Chapter 6

Statistical Measures and analysis

6.1 Measures of Central Tendency

6.1.1 Average

An average is a single significant figure which sums up the characteristics of a group of figures. It conveys a general idea of the whole group. It is generally located at the centre or the middle of distribution. Various measures of central tendency are

1. Arithmetic Mean
2. Median
3. Mode
4. Geometric Mean
5. Harmonic Mean

6.1.2 Arithmetic Mean

- Arithmetic Mean is a mathematical average.
- It is a method of representing the whole data by one figure.
- It is a simple measure and widely used.

Arithmetic Mean in Individual Series

If x_1, x_2, x_3, ... x_n are n individual values, then arithmetic mean of them is given by

$$\frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum x}{n}$$

Solved Problem : Find the arithmetic mean of the 10 values given:
10, 90, 85, 103, 11, 29, 84, 15, 35, 80

Solution: $\frac{10 + 90 + 85 + 103 + 11 + 29 + 84 + 15 + 35 + 80}{10} = \frac{542}{10} = 54.2$

6.1.3 Arithmetic Mean in Discrete frequency distribution

When the data are in discrete frequency distribution form, let x_1, x_2, x_3, ... x_n be the values of the variable with corresponding frequencies f_1, f_2, f_3, ... f_n respectively

Then arithmetic mean $= \frac{x_1.f_1 + x_2.f_2 + x_3.f_3 + ... + x_n.f_n}{f_1 + f_2 + f_3 + ... + f_n} = \frac{\sum f_i.x_i}{N}$

Example : Calculate mean from the following data

Value	5	15	25	35	45	55	65	75
Freq	15	20	25	24	12	31	71	52

Value	Frequency	f × x
5	15	75
15	20	300
25	25	625
35	24	840
45	12	540
55	31	1705
65	71	4615
75	52	3900
Total	250	12600

$$\text{Mean} = \frac{\sum f \times x}{N} = \frac{12600}{250} = 50.4$$

Mean in a Continuous Series

In the case of Continuous series A. M. is given by

$$\text{A.M.} = \frac{\sum f.m}{N}$$

where m = midpoint of the the class.

Example : Calculate the arithmetic mean of the following data.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	5	4	7	12	10	8	4

Solution:

Value	Frequency	Mid value	f × m
10-20	5	15	75
20 -30	4	25	100
30 -40	7	35	245
40-50	12	45	540
50-60	10	55	550
60-70	8	65	520
70-80	4	75	300
	N = 80		$\sum f.m = 2330$

$$\text{A. M.} = \frac{\sum f.m}{N} = \frac{2330}{80} = 29.125$$

Merits of Mean

1. Simple to understand.
2. Easy to calculate.
3. Based on all observation.
4. Capable for algebraic treatment.
5. It is more stable.
6. It can be determined in most of the cases.

Demerits

1. It is effected by extreme values.
2. Not suitable for ratios and percentages.
3. It cannot be calculated for qualitative data.
4. Misleading and absurd result in some cases.
5. It does not coincide with any observed values.

'Mean' the Best Average for the following reasons

1. Easily calculated.
2. Simple to under stand.
3. Capable for algebraic treatment.
4. Based on all observation.
5. It can located without arranging the data.
6. It is stable as it does not differ from sample to sample, when sample selected is sufficiently large.

6.2 Median

6.2.1 Median in Individual Series

Median is the middle value. When the data are arranged in ascending order or descending order the middle value is the median. It is the positional average.

Procedure for Mean in Individual Series

- First arrange the data in ascending order.
- Then determine the size of the $\frac{(n+1)}{2}^{th}$ item to get the Median.

When the value of 'N' is odd Example : Find the median of the following data:

4, 25, 45, 15, 26, 35, 55, 28, 48

Solution 4, 15, 25, 26, 28, 35, 48, 48, 55

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

$$\text{Median} = 28$$

When the value of 'N' is even Example : Find the median of the following data: 4, 25, 45, 15, 18, 26, 35, 55, 28, 48

Solution : 4, 15, 18, 25, 26, 28, 35, 48, 48, 55

$$\text{Median} = \text{Value of } \frac{(n+1)}{2}^{th} \text{ item} = \text{Value of } \frac{(10+1)}{2}^{th} \text{ item}$$

$$\text{Value of 5.5 th item} = \frac{5^{th} item + 6^{th} item}{2} = \frac{26+28}{2} = 27$$

6.2.2 Median in Discrete series

In discrete series, Median is the Value of $\dfrac{(n+1)}{2}^{th}$ item.

- First arrange the data as cumulative frequency distribution.
- Then determine the value of $\dfrac{(n+1)}{2}^{th}$ item .

Problem:

Calculate the value of median in the following series.

Size	5	8	10	15	20	25
Frequency	3	12	8	7	5	4

Solution:

Size	Frequency	C. F.	
5	3	3	
8	12	15	
10	8	23	$\leftarrow \frac{N}{2}$
15	7	30	
20	5	35	
25	4	39	

Median = Value of $\dfrac{(n+1)}{2}^{th}$ item = Value of $\dfrac{(39+1)}{2}^{th}$ item = Value of 20^{th} item. = 10

Median in Continuous Distribution Procedure

- First form the cumulated frequency distribution.
- Then determine the median class by the formula $\dfrac{N}{2}$.
- Use the formula, Median $= l_1 + \dfrac{(\frac{N}{2} - c.f.)}{f} \times C$

Where 'l_1' is the lower limit of median class, 'c.f.' is the cumulated frequency of the preceding class, 'f' is the frequency of the median class and 'C' is the class width.

Eg. Calculate median for the following data

Class	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Freq.	29	195	241	117	52	10	6	3	3

Solution:

Class	Frequency	C.F.	
0-5	29	29	
5-10	195	224	
10-15	241	465	$\leftarrow \frac{N}{2}$
15-20	117	582	
20-25	52	634	
25-30	10	644	
30-35	6	650	
35-40	3	653	
40-45	3	656	

Median class $= \dfrac{N}{2} = \dfrac{656}{2} = 328^{th}$ item

$$\text{Median} = l_1 + \frac{(\frac{N}{2} - c.f.)}{f} \times C$$
$$\text{Median} = 10 + \frac{(328 - 224)}{241} \times 5$$
$$\text{Median} = 12.2$$

Merits of Median

- It is very simple measure.
- Not affected extreme items.
- It can locate by a mere glance.
- It can be determined by graphically.

Demerits of Median

- Not based on all observation.
- Not capable of algebraic treatments.
- It requires arraying also.
- In the case of continuous series interpolation formula is used.(Only an approximate value)

6.3 Mode

6.3.1 Introduction

Mode is the value of the item of a series which occurs most frequently.

Mode is the most repeated value in a series.

According to Kenny, "The value of the variable which most frequently in a distribution is called the mode.

6.3.2 Mode in Individual Series

In individual series, the value which occurs more number of times is mode.

Example:

Find mode for the following series.

23, 35, 28, 42, 62, 53, 35, 28, 42, 35, 23, 42, 35.

In this series, 35 appears more number of times. Hence Mode = 35.

In some cases, there may not be any repeated value. In that case Mode is said to be **ill defined**. In that case Mode is calculated by the approximation formula

Mode = 3 Median - 2 mean.

Example : Find Mode for the following distribution.

40, 25, 60, 35, 81, 75, 90, 10

X
10
25
35
40
60
75
81
90
416

$$\text{Mean} = \frac{\sum x}{n} = \frac{416}{8} = 52$$

$$\text{Median} = \text{size of the } \frac{(n+1)}{2} \text{ th item.}$$
$$\text{size of the } 4.5^{th} \text{ item.} = \frac{(40+60)}{2} = 50$$
$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$
$$3(50) - 2(52) = 150\text{-}104 = 46$$

6.3.3 Mode in a Discrete Frequency Series

In the case of a discrete frequency series, the value having the highest frequency is taken as the mode.

Example :

Size	5	8	10	12	29	35	40	46
Frequency	3	12	25	40	31	20	18	7

The value **12** has the **highest frequency**. Hence **12** is the **mode**.

6.3.4 Mode in Continuous Frequency Series

In a continuous frequency distribution, Mode lies in the class having highest frequency. But since it is a continuous class, exact size cannot be determined. So from the modal class it is determined using interpolation formula.

$$\text{Mode} = l_1 + \frac{(f_1 - f_0) \times c}{2f_1 - f_0 - f_2}$$

Where 'l_1' is the lower limit of the model class, f_0, f_2 are frequencies of classes just preceding and succeeding model class, f_1 is the frequency of the model class. Problem: Calculate mode from the following data:

Size	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
Freq.	4	8	18	30	20	10	5	2

Solution:

Class	Freq.	
10-15	4	
15-20	8	
20-25	18	
25-30	30	Modal Class
30-35	20	
35-40	10	
40-45	5	
45-50	2	

$$\text{Mode} = l_1 + \frac{(f_1 - f_0) \times c}{2f_1 - f_0 - f_2}$$

$l_1 = 25$, $f_1 = 30$, $f_0 = 18$, $f_2 = 20$, $c = 5$

$$\text{Mode} = 25 + \frac{(30-18) \times 5}{60 - 18 - 20}$$
$$= 25 + \frac{60}{22} = 27.73$$

6.3.5 Merits and Demerits of Mode

Merits of Mode

1. Can be located graphically.
2. It is simple measure of central tendency.
3. Less affected by extreme values.
4. It coincide with one of the values of the series.
5. It is the best representative of series.

Demerits of the Series

1. Not base on all observations.
2. Not capable of further algebraic treatment.
3. It is ill-defined in some cases.
4. Sometimes grouping to identify the Modal values.

6.3.6 Comparison Between Mean, Median and Mode

Comparing Mean, Median and Mode

Mean	Median	Mode
1.Mathematical average	positional average	positional avera
2.Based on all observations	Middle item	Most repeated
3.Capable of further mathematical treatment	Not capable of mathematical treatment	Not capable of matical treatm

Comparison Contd...

Mean	Median	Mode
4. It may not be a value found in the series.	It is a value found in series	It is a value found in series
5. Not graphically located	Can be located graphically	Can be located graphically
6. Usually obtained by calculation	Can be obtained by mere inspection	Can be obtained by mere inspection

Comparison Contd...

Mean	Median	Mode
7. Affected by extreme values	Not affected by extreme items	Not affected by extreme items
8. Mean is always well defined	It is well defined	In some cases ill defined.

6.3.7 Relationship Between Mean, Median and Mode

Relationship B/W Mean, Median and Mode

In symmetrical distribution, Mean = Median = Mode.

In moderately unsymmetrical distributions, Mean, Median and Mode will have different values. The equation connecting them is

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

6.4 Geometric Mean

6.4.1 Geometrical Mean in Individual Series

If there are 'n' values in a series, then their Geometric Mean is defined as the n^{th} root of the product of those n values. Geometrical mean is a mathematical average. If x_1, x_2, x_3, ... x_n are n values of an individual series, then Geometric mean of the series is $\sqrt[n]{x_1 \times x_2 \times x_3 \times .... \times x_n}$

It is also given by

$$\text{Geometric Mean} = \text{Antilog of} \left[\frac{log\ x_1 + log\ x_2 + ... + log\ x_n}{n}\right]$$

$$\text{Geometric Mean} = \text{Antilog} \left[\frac{\sum_{i=1}^{i=n} log\ x_i}{n}\right]$$

Problem 1 Calculate geometric Mean for the following series, 57.5, 87.75, 53.5, 73.5, 81.75

Solution:

Size	log x
57.5	1.7597
87.75	1.9432
53.5	1.7284
73.5	1.8663
81.75	1.9125
	9.2101

$$\text{Geometric Mean} = \text{Antilog} \left[\frac{\sum_{i=1}^{i=n} log\ x_i}{n}\right] = \text{Antilog} \left[\frac{9.2101}{5}\right] = \text{Antilog of } (1.8420) = 69.51$$

6.4.2 Geometric Mean in Discrete Series

For discrete frequency distribution Geometric Mean is given by

$$\text{Geometric Mean} = \text{Antilog} \left[\frac{\sum_{i=1}^{i=n}(f \times log\ x_i)}{n}\right]$$

Eg. Find the G. M. from the following data:

Size	5	8	10	12
Frequency	2	3	4	1

Solution:

x	f	log x	f × log x
5	2	0.6990	1.3980
8	3	0.9031	2.7093
10	4	1.0000	4.0000
12	1	1.0792	1.0792
			9.1865

$$\text{G.M.} = \text{Antilog} \left[\frac{\sum(f \times log\ x_i)}{n}\right]$$

$$= \text{Antilog} \left[\frac{9.1865}{10}\right] = 8.292$$

6.4.3 G.M in Continuous Frequency Distribution

Geometric Mean in a continuous frequency distribution is given by = Antilog $\left[\frac{\sum(f \times log\ x_i)}{n}\right]$ where 'x' is the mid point of the class.

Example : Calculate G. M.for the following data:

Size	100-104	105-109	110-114	115-119	120-124	125-129
Freq.	24	30	45	65	72	84

Solution

Size	Mid Value(x)	Freq.	log x	f×log x
100-104	24	102	2.0086	48.2064
105-109	30	107	2.0294	60.8820
110-114	45	112	2.0492	92.2140
115-119	65	117	2.0682	134.4330
120-124	72	122	2.0864	150.2208
125-129	84	127	2.1038	176.7192
	320			662.6754

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum(f \times log\ x_i)}{n}\right] = \text{Antilog}\left[\frac{662.6754}{320}\right] = 117.7$$

Merits of G. M.

- It is based on all observations.
- It is rigidly defined.
- It is useful in averaging ratios and percentages.
- It is not affected by extreme values.
- It is capable of algebraic treatment.

Demerits of G. M.

- It is difficult to understand.
- It cannot be compared when there are both negative and positive values or when one or more of the values are zero.

Application Geometric mean is useful to find the average percentage increase in sales, production, etc.

6.5 Harmonic Mean

Harmonic Mean of a set of 'n' values is defined as the reciprocal of the mean of the reciprocals of those values.

Harmonic mean is used when it is desired to give greatest weightage to the smallest items.

It is applied in averaging rates, times, etc.

Expression for Harmonic Mean If x_1, x_2, x_3, ... x_n are n values of an individual series, then Harmonic Mean is given by

$$\text{Harmonic Mean} = \left[\frac{n}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ... + \frac{1}{x_n}\right)}\right] = \left[\frac{n}{\sum \frac{1}{x}}\right]$$

Example Calculate the Harmonic Mean of the following series. 2, 3, 4, 5.

x	$\frac{1}{x}$
2	0.50
3	0.33
4	0.25
5	0.20
n = 4	$\sum$= 1.28

$$\text{Harmonic Mean} = \left[\frac{n}{\sum \frac{1}{x}}\right] = \frac{4}{1.28} = 3.125$$

6.5.1 Harmonic Mean in a Discrete Series

In the case of discrete frequency series distribution, the Harmonic Mean is given by

$$\text{H. M.} = \left[\frac{n}{\sum f \times \frac{1}{x}}\right]$$

E.g. Find the Harmonic Mean from the following data:

Size	6	10	14	18
Freq.	20	40	30	10

Solution

Size (x)	Frequency	$\frac{1}{x}$	f $\times \frac{1}{x}$
6	20	0.1667	3.334
10	40	0.1000	4.000
14	30	0.0714	2.142
18	10	0.0556	0.556
	$\sum$ = 100		$\sum$ = 10.032

$$\text{H.M.} = \left[\frac{n}{\sum f \times \frac{1}{x}}\right] = \frac{100}{10.032} = 9.97$$

6.5.2 Harmonic Mean in Continuous Frequency Distribution

In continuous frequency distribution harmonic mean is given by

$$H.M. = \left[\frac{n}{\sum f \times \frac{1}{x}}\right], \text{ 'x' is the mid value of the class}$$

Eg. Find the H. M. for the following distribution.

Class	10-20	20-30	30-40	40-50	50-60
Freq.	4	6	10	7	3

Solution:

Class	Mid Value	freq.	$\frac{1}{x}$	f $\times \frac{1}{x}$
10-20	15	4	0.0670	0.2680
20-30	25	6	0.0400	0.2400
30-40	35	10	0.0290	0.2900
40-50	45	7	0.1540	0.1540
50-60	55	3	0.0180	0.0540
		$\sum = 30$		$\sum = 1.0060$

$$\text{H.M.} = \left[\frac{n}{\sum f \times \frac{1}{x}}\right] = \frac{30}{1.0060} = 29.82$$

Merits of H. M.

- It possess almost all characteristic of a good average.
- It is based on all observations.
- It is amenable to more algebraic treatment.
- It is not affected much by sampling fluctuations.
- It can be used to average relative values.

Demerits of Harmonic Mean

- It is difficult to calculate and not easy to understand.
- It gives greater weight to larger values.
- It cannot be calculated when there are both negative and positive items or when an item is zero.
- It is not a popular average.

Applications Harmonic mean is useful for:

- For computing the average rate of increase in profits.
- Average rate of speed.
- Average price etc.

6.6 Measures of Dispersion

- Dispersion refers to the variability in series of items.
- It speaks about the spread or scatter of this values in a series.
- Measure of dispersion measure the variability in a series.
- They tell us the extent to which the values of a series differ between each other or from their average.

Measures of Dispersion are classified into

1. Absolute measures. and
2. Relative Measures.

Absolute Measures Absolute measure of dispersion are expressed in the same units in which the data are collected. They measure variability in a series.

Various Absolute measures of dispersion are

1. Range
2. Quartile deviation
3. Mean Deviation. and
4. Standard Deviation.

Relative Measures A relative measure of dispersion is the ratio of the measure of dispersion to an appropriate average from which deviations are measured. It is also called coefficient of dispersion.

Relative measure of dispersion are useful for comparing two series for their variability. Greater the value of relative measure of dispersion in a series, greater is the variability in a series.

The important relative measures of dispersion are

1. Coefficient of Range.
2. Coefficient of Quartile Deviation.
3. Coefficient of Mean Deviation.
4. Coefficient of Variation.

The most commonly used relative measure of dispersion is coefficient of variation.

6.6.1 Range

1. Range is the simplest possible measure of dispersion.
2. Range is the simplest possible measure of dispersion.
3. It is the difference between the highest and lowest values in a series.
4. Range = H - L, 'H' is the Highest value and 'L' is the Lowest value.
5. Range there for measures the maximum variation in the values of a series.

Range is an absolute measure. The relative measure based on range is the coefficient of range.

Coefficient of range $= \dfrac{H-L}{H+L}$

Range in Individual Series

Eg. Find the range in the series: 25, 32, 85, 32, 42, 10, 20, 18, 28.

Solution: Range = H- L , ie 85 - 10 =75

Range in Discrete Series

Example: Calculate the range and Coefficient of range for the following distribution.

Size	5	8	10	15	20	25
Frequency	3	12	8	7	5	4

Range = H - L ; = 25 - 5 = 20

Coefficient of Range

$$\text{Coefficient of Range} = \frac{H-L}{H+L} = \frac{25-5}{25+5} = \frac{20}{30} = 0.6667$$

Uses of Range

- Range is used in certain fields to measure variability, particularly in those data where variation is not much. (e.g. Doctors are interested in range in body temperature)
- It is used in quality control.
- Range is used to study variation in prices of shares and interest rates.
- In weather forecasts, the minimum and the maximum temperature of every day are studied (To predict the range within which the temperature may vary).

6.7 Quartile Deviation

6.7.1 Definition

Quartile Deviation is defined as half the distance between the third and first quartiles.

Quartile Deviation is am absolute measure and it is given by

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

Q_1 and Q_3 are the first quartile and third quartile respectively.

The corresponding relative measure is the Coefficient of Quartile Deviation which is given by

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

QD in Individual Series

$Q_1 = \text{size of} \left(\frac{n+1}{4}\right)^{th}$ item, when the items are arranged in ascending order.

$Q_3 = \text{size of} \left(\frac{n+1}{4} \times 3\right)^{th}$ item, when the items are arranged in ascending order.

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

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

$$\text{Inter Quartile Range} = Q_3 - Q_1$$

Example Compute Quartile measure of dispersion, Interquartile range, and Coefficient of Quartile Deviation for the following series.
23, 25, 8, 10, 9, 29, 45, 85, 10, 16, 24.

Solution If values are arranged in the ascending series,
8, 9, 10, 10, 16, 23, 24, 25, 29, 45, 85.; N = 11

$Q_1 = \text{size of} \left(\frac{11+1}{4}\right)^{th}$ item = size of 3^{rd} item = 10

$Q_3 = \text{size of} \left(\frac{11+1}{4} \times 3\right)^{th}$ item = size of 9^{th} item = 29

Q. D. Quartile Deviation $= \frac{Q_3 - Q_1}{2} = \frac{29 - 10}{2} = 9.5$

Interquartile Range Inter Quartile Range = $Q_3 - Q_1 = 19$

Coefficient of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{29 - 10}{29 + 10} = \frac{19}{39} = 0.49$

6.7.2 QD in Discrete Series

Quartile Deviation $=\dfrac{Q_3 - Q_1}{2}$,

$Q_1 = \text{size of } \left(\dfrac{N+1}{4}\right)^{th}$ item, $N = \sum f$

$Q_3 = \text{size of } \left(\dfrac{3N+1}{4}\right)^{th}$ item, $N = \sum f$

Example Find the QD for the following series

Size	5	8	10	12	19	20	32
Freq.	3	10	15	20	8	7	6

Solution

Size	Freq	Cum. Freq
5	3	3
8	10	13
10	15	28
12	20	48
19	8	56
20	7	63
32	6	69
	$\sum$N=69	

$Q_1 = \text{size of } \left(\dfrac{N+1}{4}\right)^{th}$ item = 17.5 item size = 10

$Q_3 = \text{size of } \left(\dfrac{3N+1}{4}\right)^{th}$ item =52.5 item size = 19

$$\text{QD} = \frac{Q_3 - Q_1}{2} = \frac{19 - 10}{2} = 4.5$$

6.7.3 QD in a Continuous Series

Here QD is the the same, i.e. QD $= \dfrac{Q_3 - Q_1}{2}$

Q_1 & Q_3 are size of $\left(\dfrac{N}{4}\right)$th item and size of $\left(\dfrac{N}{4} \times 3\right)$th item respectively. They cannot be determined directly and is determined graphically by the interpolation formula which is given by.

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c \quad \text{and} \quad Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c$$

Where

'l_1' is the lower limit of the quartile class.

'f' is the frequency of the class.

'cf' is the cumulative frequency of the preceding class.

'c' is the class interval of the quartile class.

Example Find the quartile deviation for the following frequency distribution.

Age	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No.	15	30	53	75	100	110	115	125

Solution First you have to form the cumulative frequency distribution and determine Q_1 and Q_3

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c \qquad Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c$$

Size	f	cf	
0-10	15	15	
10-20	30	45	
20-30	53	98	
30-40	75	173	$\frac{N}{4}$
40-50	100	273	
50-60	110	383	
60-70	115	498	$\frac{3N}{4}$
70-80	125	623	
	$\sum = 623$		

$$\frac{N}{4} \text{ item size} = \frac{623}{4} \text{ th item} = 155.75$$

Lower quartile class is 30 - 40

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c = 30 + \left(\frac{(155.75 - 98)}{75}\right) \times 10 = 37.7$$

$$\frac{3N}{4} \text{ item size} = \frac{3 \times 623}{4} \text{ th item} = 467.5$$

Lower quartile class is 60 - 70

$$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c = 60 + \left(\frac{(467.5 - 383)}{115}\right) \times 10 = 67.33$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{67.33 - 37.7}{2} = 14.815$$

6.8 Mean Deviation

6.8.1 Definition

Mean Deviation is defined as the arithmetic mean of deviations of all the values in a series from their average, counting all such deviations as positive. The average selected may be mean, median or mode.

$$\therefore Mean \quad Deviation = \frac{\sum |\, d \,|}{n}$$

where '$|\, d \,|$' represents deviation from an average without sign. 'n' being number of items.

Mean deviation is an absolute measure. Its relative value is Coefficient of Mean Deviation. It is equal to the ratio of Mean deviation to Average from which Mean Deviation is computed.

If Mean is the chosen average,

$$\text{Coefficient of Mean Deviation} = \frac{Mean \quad Deviation}{Mean}$$

6.8.2 Mean Deviation in Individual Series

Example Find the mean deviation from mean and its coefficient for the following values 25,63,85,75,62,70,83

Solution First you have to find out the Mean.
Here the mean is obtained as

$$\text{Mean} = \frac{25 + 63 + 85 + 75 + 62 + 70 + 83 + 28 + 30 + 12}{10} = 53.3$$

x	|d|= x - 53.3 without sign
25	283
63	9.7
85	31.7
75	21.7
62	8.7
70	16.7
83	29.7
28	25.3
30	23.3
12	41.3
533	236.4

Then find out deviations '$\mid d \mid$ for each value i.e. 25-53.3 = 28.3

Mean Deviation $= \frac{\sum \mid d \mid}{n} = \frac{236.4}{10}$ =23.64

Coefficient of mean deviation $= \frac{Mean \;\; Deviation}{Mean} = \frac{23.64}{53.3} = 0.44$

Mean Deviation from Median Calculate mean deviation from median and its coefficient for the following values, 5, 25, 28, 33,35, 44, 82, 83, 87, 96, 99 Solution Median = Size of $\frac{(N+1)}{2}$ th item.= Size of 6th item = 44.
Find out the deviation of each value from median. Find the sum of deviation.

x	|d|=x - 44 without sign
5	39
25	19
28	16
33	11
35	9
44	0
82	38
83	39
87	43
96	52
99	55
	321

Mean Deviation $= \frac{321}{11} = 29.18$.

Coefficient of Mean Deviation $= \frac{29.18}{44} = 0.66$

6.8.3 Mean Deviation in a Discrete Series

Mean Deviation in Discrete Frequency Series In the case of discrete frequency series, MD is given by,

$$\text{MD} = \frac{\sum f \times \mid d \mid}{N}$$

Where 'f' is the frequency corresponding to the given size, '$\mid x \mid$' and 'N' is the total frequency.

Example Calculate the MD for the following series from mean.

Size	0	1	2	3	4	5	6
Frequency	171	82	50	25	13	7	2

x	f	fx	|d| =|x – 1.02|	f|d|
0	171	0	1.02	174.42
1	82	82	0.02	1.64
2	50	100	0.98	49.00
3	25	75	1.98	49.5
4	13	52	2.98	38.74
5	7	35	3.98	27.86
6	2	12	4.98	9.96
	350	356		351.12

solution Mean of the series $= \frac{\sum f \times x}{N} = \frac{356}{350} = 1.02$

Mean Deviation $= \frac{\sum f \times \mid d \mid}{N} = \frac{351.52}{350} = 0.98$

6.8.4 MD in Continuous Frequency Distribution

MD in continuous frequency distribution is given by:

Mean Deviation $= \frac{\sum f \times \mid d \mid}{N}$

It is same as discrete frequency series. The only difference is that here the mid value of the class is taken as 'x'.

Example Calculate the MD from mean for the following frequency distribution.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	6	10	20	10	6	4

Mean of the series $= \frac{\sum f \times x}{N} = \frac{2100}{60} = 35$

Mean Deviation $= \frac{\sum f \times \mid d \mid}{N} = \frac{680}{60} = 11.33$

6.8.5 Merits and Demerits of Mean Deviation

Merits

- Mean deviation is a very simple and an easy measure of dispersion.
- It is easily understood.
- It is based on all the items of the series. So it is more representative.
- Mean deviation is less affected by extreme values.

Marks	f	Mid x	fx	|d| =|x – 35|	f|d|
0-10	4	5	20	30	120
10-20	6	15	90	20	120
20-30	10	(25)	250	10	100
30-40	20	35	700	0	0
40-50	10	45	450	10	100
50-60	6	55	330	20	120
60-70	4	65	260	30	120
	60		2100		680

Demerits

- Mean deviation suffers from inaccuracy because '+' or '—' signs are ignored.
- Mean deviation is not capable of any further algebraic treatment.
- Mean deviation is not reliable measure when calculated from Mode as the Mode is uncertain in some cases.

Uses of MD

- Mean deviation is significantly used (or measuring variability of the series relating to Economic and Social phenomena.
- Variability in the distribution of wealth and income is generally measured in terms of Mean deviation.

6.9 Standard Deviation

6.9.1 Meaning of the Term

Definition According to Spiegel, "Standard Deviation is the square root of the mean of the squares of the deviations of all values of a series from their Arithmetic Mean"

6.9.2 Standard Deviation in Individual Series

If x_1, x_2, x_3, ...,x_n are 'n' values, then the standard deviation '

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n}}$$

$$Standard \quad Deviation = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}; \quad \text{'n' is the no. of items.}$$

$$On \quad simplification \quad \sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$Variance \;=\; \sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$\text{Coefficient of variation} = \frac{SD}{Mean}$$

6.9.3 SD in Individual Series

Eg. Find the standard variation, variance and coefficient of variation for the following series.

5, 8, 7, 11, 9, 10, 8, 2, 4, 6.

x	x²
5	25
8	64
7	49
11	121
9	81
10	100
8	64
2	4
4	16
6	36
70	560

$$SD = \sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \quad SD = \sqrt{\frac{560}{10} - \left(\frac{70}{10}\right)^2} = \sqrt{56-49} = \sqrt{7} = 2.646$$

$$\text{Variance} = \sigma^2 = 7$$

$$\text{Coefficient of Variation} = \frac{2.646}{7} \times 100 = 37.8$$

6.9.4 SD in Discrete Frequency Series

$$SD = \sigma = \sqrt{\frac{\sum f \times x^2}{n} - \left(\frac{\sum f \times x}{n}\right)^2}$$

Example Find the standard deviation for the following series:

Mark	2	4	6	8	10
No. of Students	8	10	16	9	7

x	f	fx	x²	fx²
2	8	16	4	32
4	10	40	16	160
6	16	96	36	576
8	9	72	64	576
10	7	70	100	700
	50	294		2044

$$\sigma = \sqrt{\frac{\sum f \times x^2}{n} - \left(\frac{\sum f \times x}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{2044}{50} - \left(\frac{294}{50}\right)^2}$$

$$= \sqrt{40.88 - (5.88)^2}$$

$$= \sqrt{40.88 - 34.57} = \sqrt{6.31} = 2.51$$

SD in Continuous Series Here the only difference is that the mid value of the class is taken as the size(x).

$$\sigma = \sqrt{\frac{\sum f \times x^2}{n} - \left(\frac{\sum f \times x}{n}\right)^2}$$

Example 1

Find the standard deviation, variance and coefficient of variation.

Size	0-2	2-4	4-6	6-8	8-10	10-12
Frequency	2	4	6	4	2	6

Solution:

Size	f	Mid X	fx	x^2	fx^2
0-2	2	1	2	1	2
2-4	4	3	12	9	36
4-6	6	5	30	25	150
6-8	4	7	28	49	196
8-10	2	9	18	81	162
10-12	6	11	66	121	726
	24		156		1272

$$\sigma = \sqrt{\frac{\sum f \times x^2}{n} - \left(\frac{\sum f \times x}{n}\right)^2} \quad \sigma = \sqrt{\frac{1272}{24} - \left(\frac{156}{24}\right)^2}$$

$$= \sqrt{53 - (6.5)^2} = \sqrt{53 - 42.25} = \sqrt{10.75} = 3.28$$

$$\text{Variance} = \sigma^2 = 10.75$$

Mean $= \frac{\sum fx}{N} = \frac{156}{24} = 6.5$; Coefft of variation $= \frac{\sigma}{\bar{x}} \times 100 = \frac{3.28}{6.5} \times 100 = 50.46$

Example 2

The scores of two batsmen 'Sachin' and 'Dhoni' during a certain match are as follows.

Sachin	10	12	80	70	60	100	0	4
Dhoni	8	9	7	10	5	9	10	8

Examine which of the batsmen is more consistent in scoring. Who is more efficient batsmen?

Solution

The batsman who has the low coefficient of variation is more consistent. Also the batsman who has higher average score is more efficient.

$$\bar{x} = \frac{\sum x}{n} = \frac{336}{8} = 42$$

(More efficient)

$$\bar{x} = \frac{\sum x}{n} = \frac{66}{8} = 8.25$$

Sachin

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sigma = \sqrt{\frac{25160}{8} - \left(\frac{336}{8}\right)^2}$$

$$= \sqrt{3145 - (42)^2} = \sqrt{3145 - 1764} = \sqrt{1381} = 37.16$$

Coefft. of variation $= \frac{SD}{Mean} \times 100 = \frac{37.16}{42} \times 100 = 88.48$

Table 6.1: Sachin

x	x^2
10	100
12	144
80	6400
70	4900
60	3600
100	1000
0	0
4	16
336	25160

Table 6.2: Dhoni

x	x^2
8	64
9	81
7	49
10	100
5	25
9	81
10	100
8	64
66	564

Dhoni

$$\sigma = \sqrt{\frac{564}{8} - \left(\frac{66}{8}\right)^2} = \sqrt{70.5 - (8.25)^2} = \sqrt{2.44} = 1.56$$

Coefft. of Variation $= \dfrac{1.56}{8.25} \times 100 = 18.93$ (More Consistent)

Why SD is Best Measure?

Standard Deviation possesses most of the important characteristics which an ideal measure of dispersion should have.

- Standard Deviation is rigidly defined.
- It is based on all the observations of the data.
- It is amenable to more algebraic treatment.
- It possesses many mathematical properties.
- It is not much affected by sampling fluctuations.
- It does not ignore the signs of the deviations.
- It is possible to find out S. D. of two or more groups.
- Coefficient of Variation is based on Standard Deviation and it can be used to compare variability of two series.

- Standard Deviation is used to find out statistical measures like coefficient of Skewness, coefficient of Correlation, Regression equations etc.

So Standard Deviation is the best measure of dispersion.

6.9.5 Merits and Demerits of SD

Merits

- SD is based on all values of a series.
- It is rigidly defined.
- It is capable of further mathematical treatment.
- It is not much affected by sampling fluctuations.

Demerits

- It is difficult to calculate.
- Signs of deviations are not ignored.
- It gives weight for extreme values.

6.10 Skewness and Kurtosis

6.10.1 Symmetric Distributions

A frequency distribution is said to be symmetric if the frequencies are distributed symmetrically or evenly on either side of an average.

In a symmetrical frequency distribution, the number of items above the mean and below the mean would be the same and the items are symmetrically arranged about the mean.

Further, for symmetric distribution, Q_3 and Q_1 are equidistant from median.

Skewness

Skewness means lack of symmetry. The word skewness literally denotes asymmetry.

If a frequency distribution is skewed, there will be more items on one side of the mode than the other side

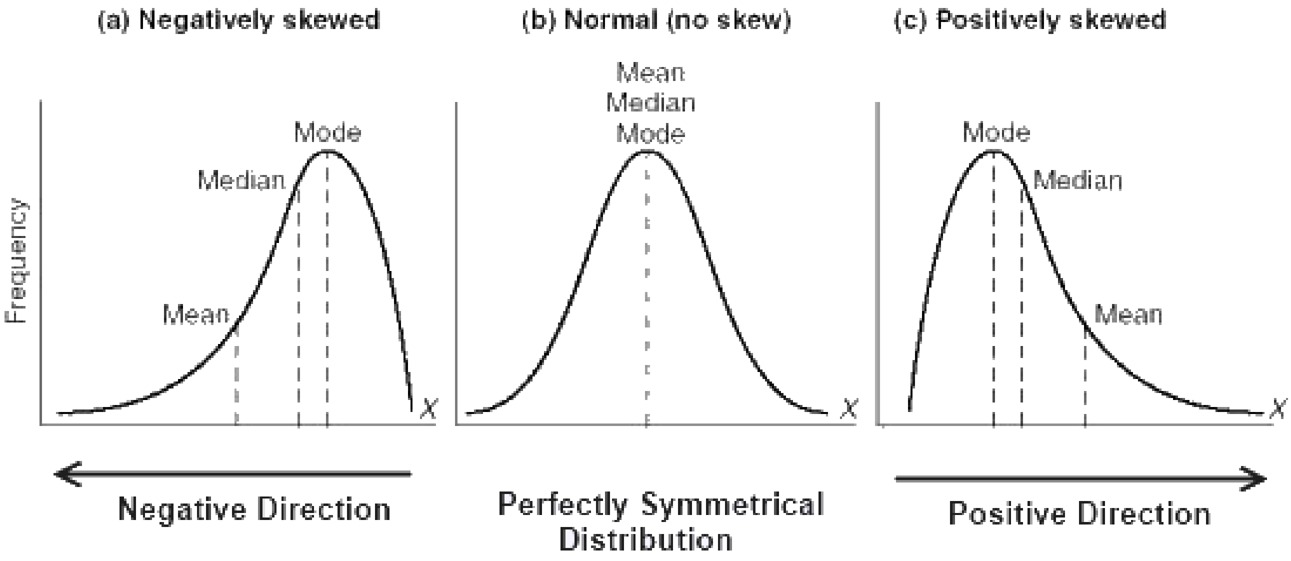
In the case of a skewed distribution, the mean and the median are pulled away from the mode. That is, for a skewed distribution mean, median and mode are not equal.

- For a skewed distribution Q_1 and Q_3 will not be equidistant from median.
- It has a long tail on one side and a short tail on the other side.
- Most of the economic data have skewed distributions.

Example : Income, Savings, etc. have skewed distributions.

+ve and -ve Skewness

- Skewness may be either positive or negative.
- Skewness is said to be positive when mean is greater than median and median is greater than mode. In this case the curve is skewed to the right.
- Here more than half the area falls at the right side of the highest ordinate.
- Skewness is said to be negative when mean is less than median and median is less than mode (the curve is skewed to the left).
- Here more than half the area falls at the left of the highest ordinate.
- For a positively skewed longer tail at the right and for a negatively skewed curve, tail at the left.



6.10.2 Measures of Skewness

Ist Measure of Skewness

Karl Pearson's Coefficient 'J' = $\dfrac{Mean - Mode}{Standard\ Deviation}$

In case, if in a frequency distribution, the Mode is ill defined, then Mean-Mode is taken as 3(Mean-Median)

If Mode Ill Defined 'J' = $\dfrac{3(Mean - Median)}{StandardDeviation}$; value of 'J' will be -3 < J > 3

2nd Measure of Skewness(Bowley's Coefft.)

Bowley's coefft. 'J' = $\dfrac{(Q_3 + Q_1 - 2M)}{Q_3 - Q_1}$

Where 'M' is the Median.

Third Measure of Skewness(Kelly's Coefft.)

'J' = $\dfrac{(D_9 + D_1 - 2\ Median)}{(D_9 - D_1)}$ or $\dfrac{(P_{90} + P_{10} - 2\ Median)}{(P_{90} - P_{10})}$

Where 'D' stands for decile and 'P'stands for Percentile.

Fourth Measure of Skewness

On the basis of central moments, Coefficient of skewness is given by 'J' = $\dfrac{\mu_3}{\sqrt{\mu_2^2}}$

Where'μ_3 and μ_2 are third moment and second moment respectively.

Central Moment μ_r

Central Moment 'μ_r' = $\dfrac{\sum(x - \bar{x})^r}{n}$ for individual series and $\dfrac{\sum f \times (x - \bar{x})^r}{N}$ for a frequency distribution.

6.10.3 Measures of Kurtosis

Kurtosis The term 'Kurtosis' indicates whether a distribution is flat topped or peaked.

Measure of Kurtosis is therefore measure of peakedness.

Mesokurtic When a curve is neither peaked nor flat topped, it is called mesokurtic(normal).

Leptokurtic When a frequency curve is **more peaked** than the normal curve it is called **leptokurtic**.

Platykurtic When a frequency curve is **more flat topped** than the normal curve, it is called **platykurtic**.

Measure of Kurtosis

Measure of Kurtosis is derived from moments.

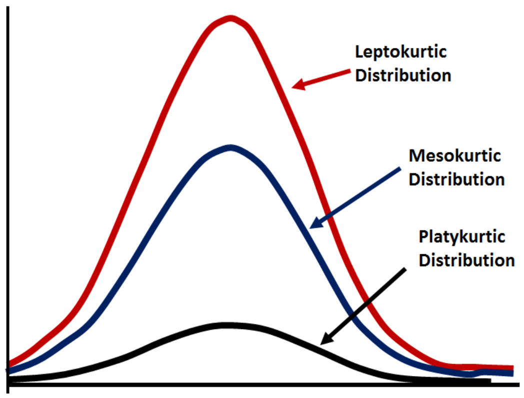


Table 6.3: The first four moments are as follows

Individual Series	Frequency Distribution
$`\mu_1' = \frac{\sum(x-\bar{x})^1}{n}$	$`\mu_1' = \frac{\sum f \times (x-\bar{x})^1}{N}$
$`\mu_2' = \frac{\sum(x-\bar{x})^2}{n}$	$`\mu_2' = \frac{\sum f \times (x-\bar{x})^2}{N}$
$`\mu_3' = \frac{\sum(x-\bar{x})^3}{n}$	$`\mu_3' = \frac{\sum f \times (x-\bar{x})^3}{N}$
$`\mu_4' = \frac{\sum(x-\bar{x})^4}{n}$	$`\mu_4' = \frac{\sum f \times (x-\bar{x})^4}{N}$

Measure of Kurtosis $`\beta_2' = \frac{\mu_4}{(\mu_2)^2}$

When $\beta_2 = 3$, the distribution will be mesokurtic, if it is lower than 3, the distribution is platykurtic and if it is greater than 3, it is leptokurtic.

Central Moments

Example : Find the coefficient of Skewness and Measure of Kurtosis for the following frequency distribution.

Class	0-2	2-4	4-6	6-8	8-10
Frequency	2	3	3	1	1

Solution

Class	x	f	fx	$x-\bar{x}$	f$(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
0-2	1	2	2	-3.2	-6.4	20.48	-65.54	209.73
2-4	3	3	9	-1.2	-3.6	4.32	-5.18	6.22
4-6	5	3	15	0.8	2.4	1.92	1.54	1.23
6-8	7	1	7	2.8	2.8	7.84	21.95	61.46
8-10	9	1	9	4.8	4.8	23.04	110.59	530.83
		10	42	4	0	57.60	63.36	809.47

$$\bar{x} = \frac{f \times \sum x}{N} = \frac{42}{10} = 4.2$$

$$`\mu_1' = \frac{\sum f \times (x-\bar{x})^1}{N} = \frac{0}{N} = 0$$
$$`\mu_2' = \frac{\sum f \times (x-\bar{x})^2}{N} = \frac{57.60}{10} = 5.76$$
$$`\mu_3' = \frac{\sum f \times (x-\bar{x})^3}{N} = \frac{63.36}{10} = 6.34$$
$$`\mu_4' = \frac{\sum f \times (x-\bar{x})^4}{N} = \frac{809.47}{10} = 80.95$$

Coefficient of skewness 'J' $= \frac{\mu_3}{\sqrt{(\mu_2)^3}} = \frac{6.36}{\sqrt{(5.76)^3}} = 0.46$

$\therefore$ the distribution is positively skewed.

Measure of Kurtosis = 'β_2' $= \frac{\mu_4}{(\mu_2)^2} = \frac{80.95}{(5.76)^2} = 2.44$

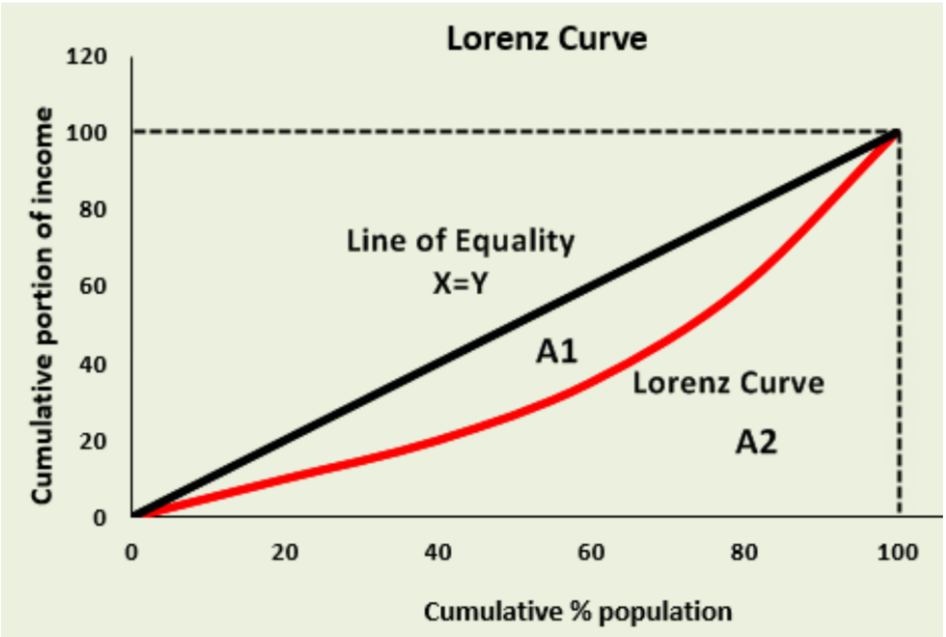
$\therefore$ the distribution is platykurtic.

6.10.4 Lorenz Curve

Lorenz Curve

- Lorenz curve is a graphical method of studying dispersion in a series.
- It is used in Business to Study the disparities of the distribution of wages, turn over, production, population, etc.
- In Economics it is useful to measure inequalities in the distribution of income between different countries, or between different periods of time.
- Lorenz curve is a graph drawn to a frequency distribution taking the cumulated percentage values of frequencies along the X-axis and cumulated percentage values of the variable along the Y- axis.

A typical Lorenz curve is given by



Measurement of Variability

- If there is no equality in the distribution Lorenz curve will coincide with the line of equal distribution.
- The more the Lorenz curve is away from the line of equal distribution the greater is the inequality or variability.

Lorenz curve is useful to:

- Study the variability in distribution.
- Compare the variability relating to phenomenon for two regions.
- study the changes in variability over a period.

6.11 Index numbers

6.11.1 Introduction

Meaning Definition by Croxton and Cowden Index numbers are devices for measuring differences in the magnitude of a group of related variables.

Definition by 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.

Definition by Wessel Willet and Simone An index number is a special type of average that provide in measurement of relative changes from time to time or from place to place.

Example An index number, therefore, is a statistical device for measuring changes in the magnitude of a group of related variables during a specific period in comparison to their level on some other period. i.e. index numbers measure changes in group of related variables over time.

Characteristics of Index Numbers

- Index numbers are specialised averages.i.e. we consider different types of items, units, price quotations, etc.
- Index numbers are expressed in percentages of relative changes; but the sign '%' is never used.
- Index numbers measure changes not capable of direct measurement. e.g. 'price level', 'Economic activity', 'Cost of living', etc. are not capable of direct measurement.
- Index numbers are meant for comparison. Comparison is made over different intervals of time with reference to a particular base year.
- Index numbers have universal application. They are used to to ascertain changes in different sectors of economy.

Important Uses of Index Numbers

- Index numbers measure changes in price level.
- Index numbers are indicators of inflationary or deflationary tendencies. Index numbers are useful to arrive at the real value of money or purchasing power of money.
- Index numbers can be used to make measuring adjustments in wages of employees. When there is increase in cost of living index number the wages of workers are also to be increased accordingly.
- Many of the economic and business policies are guided by index numbers. eg. Index numbers of Industrial Production shows relative changes in physical outputs.
- Index numbers can be used to study trends and tendencies of various phenomena.

Index Numbers are Economic Barometers

- Barometer is an instrument which is used for recording the changes in atmospheric pressure.
- Wholesale price index numbers record the changes in the price level of a country.
- Index numbers of agricultural production and industrial production reveal the changes in the progress in the fields of agriculture and industry.

Index Number is Economic Barometer

- Index numbers of business activity throw light on the economic progress that various countries have made.

Since index numbers make a complete study of the economic changes in the country, they are called economic barometers.

Limitations of Index Numbers

- Index numbers are based on samples so they do not take into account each and every item. Hence they are not perfect.
- Index numbers are only approximate indicators of the relative level of a phenomenon.
- All index numbers are not good for all purposes. An index number calculated for one purpose cannot be used in other places where they may not be fully appropriate.
- Index numbers are specialised type of average that they are subject to all limitations also which an average suffers.
- There are different methods for calculations of index numbers. So if a suitable method is not selected, the result obtained may not be accurate.
- Index numbers are liable to be misused by choosing an abnormal base year or irrational weights.

Problems in Constructing Index Numbers

- Purpose of index :- If not properly defined, would lead to confusions.
- Selection of the base period :- If not properly chosen, IN will be useless.
- Selection of items :- Difficult to include all items.
- Obtaining price quotation :- Should collect from a number of markets.
-
- Selection of an average :- Though Geometric Mean is the best average, arithmetic mean is usually used since it is simple.
- Selection of appropriate weight. If weights are wrongly chosen, IN will be inappropriate.
- Selection of an appropriate method and formula.

6.11.2 Simple and Weighted Index Numbers

Simple Index Number

Simple Index numbers are those in the calculation of which all the items are treated equally important. No item has importance than other.

Weighted Index Number

Weighted index numbers are those index numbers in the calculation of which each item assigned a particular weight. The importance of each item differs.

Meaning of the Term Weight

- Weight refers to the relative importance of the item.
- The system of weighing may be arbitrary or rational.
- In arbitrary way, the statistician is free to assign weights.
- In rational weighing some criteria have been fixed for assigning weights.
- The weights may be on the basis of the value or quantity purchased, consumed or sold.
- When quantity forms the basis, it is called ‘Quantity weighing’ and in case of value, it is called ‘Value weighing’.

Price Index Number

Price index numbers measure changes in the price of a commodity for a given period.

Example : Wholesale Price Index Number, cost of living index number, etc.

6.11.3 Methods for Construction of Price Index number

- Unweighted Index Numbers
- Weighted Index Numbers.

Unweighted or Simple Index Numbers

Simple index numbers are those index numbers in which all items are treated equally. There are two methods available:

- Simple Aggregate Method
- Simple Average Price Relative Method.

Simple Aggregate Method

Index Number (P_{01}) $= \frac{\sum P_1}{\sum P_0} \times 100$

where P_1 and P_2 are price for the current year and price for the base year.

Example :

Construct index number for 2015 on the basis of price of 2010.

Commodities	Price of 2010 (P_0)	Price of 2015 (P_1)
A	115	130
B	72	89
C	54	75
D	60	72
E	80	105
Sum	$\sum P_0 = 381$	$\sum P_1 = 471$

$$\text{Index Number } (P_{01}) = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{471}{381} \times 100 = 123.62$$

2. Simple Average Price Relative Method

Here Index number corresponding to each item is found out first and then average is taken.

$$\text{Price Index Index Number} = \frac{\sum I}{N}$$

Where 'I' $= \frac{P_1}{P_0} \times 100$ for each item and 'N' is the total number of items.

Example :

Calculate simple index number by average relative method.

Items	Price in Base Year	Price in Current Year
A	5	7
B	10	12
C	15	25
D	20	18
E	8	9

Solution

Items	Price in Base Year	Price in Current Year	$I = \frac{P_1}{P_0} \times 100$
A	5	7	140
B	10	12	120
C	15	25	166.7
D	20	18	90
E	8	9	112.5
N=5			$\sum I = 629.2$

Index Number $= \frac{\sum I}{N} = \frac{629.2}{5} = 125.84$

6.11.4 Weighted Index Number

In this method quantity consumed is also taken into account.
There two such index numbers

- Weighted Aggregate Method.
- Weighted Average of Price Relative Method.

I. Weighted Aggregate Method

This method is based on the weight of the prices of the selected commodities. Commonly used methods are:

1. Laspeyre's Method
2. Paasche's Method
3. Bowley-Dorbish Method
4. Fischer's Ideal Method
5. Kelly's Method
6. Marshal-Edgeworth Method

1. Laspeyre's Method

Here the index number is given by

$$\text{Index Number } (P_{01}) = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

where P_1, P_0 and q_0 are Price of current year, Price of base year and quantity of base year respectively.

2. Paasche's Method

Here the index number is given by

$$\text{Index Number } (P_{01}) = \frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$$

where P_1, P_0 and q_1 are Price of current year, Price of base year and quantity of current year respectively.

3.Fischer's Ideal Method

Here the index number is given by

$$\text{Index Number } (\mathrm{P}_{01}) = \sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1}} \times 100$$

where P_1, P_0 and q_1, q_0 are Price of current year, Price of base year and quantity of current year, quantity of base year respectively.

i.e. index number $= \sqrt{L \times P}$

where 'L' and 'P' are Laspeyre's and Paasche's index numbers respectively.

4. Bowley-Dorbish Method

Here the index number is given by average of Laspeyre's and Paasche's index numbers. i.e. Index Number $= \dfrac{L+P}{2}$

5. Kelly's Method

Index Number, $(\mathrm{P}_{01}) = \dfrac{\sum P_1 q}{\sum P_0 q} \times 100$; and 'q' $= \dfrac{q_1 + q_0}{2}$

6. Marshall-Edgeworth Method

Index Number $= \dfrac{\sum P_1 q_1 + \sum P_1 q_0}{\sum P_0 q_1 + \sum P_0 q_0} \times 100$

Problem

Calculate Laspeyre's Index Number, Paasche's Index Number and Fischer's Index Number for the following data related with the prices and quantities consumed for 2010 and 2015.

Commodity	2010 Price(P_0)	2010 Quantity(q_0)	2015 Price (P_1)	2015 Quantity(q_1)
Rice	5	15	7	12
Wheat	4	5	6	4
Sugar	7	4	9	3
Tea	52	2	55	2

Solution

Commodity	P_0	q_0	P_1	q_1	$\mathrm{P}_1 q_0$	$\mathrm{P}_0 q_0$	$\mathrm{P}_1 q_1$	$\mathrm{P}_0 q_1$
Rice	5	15	7	12	105	75	84	60
Wheat	4	5	6	4	30	20	24	16
Sugar	7	4	9	3	36	28	27	21
Tea	52	2	55	2	110	104	110	104
sum $\sum$					281	227	245	201

$$\text{Laspeyre's Index Number } (\mathrm{P}_{01}) = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

$$= \frac{281}{227} \times 100 = 123.79$$

Paasche's Method

$$\text{Paasche's Index Number } (\mathrm{P}_{01}) = \frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$$

$$= \frac{245}{201} \times 100 = 121.89$$

Fischers Method

$$\text{Fischer's index number } (\mathrm{P}_{01}) = \sqrt{L \times P}$$

$$= \sqrt{123.79 \times 121.89} = 122.84.$$

6.11.5 II. Weighted Average Price Relative Method

Here Index Number is given by

$$\text{Index Number} = \frac{\sum I \times V}{\sum V}$$

where 'I' $= \frac{P_1}{P_0} \times 100$ which is called 'index'.
'V' is the weight.

Sample Problem

Calculate the index number of Price for 2018 on the basis of 2014 from the data given below:

Commodity	Weight	Price of 2014	Price of 2018
A	40	16	20
B	25	40	60
C	5	2	2
D	20	5	6
E	10	2	1

Solution

Commodity	Weight	P_0	P_1	I $= \frac{P_1}{P_0} \times 100$	IV
A	40	16	20	125	5000
B	25	40	60	150	3750
C	5	2	2	100	500
D	20	5	6	120	2400
E	10	2	1	50	500
	$\sum$V=100				$\sum$IV=12150

$$\text{Index Number} = \frac{\sum I \times V}{\sum V} = \frac{12150}{100} = 121.5$$

6.11.6 Quantity Index Number

In Price index number weight is given to quantity where as in Quantity index number weight is given to Price.

Laspeyre's Quantity Index Number Laspeyre's Quantity Index Number(q_{01}) $= \frac{\sum q_1 P_0}{\sum q_0 P_0} \times 100$

Paasche's Quantity Index Number

$$\text{Index Number } (q_{01}) = \frac{\sum q_1 P_1}{\sum q_0 P_1} \times 100$$

Problem

From the following data compute quantity index number according to Laspeyre's method and Paasche's method.

	2010	2010	2015	2015
Commodity	Price	Total Value	Price	Total Value
Rice	8	80	10	110
Sugar	10	90	12	108
Pulses	16	256	20	340

Solution

Commodity	P_0	q_0	P_1	q_1	q_1P_0	q_0P_0	q_1P_1	q_0P_1
Rice	8	10	10	11	88	80	110	100
Sugar	10	9	12	9	90	90	108	108
Pulses	16	16	20	17	272	256	340	320
Total					450	426	558	528

Laspeyre's Quantity index number $q_{01} = \frac{\sum q_1 P_0}{\sum q_0 P_0} \times 100$

$= \frac{450}{426} \times 100 = 105.63$

Commodity	P_0	q_0	P_1	q_1	q_1P_0	q_0P_0	q_1P_1	q_0P_1
Rice	8	10	10	11	88	80	110	100
Sugar	10	9	12	9	90	90	108	108
Pulses	16	16	20	17	272	256	340	320
Total					450	426	558	528

Paasche's Quantity index number $q_{01} = \frac{\sum q_1 P_1}{\sum q_0 P_1} \times 100$

$= \frac{558}{528} \times 100 = 105.68$

6.11.7 Types of Index Numbers

Generally index numbers of are of three types

1. Price Index Number

They indicate the general levels of prices of articles, commodities, services, in the current period as compared to a base period.

Example : Whole sale price index number, Consumer Price index number, etc.

2. Quantity Index Numbers

Quantity index numbers indicate general level of quantities of goods produced or consumed in the current period compared to a base period.

Example :

- Index numbers of quantity of goods imported
- Index numbers of quantity of goods exported
- Index numbers of quantity of agricultural production. etc.

3. Value Index Numbers

They measure total money value of transaction taking place.

6.11.8 Tests for Index Numbers

A good index number should satisfy the following tests.

1. Time Reversal Tests
2. Factor Reversal Tests and
3. Unit Test

1. Time Reversal Test (Irving Fischer)

An index number should be such that, when the base year and current year are interchanged(reversed), the resulting index number should be the reciprocal of the earlier.

Let P_{01} be the index number for the period '1' with respect to the base period '0' and let P_{10} is the index number for the period '0' with respect to period '1',then the particular index number satisfies time reversal test if:

$$P_{01} \times P_{10} = 1$$

P_{01} and P_{10} are mere ratios & should not be represented as percentages.

Time reversal test is not satisfied by Laspeyre's and Paasche's Index number. But Fischer index number and Marshal and Edgeworth index number satisfy this condition.

Fischer's index number is given by

$$\text{Index Number } (P_{01}) = \sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1}} \times 100$$

According to time reversal test

$$P_{01} \times P_{10} = 1$$

According to time reversal test:

$$P_{01} \times P_{10} = 1$$

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum P_0q_1}{\sum P_1q_1} \times \frac{\sum P_0q_0}{\sum P_1q_0}} = 1$$

It can also be verified experimentally.

Factor Reversal Test(Proposed by Irving Fischer)

Here the argument is that the index number should be such that the price index number and quantity index number calculated according to the formula should be quality effective in indicating changes.

Factor reversal test require that the product of the price index number and quantity index number should indicate the net change in value taking place in between the two periods.

According to factor reversal test :

$$P_{01} \times q_{01} = \frac{\sum P_1q_1}{\sum P_0q_0}; \quad \text{RHS part is the value index ratio.}$$

Only Fischer's Index number satisfy this condition.

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum q_1P_0}{\sum q_0P_0} \times \frac{\sum q_1P_1}{\sum q_0P_1}}$$

$$\text{i.e} = \frac{\sum P_1q_1}{\sum P_0q_0}$$

Unit Test

This test requires that the index number formula should be independent of units in which prizes or the quantities of various commodities are quoted.

All index numbers except simple aggregate of prizes and simple aggregate of quantities satisfy this test.

Example

From the following data, find Fisher's index number and show that the time and factor reversal tests are satisfied by it.

Commodity or Items	Base Year Price	Base Year Expenditure	Current Year Price	current Year Expenditure
A	8	80	10	120
B	10	120	12	96
C	5	40	5	50
D	4	56	3	60
E	20	100	25	150

Solution

$$\text{Fischer's Index Number } (\mathrm{P}_{01}) = \sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1}} \times 100$$

Time reversal test:

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum P_0q_1}{\sum P_1q_1} \times \frac{\sum P_0q_0}{\sum P_1q_0}} = 1$$

Factor Reversal Test:

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum q_1P_0}{\sum q_0P_0} \times \frac{\sum q_1P_1}{\sum q_0P_1}} = \frac{\sum P_1q_1}{\sum P_0q_0}$$

Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
A	8	10	10	12	80	96	100	120
B	10	12	12	8	120	80	144	96
C	5	8	5	10	40	50	40	50
D	4	14	3	20	56	80	27	60
E	20	5	25	6	100	120	125	150
$\sum$					396	926	436	476

$$\text{Fischer's Index Number } (\mathrm{P}_{01}) = \sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1}} \times 100$$

$$= \sqrt{\frac{436}{396} \times \frac{476}{926}} \times 100 = 75.23$$

Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
-	-	-	-	-	-	-	-	-
$\sum$					396	926	436	476

Time reversal test:

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum P_0q_1}{\sum P_1q_1} \times \frac{\sum P_0q_0}{\sum P_1q_0}} = 1$$

$$= \sqrt{\frac{436}{396} \times \frac{476}{926} \times \frac{926}{476} \times \frac{396}{436}} = 1$$

Commodity	P_0	q_0	P_1	q_1	P_0q_0	P_0q_1	P_1q_0	P_1q_1
-	-	-	-	-	-	-	-	-
$\sum$					396	926	436	476

Factor Reversal Test:

$$\sqrt{\frac{\sum P_1q_0}{\sum P_0q_0} \times \frac{\sum P_1q_1}{\sum P_0q_1} \times \frac{\sum q_1P_0}{\sum q_0P_0} \times \frac{\sum q_1P_1}{\sum q_0P_1}} = \frac{\sum P_1q_1}{\sum P_0q_0}$$

$$= \sqrt{\frac{436}{396} \times \frac{476}{926} \times \frac{926}{396} \times \frac{476}{436}} = \frac{476}{396} = \frac{\sum P_1q_1}{\sum P_0q_0}$$

6.11.9 Consumer Price Index Number

'Consumer Price Index Number' is also known as 'Cost of Living Index Number' or 'Retail Price Index Number'.

Consumer Price Index Numbers are generally used to represent the average change over time in the pries paid by the ultimate consumer for a specified basket of goods and services.

It represents the average change in prices over a period of time, paid by the consumer for goods and services.

Steps for Construction of Consumer Price Index

- Determination of class of people for whom the index number is to be constructed.
- Selection of base period.
- Conducting a family budget enquiry.
- Obtaining price quotations.
- Selecting Proper weights.
- Selection of suitable methods for constructing index.

Usually Aggregate Expenditure Method (By Laspeyre formula) and Family Budget Method (Average Relative Method) is used to construct the cost of living index number.

1. Aggregative Expenditure Method : $\frac{\sum P_1q_0}{\sum P_0q_0}\times 100$
2. Family Budget Method : $\frac{\sum IV}{V}$

Example : An enquiry into the budgets of middle class families in Thrissur City gave the following information. What changes in the cost of living of 2016 as compared to 2012 ?

Expenses	Food	Rent	Clothing	Fuel	Misc.
≈ Proportion	35%	15%	20%	10%	20%
Price(2012)	150	30	75	25	40
Price(2016)	145	30	65	23	45

Solution

Expenses	Weight	P_0	P_1	$I = \frac{P_1}{P_0}\times 100$	IV
Food	35	150	145	96.67	3383.45
Rent	15	30	30	100	1500.00
Clothing	20	75	65	86.67	1733.40
Fuel	10	25	23	92	920.00
Misc.	20	40	45	112.50	2250.00
$\sum$	100				9786.85

Cost of Living Index Number : $\frac{\sum IV}{\sum V} = \frac{9786.85}{100} = 97.87$ (-2.13%)

Utility of Consumer Price Index Number

- Important use is for wage negotiations and and wage contracts.
- For Govts, they are used for wage policy, Price control, Rent control, Taxation, etc.
- Used for measuring purchasing power of money.
- Used for analysing markets for particular kind of goods and services.
- They are used for estimating the real wages of the workers from the wages paid to them.

6.11.10 Purchasing Power of Money

Purchase of power of money is also called 'Value of Rupee'. As the prize of commodities increases the value of Rupee decreases.

Value of Rupee for a period is the reciprocal of Price Index of that period multiplied by 100.

$$\text{Current Purchasing Power of Money} = \frac{1}{CurrentPriceIndex} \times 100$$

6.12 Analysis of Time Series

6.12.1 Time Series

Time series is the arrangement of data according to their occurrence.

It helps to find out the variations to the value of data due to changes in time.

Importance

- It helps for understanding past behaviour.
- It facilitates forecasting and planning.
- It facilitates comparison.
- It helps in evaluating current programmes.

6.12.2 Components of Time Series

Statistical Series are usually affected by multiplicity of causes like tastes and habits of people, changes in population, changes in the cost of production, changes in the income of people, etc.

The effects of these factors on time series are called components of time series. They are:

1. Secular Trend
2. Seasonal Variation
3. Cyclic Variation.
4. Irregular Fluctuations.

1. Secular Trend

Secular Trend may be defined as the changes over a long period of time. The significance is greater when the period of time is larger.

It is the general tendency of a statistical data

Example : Population increases, Illiteracy decreases.

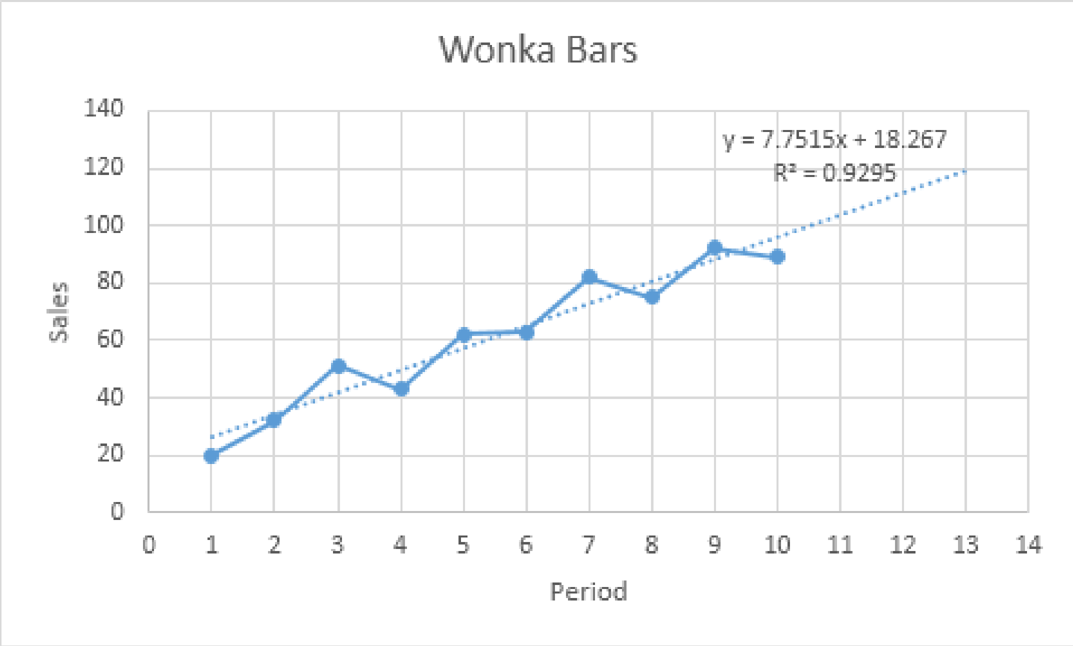
2. Seasonal Variations

Seasonal variations are measured for one calendar year. It is the variations which occur with some degree of regularity.

Example : Climate conditions, Social customs, etc.

3. Cyclic Variations

Cyclic variations are those variations which offer on account of business cycle for greater than one year. These cyclic moments pass through different stages of prosperity, recession, depression and recovery.



h!

6.12.3 4. Irregular Fluctuations

Irregular fluctuations are caused by unusual, unexpected, and accidental causes like earthquake, flood, strike, etc.

6.12.4 Methods for Measuring Trend

1. Graphical Method(Free Hand Method)

- This is the simplest method for measuring Trend.
- Under the method original data are plotted on the graph.
- The plotted points should be joined to get a curve.
- A straight line should be drawn through the middle area of the curve.
- Such line will describe the tendency of the data.

Semi Average Method

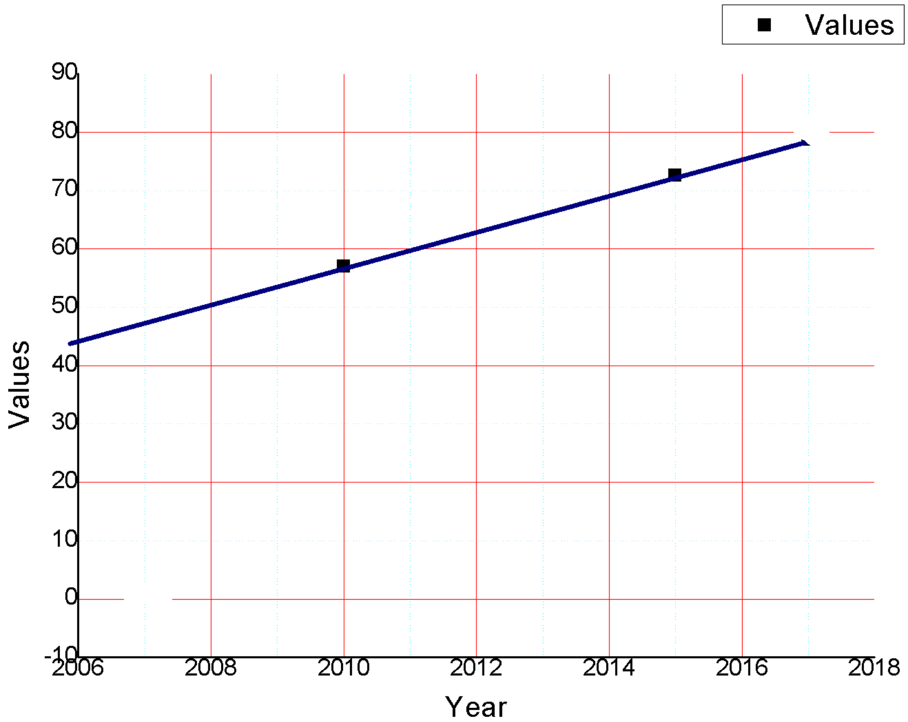
- The whole data are divided into two parts in the chronological order.
- If the number of years are odd, the middle year is omitted.
- Then average of these two groups are calculated.
- The two averages are then plotted in the graph against the middle year of each group.
- The two points are then joined together so as to get a straight line.
- This line is called **ward live**.
- Trend values are then calculated from this trend line corresponding to each year.

Problem

Determine trend by applying method of semi average.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Value	45	58	62	50	70	72	68	70	78	75

Year	Values	Semi Average
2008	45	
2009	58	
2010	62	$\frac{285}{5} = 57$
2011	50	
2012	70	
2013	72	
2014	68	
2015	70	$\frac{363}{5} = 72.6$
2016	78	
2017	75	



Then trend values are read from curve

Year	Values
2008	50.5
2009	53.2
2010	57
2011	60
2012	63
2013	65.8
2014	70
2015	72.6
2016	75
2017	78.5

3. Method of Moving Average

- Moving average method is an improvement over free hand curve and semi average method.
- is quite simple and is used for smoothing the fluctuations in curves.
- The trend values obtained by this method are more accurate.
- Under this method, a series of successive average should be calculated from a series of values.

- Moving average is calculated for 3, 4, 5, 6, or 7 year periods.
- The period of moving average is so chosen that it is nether be too long or too short so that trend values are not distorted or irregular fluctuations are not significant.

The moving average can be calculated as follows:

3 Year Moving Average

3 year period moving average is calculated as $\frac{a+b+c}{3}$, $\frac{b+c+d}{3}$, $\frac{c+d+e}{3}$, $\frac{d+e+f}{3}$,.. and so on

5 Year Moving Average

5 Year moving average is calculated as $\frac{a+b+c+d+e}{5}$, $\frac{b+c+d+e+f}{5}$, $\frac{c+d+e+f+g}{5}$, and so on.

Example:

Compute a 3 year moving average for the following data:

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Sales	55	47	59	151	79	36	45	72	83	89	102

Solution

Year	sales	3 yearly moving total	3 yearly moving average
2002	55	–	–
2003	47	161	53.67
2004	59	257	85.67
2005	151	289	96.33
2006	79	266	88.67
2007	36	160	53.33
2008	45	153	51
2009	72	200	66.67
2010	83	244	81.33
2011	89	274	91.33
2012	102	–	–

4. Method of Least Squares

- This is the popular method of drawing trend line.
- The trend line obtained by this method is called line of best fit.
- The trend line is represented as ‘ y = a+bx’
- The value of ‘a’ and ‘b’ are determined by solving the following two normal equations.
- To find ‘a’ and ‘b’, we apply the method of least squares.
- Let ‘E’ be the sum of the squares of the deviations of all original values from their respective values derived from the equations,so that ‘E = [y-(a+bx)]^2’.
- As per calculus method, for minimum ‘$\frac{\partial E}{\partial a}$’ = 0 and ‘$\frac{\partial E}{\partial b}$’ = 0.
- On solving we get two equations: $\sum y = Na + b\sum x$ and $\sum xy = a\sum x + b\sum x^2$
- Where ‘x’ and ‘y’ represents the time and the value of the variable, ‘a’ and ‘b’ are constants an ‘N’ represents the total number.

- Solving these equations we get ‘a’ and ‘b’.
- Substituting these values in the equation ‘y = a + bx’ will give us the trend line.

When middle year is taken as the origin, then $\sum x = 0$, then normal equation would be

$$\sum y = Na + b\sum x \text{ ; } \sum y = \text{Na ; } \therefore\ a = \frac{\sum y}{N}$$

$$\sum xy = a\sum x + b\sum x^2 \text{ ; } \sum xy = \text{b}\sum x^2 \therefore \text{b} = \frac{\sum xy}{\sum x^2}$$

Then graph may be drawn to represent trend line.

Example

Calculate the trend values through the method of least squares and also forecast the production in 2013 and 2015.

Year	2006	2007	2008	2009	2010	2011	2012
Production	47	64	77	88	97	109	113

Solution:

Year ‘t’	Production ‘y’	x=(t-2009)	xy	x^2
2006	47	-3	-141	9
2007	64	-2	-128	4
2008	77	-1	-77	1
2009	88	0	0	0
2010	97	1	97	1
2011	109	2	218	4
2012	113	3	339	9
$\sum$	595		308	28

Calculation of ‘a’

$$\text{a} = \frac{\sum y}{N} = \frac{595}{7} = 85$$

Calculation of ‘b’

$$\text{b} = \frac{\sum xy}{\sum x^2} = \frac{308}{28} = 11.$$

The equation for straight line will be ‘y = a + bx’
a= 85 ; b = 11 ; ∴ y = 85 + 11x.
Then for each year trend values (‘y’) can be calculated by putting the value of x for that year in the above equation for straight line.

Year	‘x’	‘y = a + bx’	Trend value, ‘y’
2006	-3	y = 85 + 11(-3)	52
2007	-2	y = 85 + 11(-2)	63
2008	-1	y = 85 + 11(-1)	74
2009	0	y = 85 + 11(0)	85
2010	1	y = 85 + 11(1)	96
2011	2	y = 85 + 11(2)	107
2012	3	y = 85 + 11(3)	118
2013(Expected)	4	y = 85 + 11(4)	129
2015(Expected)	6	y = 85 + 11(6)	151

Graph may drawn with above actual values and trend values for verification.

6.12.5 Seasonal Variations

- Seasonal variations are quite regular and uniform and can be predicted with some amount of accuracy.
- Our past experience is the best guide for such forecasts.

Example : Prices of rice will come down during harvesting season and there it will go up in the sowing season.

- One is interested in seasonal fluctuations in order to take the advantage of it.
- We try to purchase rice and stock it.
- At the peak of the season, the price is low and quality may be high.
- It helps in definite forecasting and adjustment of supply to possible demand of the people.

Factors affecting Seasonal Variation

- Seasonal variation occurs because of natural factors and man made conventions most of the business and economic activities vary in systematic pattern during the season of the year.
- The major factors responsible for seasonal variations are
  - ⋆ Climate and weather conditions.
  - ⋆ Customs, traditions and habits.
- The most important factor causing seasonal variation is the changes in the climate and weather conditions such as rainfall, temperature, etc.
- They act on different products differently. e.g. Sales of umbrella will be more in rainy season and also greater demand for cold drinks during summer.
- The nature is primarily responsible for seasonal variations.
- Custom tradition and habits have their in back
- For example on certain occasions like Deepavali, Christmas etc., there is big demand for sweets and gifts, during Onam, there will be higher sales of textile goods there is huge demand for gold during marriage season.

Measurement of Seasonal Variation

Methods for measuring Seasonal Variation.

1. Method of averages.
2. Ratio to trend method.
3. Ratio to moving average method.
4. Method of link relatives.

Sample Problem

Fit a straight line trend by the method of least squares and estimate the value of 2021.

Year	2012	2013	2014	2015	2016	2017	2018	2019
Values	380	400	650	720	690	600	870	930

* * * * * * *