 4 & -2 & -4 \\ -12 & -36 & 40 \end{bmatrix} = \begin{bmatrix} \frac{-8}{-56} & \frac{18}{-56} & \frac{20}{-56} \\ \frac{4}{-56} & \frac{-2}{-56} & \frac{-4}{-56} \\ \frac{-12}{-56} & \frac{-36}{-56} & \frac{40}{-56} \end{bmatrix} = \begin{bmatrix} \frac{1}{7} & \frac{-9}{28} & \frac{5}{14} \\ \frac{-1}{14} & \frac{1}{28} & \frac{1}{14} \\ \frac{3}{14} & \frac{9}{14} & \frac{-5}{7} \end{bmatrix}$$



## Solution of a System of Linear Equations

### By Matrix Inversion Technique

$$AX = B$$

$$X = A^{-1} B$$

**Illustration 8:** Solve:

$$2x_1 + 3x_2 - x_3 = 9; x_1 + x_2 + x_3 = 9; 3x_1 - x_2 - x_3 = -1$$

**Solution:**

$$AX = B$$

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 1 & 1 \\ 3 & -1 & -1 \end{bmatrix} X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} B = \begin{bmatrix} 9 \\ 9 \\ -1 \end{bmatrix}$$

$$X = A^{-1} B$$

$$A^{-1} = \frac{1}{|A|} \text{adj}A$$

$$|A| = \begin{vmatrix} 2 & 3 & -1 \\ 1 & 1 & 1 \\ 3 & -1 & -1 \end{vmatrix} = 2 \begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} - 3 \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix}$$

$$= 2[(1 \times -1) - (-1 \times 1)] - 3[(1 \times -1) - (3 \times 1)] - 1[(1 \times -1) - (3 \times 1)]$$

$$= 2(-1 + 1) - 3(-1 - 3) - 1(-1 - 3) = 2(0) - 3(-4) - 1(-4)$$

$$= 0 + 12 + 4 = 16 \neq 0$$

$\therefore A^{-1}$  exists.

Adj A = (Cofactor of A)'





Cofactor of A:

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 1 & 1 \\ -1 & -1 \end{vmatrix} = 0$$

$$C_{12} = (-1)^{1+2} \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} = +4$$

$$C_{13} = (-1)^{1+3} \begin{vmatrix} 1 & 1 \\ 3 & -1 \end{vmatrix} = -4$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 3 & -1 \\ -1 & -1 \end{vmatrix} = +4$$

$$C_{22} = (-1)^{2+2} \begin{vmatrix} 2 & -1 \\ 3 & -1 \end{vmatrix} = +1$$

$$C_{23} = (-1)^{2+3} \begin{vmatrix} 2 & 3 \\ 3 & -1 \end{vmatrix} = +11$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} 3 & -1 \\ 1 & 1 \end{vmatrix} = +4$$

$$C_{32} = (-1)^{3+2} \begin{vmatrix} 2 & -3 \\ 1 & 1 \end{vmatrix} = -3$$

$$C_{33} = (-1)^{3+3} \begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = -1$$

$$\text{Cofactor of A} = \begin{bmatrix} 0 & 4 & -4 \\ 4 & 1 & 11 \\ 4 & -3 & -1 \end{bmatrix}$$

$$\therefore \text{Adj A} = (\text{Cofactor of A})^T = \begin{bmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{16} \begin{bmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{bmatrix}$$



$$\mathbf{A}^{-1} \mathbf{B} = \frac{1}{16} \begin{bmatrix} 0 & 4 & 4 \\ 4 & 1 & -3 \\ -4 & 11 & -1 \end{bmatrix} \begin{bmatrix} 9 \\ 9 \\ -1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 32 \\ 48 \\ 64 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

$$\text{But } \mathbf{X} = \mathbf{A}^{-1} \mathbf{B} \text{ i.e., } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

$$x_1 = 2, x_2 = 3 \text{ and } x_3 = 4.$$

Substitute these values of  $x_1$ ,  $x_2$ , and  $x_3$  in the equations and see whether they are satisfied to ensure that the solution is correct.

\*\*\*\*\*

### Practical Problems:

1. Find  $\mathbf{A} + \mathbf{B}$  and  $\mathbf{A} - \mathbf{B}$  if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} -1 & -2 & -4 \\ -1 & -2 & -4 \\ 1 & 2 & 4 \end{bmatrix}$$

2. Find  $\mathbf{AB}$  and  $\mathbf{BA}$  when

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} -1 & -2 & -1 \\ 6 & 12 & 6 \\ 5 & 10 & 5 \end{bmatrix}$$

3. If  $\mathbf{A} = \begin{bmatrix} 2 & 5 & -1 \\ 3 & -1 & 2 \\ 7 & 2 & -3 \end{bmatrix}$  find  $\mathbf{A}^{-1}$

4. Solve by using matrix inversion method.

$$2x - y + 3z = 1;$$

$$x + y + z = 2;$$

$$x - y + z = 4.$$

\*\*\*\*\*



Sub. Code: DB 1 C/ DNB 1 C

**BBA DEGREE EXAMINATION, MAY 2016.**

**First Year/Non-Semester**

**Business Administration - Main (DD & CE)**

**QUANTITATIVE METHODS/ BUSINESS MATHEMATICS AND STATISTICS**

**Time: Three hours**

**Maximum: 100 marks**

**PART A – (5 x 5 = 25 marks)**

**Answer any FIVE questions out of Eight.**

1. Define. What are the various methods of measuring dispersion?
2. Why are there in general two regression equations?
3. Evaluate  $\lim_{x \rightarrow 3} \left( \frac{x^2 - 9}{x - 3} \right)$
4. Find  $\int \frac{1}{(1 + \cos x)} dx$ .
5. Is  $AB = BA$ ? (Where A and B are 3 x 3 matrices). Exemplify your answer.
6. Calculate the S.D. for the data 9, 27, 18, 54, 45, 72, 36, 63, 87.
7. Find out the regression equation of sales on profit from the given data:

	Average	S.D.
Sales (in '000)	33	8.6
Profit (in '000)	12	17.4

8. Evaluate:  $\lim_{n \rightarrow \infty} \frac{1+2+\dots+n}{n^2}$ .

**PART B – (5 x 15 = 75 marks)**

**Answer any FIVE questions out of Eight.**

9. The coefficient of Q.D. in a symmetric distribution is 0.7. The first and 3<sup>rd</sup> quartiles are 21 and 42. Calculate the value of median.



10. For two variables  $x$  and  $y$  the equation of regression lines are  $9y - x - 288 = 0$  and  $x - 4y + 38 = 0$ , find

(a)  $\bar{x}$  and  $\bar{y}$

(b) The coefficient of correlation between  $x$  and  $y$

(c) The ratio of the S.D. of  $y$  to that of  $x$ .

11. If  $y = (\log x)^2$  show that  $x^2 y_2 + xy_1 = 2$ .

12. Integrate  $\frac{1}{x(\log x)^n}$  w.r.t.  $x$ .

13. Solve for  $X$  and  $Y$  if  $3X + 2Y = I$  and  $2X - Y = 0$  where  $I, 0$  are unit and null matrix of order three.

14. Calculate MD and its co-efficient from the data given below:

<b>X</b>	10	20	30	40	50	60	70	80	90
<b>F</b>	35	47	52	69	78	57	43	32	15

15. Given data: variance of  $x = 25$ ;  $y$  on  $x \rightarrow 5x - y = 28$ ;  $y$  on  $64x - 45y = 24$ . Find mean value of  $x$  and  $y$ ; coefficients of correlation between  $x$  and  $y$ , and standard deviation of  $y$ . Assume other data if required.

16. If  $x = a(\cos t + t \sin t)$ ,  $y = a(\sin t - t \cos t)$  find  $\frac{d^2 y}{dx^2} = ?$

\*\*\*\*\*



Sub. Code: DB 1 C/ DNB 1 C

**BBA DEGREE EXAMINATION, DECEMBER 2016.**

**First Year/Non-Semester**

**Business Administration – Main (DD & CE)**

**QUANTITATIVE METHODS/BUSINESS MATHEMATICS AND STATISTICS**

**Time: Three hours**

**Maximum: 100 marks**

**PART A – (5 x 5 = 25 marks)**

**Answer any FIVE questions out of Eight.**

1. Define G.M. Calculate the G.M. for 25, 37, 356, 545, 976, 500, 750.
2. Calculate the S. D. for the data: 9, 27, 18, 54, 45, 72, 36, 63, 87.
3. What is correlation analysis? Name the types of correlation.
4. Distinguish between correlation and regression.
5. Prove:  $\frac{d(x^n)}{dx} = nx^{n-1}$
6. Find:  $\frac{dy}{dx} = y^2 = ax$ .
7. Integrate the following with respect to  $x$ :  $\left(x + \frac{1}{3}\right)^3$
8. If  $X = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$   $Y = \begin{bmatrix} 4 & 2 \\ 2 & 4 \\ 1 & 1 \end{bmatrix}$ : find  $X, Y$ .

**PART B – (5 x 15 = 75 marks)**

**Answer any FIVE questions out of Eight.**

9. Briefly explain the various measures of control tendency.
10. Calculate the G.M. from the following data:



<b>Income (Rs.) :</b>	45	52	59	62	75	80
<b>No. of families :</b>	3	9	6	5	2	1

11. Calculate Co – efficient of Q.D. for the following data:

<b>Wages ( per week)</b>	<b>No. of wage earners</b>
Less than 349	14
350 – 379	62
380 – 409	99
410 – 430	18
Over 430	7

12. Find the rank correlation coefficient for the following data:

<b>x:</b>	92	89	87	86	86	77	71	63	53	50
<b>y:</b>	86	83	91	77	68	85	52	82	37	57

13. Find the coefficient of correlation between  $x$  and  $y$  from the following data:

<b>x:</b>	15	20	25	30	35	40
<b>y:</b>	90.5	100.8	120.7	140.9	160.1	170.8

14. What are the various properties of limits?

15. Find matrices  $M$  and  $G$  of order 2 such that  $2M - 3G = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$ ;  $2M + 3G = \begin{bmatrix} 7 & 1 \\ 4 & 5 \end{bmatrix}$ .

16. If  $K = \begin{bmatrix} 5 & 4 & -2 \\ 4 & 5 & -2 \\ -2 & -2 & 2 \end{bmatrix}$  then, show that  $(A - 10I) (A - I) = 0$ .

\*\*\*\*\*

**Prepared by**

**Miss. K. MEENA, MBA, M.Phil., UGC-NET-MGT.,**

Assistant Professor, Department of Business Administration,

Manonmaniam Sundaranar University Constituent College of Arts and Science,

Sathankulam - 628 704, Thoothukudi - District.