(iv) $I = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$ is ..... weighted index number.

Q 5: Calculate quantity index by (i) Laspeyre's method (ii) Paasche's method (iii) Fisher's method.

Commodity	Price (p_0)	2004 Total value (p_0q_0)	Price (p_1)	2006 Total value (p_1q_1)
A	10	100	12	144
B	12	144	14	196
C	14	196	16	256
D	16	256	18	324
E	18	324	20	400

Q 6: From the following date calculate Price index numbers for 2007 as base by (1) Laspeyre's method (2) Paasche's method and (3) Fisher's method.

Commodity	2006		2007	
	Price	Quantity	Price	Quantity
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

11.8 TEST OF ADEQUACY OF INDEX NUMBERS

We have seen in 11.5 section that the relatives have certain important properties. What is true for an individual commodity should also true for a group of commodities. The index numbers as an aggregative relative should also satisfy the same set of properties. We shall examine the properties that a good index number should have.

1. Time Reversal Test or Property:

If the two periods, the base period and the reference period are interchanged, the product of the two index numbers should be unity. In other

words, one should be the reciprocal of the other.

Denote the index number of the given year (1) relative to the base year (0) by $I_{0/1}$. Then $I_{1/0}$ is the index number of the year (0) relative to the year (1), that is, with the years interchanged time reversal test requires that

$$I_{0/1} \times I_{1/0} = 1, \qquad or \qquad I_{1/0} = \frac{1}{I_{0/1}}$$

An index number which satisfies the above property is said to satisfy time reversal test.

Remark: As we have seen in section 11.5, all the relatives satisfy the time reversal property.

Let us now examine whether Laspeyre's index number satisfy this property or not. We know that Laspeyre's index number (without percentage) is $\frac{\sum p_1 q_0}{\sum p_0 q_0}$. Interchanging the time subscripts, we get $\frac{\sum p_0 q_1}{\sum p_1 q_1}$ and hence their product is $\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_0 q_1}{\sum p_1 q_1} \neq 1$.

Consequently, Laspeyre's price index number does not satisfy the time reversal property. Similarly, it can be seen that Laspeyre's quantity index number also does not satisfy the time reversal property. Also, Paasche's price and quantity index numbers do not satisfy this property.

Now consider the Fisher's price index number. Interchanging the time subscripts we find another index number. The product of the two is given by

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

which is equal to 1. Hence Fisher's price index number satisfies the time reversal property.

2. Factor Reversal Test or Property:

If the two factors *p* and *q* in a price index formula are interchanged, so that a quantity index number is obtained, then the product of the two index numbers should give the true value ratio

$$\frac{\sum p_1 q_1}{\sum p_0 q_0}$$

In other words, the price index number multiplied by the corresponding quantity index number should give the true ratio of value in the given year (1) to the value in the base year (0). This property holds for a single commodity, since

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

but it does not hold for most of the index numbers. For example,

$$P_L \times Q_L = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum q_1 p_0}{\sum q_0 p_0} \neq \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Now consider the Fisher's index number. We have from the formulas (1) and (2),

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

This proves that Fisher's index number also satisfy factor reversal test. Fisher's index number is one of the rare index numbers which satisfy both the time reversal and factor reversal tests. So it is called "Fisher's Ideal Index Number".

3. Circular Test:

This test is based on the shifting of the base period. If $I_{i/j}$ denotes the index number for the given period *j* with respect to the base period *i*, then this test requires that

$$I_{a/b} \times I_{b/c} \times I_{c/a} = 1$$

Remark: As we have seen in section 11.5, all the relatives satisfy the circular property.

Remark: Verify that none of the weighted index numbers satisfy circular test.



11.9 LET US SUM UP

- Index number is a statistical measure or device with a purpose of showing average change in one or more related variables over time and space.
- Price index numbers are more commonly used. It measures relative changes in prices over a time period.
- We can have either simple or weighted index number. Simple index is also called unweighted index number or index number with equal weights. In this unit, we have discussed about the simple index numbers as well as weighted index number.
- We have discussed four properties of relatives: Identity property, time reversal property, circular property and modified circular property.
- We have learnt about two methods of constructing simple index numbers: Simple aggregate method and Method of simple average of relatives. The formula to find index number using simple aggregate method is given by $P_{0/1} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$.

  Again, the formula to find price index number using the method of simple averages of relatives is given by $p_{0/1} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N}$.
- We have also leant about weighted index numbers. There are two groups of methods to calculate index number of this category: (a) Weighted aggregate method and (b) Weighted average of price relative methods. We briefly discussed about three types of weighted aggregate method. These methods are popularly known as (1) Laspeyre's Method, (2) Paasche's Method and (3) Fisher's Method.
- We have also learnt about the test of adequacy of index numbers.

We mentioned three properties that a good index number should have. Fisher's index number satisfies both the time reversal and factor reversal tests and hence it is known as Fisher's Ideal Index Number.



11.10 FURTHER READINGS

1) Agarwal, D. R. (2006). Business Statistics. Delhi, Vrinda Publications.
2) Gupta S.C. (1994). Fundamentals of statistics. New Delhi, Himalayan Publishing House.
3) Rajagopalan, S. P. & Sattanthan R. (2009).Business Statistics and Operations Research. New Delhi, Tata McGraw-Hill
4) Sharma, J. K. (2007). Business Statistics. New Delhi. Pearson Education Ltd.
5) Verma, A. P. (2007). Business Statistics. Guwahat, Asian Books Private Limited.



11.11 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: T, T, F, F, T.

Ans to Q No 2: (i) average (ii) specialized (iii) related (iv) 1990, 1998 (v) equal (vi) base year price (vii) 200%

Ans to Q No 3: T, T, F, F, F

Ans to Q No 4: (i) base year (ii) current year (iii) Laspeyre's (iv) Paache's.

Ans to Q No 5: Construction of various Indices

p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_1	p_1q_0
10	10	12	12	100	120	144	120
12	12	14	14	144	168	196	168
14	14	16	16	196	224	256	224

16 18	16 18	18 20	18 20	256 324	288 360	324 400	288 360
			Total	1020	1160	1320	1160

(i) Laspeyr's Price Index is

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

$$= \frac{1160}{1020} \times 100$$

$$= 113.7$$

(ii) Paasche's Price Index is

$$\frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

$$= \frac{1320}{1160} \times 100$$

$$= 113.8$$

(iii) Fisher's Price Index is

$$\sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

$$= \sqrt{113.7 \times 113.8}$$

$$= 113.75$$

Ans to Q No 6: Computation of Index Numbers

Comm-odity	2006		2007					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	20	8	40	6	320	160	240	120
B	50	10	60	5	600	500	300	250
C	40	15	50	15	750	600	750	600
D	20	20	20	25	400	400	500	500
					$\sum p_1q_0$ 2070	$\sum p_0q_0$ 1660	$\sum p_1q_1$ 1790	$\sum p_0q_1$ 1470

(i) Laspayer's Index

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

$$= \frac{2070}{1660} \times 100$$

$$= 124.70$$

(ii) Paasche's Index

$$\frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

$$= \frac{1790}{1470} \times 100$$

$$= 121.77$$

(iii) Fisher's Index

$$\sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

$$= \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= \sqrt{1.5184} \times 100$$

$$= 1.2322 \times 100$$

$$= 123.22$$



11.12 MODEL QUESTIONS

Q 1: Explain the meaning of an index number and its main characteristics.

Q 2: "Index numbers are economic barometers". Why are index numbers so called? Explain.

Q 3: What is price relative? Show with the help of an example how it is calculated?

Q 4: Define quantity and value relatives. What do they measure?

Q 5: State the important properties of relatives.

Q 6: Explain the method of construction of unweighted index numbers by Simple Aggregate Method.

Q 7: Explain the method of construction of unweighted index numbers by Simple Aggregate of Price Relative Method.

Q 8: The exchange rate of a US dollar was Rs.44.00 in July 2008 and was

Rs.50.00 in February 2009. Find the price relative.

Q 9: The production of rice in Assam in the month of January 2008 was 3608 tonnes and that in February 2009 was 3700 tonnes. Find the quantity relative.

Q 10: The following table shows the prices of the commodities consumed in 1970 and 1978.

Commodities	Unit	1970 Price (Rs)	1978 Price (Rs)
A	kilogram	4.00	4.60
B	litre	8.00	8.75
C	metre	7.00	8.00
D	kilogram	14.00	16.00
E	litre	3.00	4.00

Find the price index number using:

(a) simple aggregate method

(b) method of simple aggregate of price relatives.

Q 11: The production of rice, milk, cloth, wheat and petrol are given in the following table. Taking 1980 as base year, compute quantity index number using:

(a) simple aggregate method

(b) method of simple aggregate of quantity relatives.

Commodities	Unit	Quantity produced	
		1980	1990
Rice	kilogram	400000	700000
Milk	litre	800000	905000
Cloth	metre	100000	105000
Wheat	kilogram	140000	160000
Petrol	litre	300500	400000

Q 12: What do you mean by tests of index numbers?

Q 13: Explain the important criterion or test that a good index number should satisfy.

Q 14: The prices and quantity of five commodities are given in the following table. Taking 2000 as base year, compute value index number using:

(a) simple aggregate method

(b) method of simple aggregate of value relatives.

Comm-odities	2000		2002	
	Price (Rs)	Quantity (Kg)	Price (Rs)	Quantity (Kg)
A	400	10	700	12
B	800	15	1905	17
C	100	5	125	5
D	140	7	160	7
E	300	6	400	6

Q 15: Find Laspeyre's and Paasche's price and quantity index number from the following data. Also find the Fisher's price and quantity index numbers.

Commo-dities	Base year		Current year	
	Price (Rs) p_0	Quantity (kg) q_0	Price (Rs) p_1	Quantity (kg) q_1
A	5	25	6	30
B	10	5	15	4
C	3	40	2	50
D	6	30	8	35

Q 16: The following table shows the price quantity of the commodities consumed in 2000 and 2006.

Commo-dities	2000		2006	
	Price (Rs)	Quantity (kg)	Price (Rs)	Quantity (kg)
I	400	35	460	35
II	800	10	1000	10
III	300	15	50	12
IV	4000	10	5000	8
V	50	20	300	20

Find the following index numbers.

(1) Laspeyre's Price and Quantity index numbers.

(2) Paasche's Price and Quantity index numbers.

(3) Fisher's Price and Quantity index numbers.

*** ***** ***

UNIT 12 : TIME SERIES

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Definition of Time Series

12.4 Importance of Time Series Analysis

12.5 Components of a Time Series

12.6 Methods of Measuring Secular Trend

12.7 Estimation of the Trend by the Method of Moving Average

12.8 Let Us Sum Up

12.9 Further Readings

12.10 Answers To Check Your Progress

12.11 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will able to-

- understand the concept of time series
- explain the importance of time series analysis
- known about the various factors which affect a time series
- know about the methods for determining secular trend in time series.

12.2 INTRODUCTION

In Economics, Business and Commerce, it is important to make estimate for the future e.g. an economist is interested to know the figures of national income, prices and wages, population etc. for his future planning or a businesman likes to estimate his likely sales in the coming years to adjust his production accordingly. For making such estimates, one has to collect information from the past i.e. one has to deal with statistical data collected and recorded at successive intervals of time (or points of time). Such statistical data relating to time all refered to as Time Series. The analysis of time series has an important role in Economics, Business

and Commerce. Basically, most of the statistical technigues for the analysis of time series data have been developed by economists. However, these techniques can also be applied for study of behaviour of any phenomenon collected chronologically over a period of time in any discipline relating to natural and social sciences, though not directly related to business and economics.

12.3 DEFINITION OF TIME SERIES

A Time Series is a set of observations taken at specified times, usually (but not always) at equal intervals. Thus a set of data depending on time (which may be year, half-year, quarter, month, week, days etc.) is called a time series.

Some examples of time series are–

(i) The annual production of steel in India over the last ten year.

(ii) The sale of a departmental store in different months of the year.

(iii) The daily closing price of a share in the Bombay Stock exchange for a month.

(iv) Hourly Temperature recorded by the meteorological office in a city.

(v) Quarterly Consumer Price Index numbers of a country for last three years.

(vi) Daily rainfall of a particular city in the month of May, June, July and August.

Mathematically, a time series is defined by the functional relation ship – $Y = f(t)$

Where y is the value of the phenomenon (or variable) under consideration at time t. Thus, if the values of a phenomenan (variable) at times $t_1, t_2, \ldots\ldots\ldots t_n$ are $y_1, y_2, \ldots\ldots\ldots y_n$ respectively, then the series

t : t_1 t_2 $t_3 \ldots\ldots\ldots t_n$

y : y_1 y_2 $y_3 \ldots\ldots\ldots y_n$

Constitutes a time series.

12.4 IMPORTANCE OF TIME SERIES ANALYSIS

Analysis of time series is of special importance to Businessman, Economist, Scientist, Sociologist, Geologists and Research workers in various discipline. The following points indicate the importance of time series analysis

1. **Past Behaviour :** It helps us to understand the past behaviour of a variable. Analysis of past data discloses effect of various factors on the variable under study.
2. **Forecasting :** With the knowledge of past behaviour, it would be possible, within certain limits, to forecast the future behaviour of the variable under study. This helps in making future plans of action. Various five year plans of our country are based on the analysis of past data.
3. **Evaluation :** It helps in evaluation of current achievements. The review and evaluation of progress made on the basis of a plan are done on the basis of time series data. For example, the progress of various plans in our country is judged by the yearly rate of growth in GNP (Gross National Product).
4. **Comparison :** Analysis of time series helps in making comparison between one time period to another.It also facilitate to compare the actual values with the expected values. It provides a scientific basis for making comparisons by studying and isolating the causes of variation of the variable under study.

12.5 COMPONENTS OF TIME SERIES

Analysis of time series shows that if the values of a phenomenon are observed at different periods of time, the values so obtained are always fluctuating. The reason behing this type fluctuation is the joint effect of various factors (or forces). For example, the sales (y) of a product depends on (i) advertisement expenditure (ii) the price of the product (iii) the income of the consumer (iv) other competitive product in the market (v) tastes, fashions, habits and customs of the people and so on similarly, the price of a particular

product depends on its (i) demands (ii) various competitive products in the market (iii) availability of raw matirials (iv) transportation cost etc.

To know the proper nature of a time series, the statisticians have broadly classified the various operating factors (forces) affecting the values of a shenomenon in a time series into four parts. These are commonly known as **Components of a time series.** The four components of a time series are :

(a) Secular Trend or Trend (T)

(b) Seasonal Variation (S)

(c) Cyclical Variation (C)

(d) Irregular or Random Variation (I)

N. B. : Instead of the word 'variation', the word 'fluctuation' or 'movement' may also be used.

The changes in time series data are the result of the combined effect of these four components. Now, we proceed to a brief discussion on the various components of time series.

(a) Secular Trend or Trend (T) : The general tendency of the time series data showing continous growth (increase), stagnation or decline during a long period of time is called the **Secular Trend** or simply **Trend.** This phenomenon is usually observed in most of the time series relating to Business and Economics. e.g. an upward tendency is usually observed in time series relating to population, prices, income, production and sales of products, money in circulation etc, while a downward tendency is noticed in time series relating to birth rate, death rate, death rate due to epidemics etc due to the advancement in medical science, improved in medical facilities, better sanitation diet etc. Simpson and Kafka defined secular trend as

> **'Trend, also called Secular Trend or Long term trend, is the basic tendency of a time series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movement over a long time.**

Remarks :

It should be understood that the term 'long period' is a relative term and can not be defined exactly. In certain phenomenon, a period of few hours may be long enough whereas in some others even a period as long as 3-4 years may not be regarded as a long period. For example, if some agricultural or industrial production of a partcular product shows an increase over the past 3 years, this can not be termed as a long term tendency, for which we must have data for 7-8 years. On the other hand, while a patient is having high fever and as per advice of the attending physician, temperature reading are needed to be recorded on every 4 hours, in order to observe whether the effect of a particular medicine shows a decline in temperature or not, then in such a case, taking reading for a period of 48 hrs to 72 hrs would be termed as a long period.

Uses of Trend :

The study of the data over a long period of time enables us to have a general idea about the pattern of the behaviour of the phenomenon under consideration. This helps in business forecasting and planning of the future operations. Trend analysis enables us to compase two or more time series over different periods of time and draw important conclusions about them.

(b) Seasonal Variations (S) : The fluctuations or variations in the values of a time series over a period of one year or less are termed as seasonal variation. This variation occurs due to forces which are rhythmic in nature and which repeat themselves periodically over a span of less than a year. Such type of variations are present in a time series if the data are recorded quarterly (after every three months), monthly, weekly or daily basis. **Thus in a time series data where only annual figures are given, there can not be seasonal variation.** Most of the economic time series are influenced by swings e.s. prices, production and consumption of commodities, sales and profits of departmental stores, bank clearings and bank deposits etc. are all affected by seasonal variation. Seasonal variation takes place due to the following two causes–

(i) **Narural Causes :** As the name suggests, the various seasons or weather conditions and climatic changes play an important role in seasonal variation. For example, the sale of umbrella picks up very fast in raing season, the sale of woollen garment is more in winter, during summer demand for cold drinks and ice-cream increases. The prices of agricultural commodities always go down at the time of harvest and then pick up gradually.

(ii) **Rituals and Social Customs :** Man made rituals, social customs and teaditions are also responsible for seasonal fluctuation of a time series. For example, sale of jewellary goes up during diwali and marriage season, in the beginning of an academic session sale of books, papers, uniforms etc. go up, on the eve if new year, sale of greetings card increases.

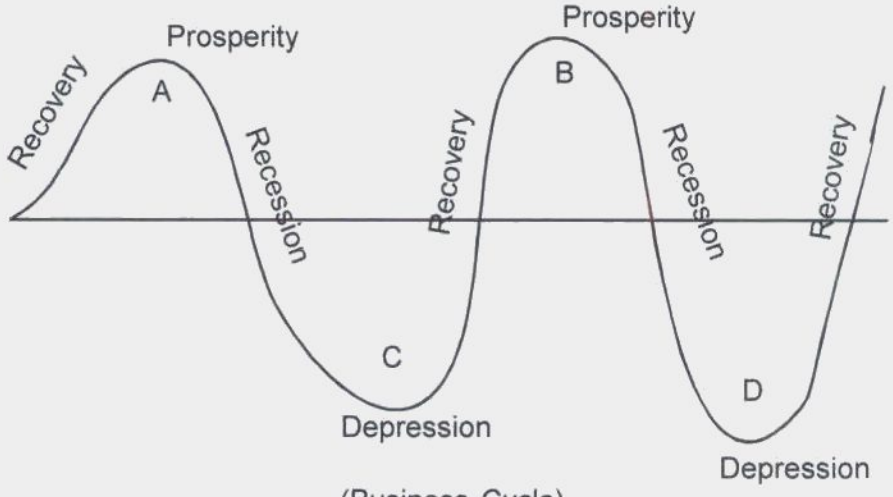
Uses of Seasonal Variation :

The main objective of seasonal variations is to isolate them from trend and study their effects. A study of the seasonal pattern is extremely useful for a businesman, producers sales managers etc. in planing future operations and in formulation of policy decisions regarding purchase, productions, inventory control and advertising programmes. In the absence of any knowledge of seasonal variation, a seasonal upswing may be mistaken as indicator of better business conditions while a seasonal slump may be misinterpreted as deteriorating business conditions.

(c) **Cyclical Variation (C) :** Cyclical variations are the oscillatory movement in a time series with period of oscillation greater than one year. The cyclical variations though more or less regular, are not necessairly uniformly periodic i.e. they may or may not follow exactly similar patterns after equal intervals in time one cyclic period normally lasts from 7 to 9 years.

Cyclical variations are found to exist in almost all business and economic time series where it is known as **business cycle** or **trade cycle**. The ups and downs (or rises and decline) in business recurring at intervals of time are the effects of cyclical variation. A **business cycle consists of four phases** namaly **prosperity** (boom), **recession, depression** and

recovery. Each phase changes gradually into the next phase in the given order until one business cycle is completed



(Business Cycle)

The time period between two successive booms or depressions (e.g. time period between A and B or C and D) is known as the length of a cycle.

Uses of Cyclical Variations : The study of cyclical variations is of great importance to business executives in the formulation of policies aimed at stabilising the level of business activity. A knowledge of cyclical variations enables a businessman to have an idea about the periodicity of booms and depressions and accordingly timely steps may be taken for maintaining stable market for his product.

Note : Though seasonal variations and cyclical variations are both periodic in nature, there is a significant difference between the two types of variations. These are mentioned below :

Seasonal variations	**Cyclical Variations**
(i) Seasonal variation takes place within one yeas.	(i) Cyclical variation takes place in a period more than one year. It usually takes place in 3 to 10 years time period.
(ii) In seasonal variation the periodic time remains the same.	(ii) In cyclical variation, the periodic time does not remain the same. It takes place only with some rough regularity.

(iii) It is mainly attributed to climatic changes and man made rituals and social customs.	(iii) Economic factors like price, production, sales, demand etc. are responsible for cyclical variations.

(d) Irregular Variations (I) : Irregular or Random Variations are such variations inherent in a time series which are caused by factors irregular (or erratic) in nature. These are purely random and unpredictable. These include all movements not already covered in trend, seasonal variation and cyclical variation. Irregular variations are caused by unforeseen factors like flood, famines, wars, earth. quakes, fire, epidemics, strike of trade union, terrorisim etc. As there are unpredictable, they can not be controlled by human hand. For almost all the time series, irregular variations are inevitable. Normally, these are short term fluctuation but sometimes their effect is so intense that they may give rise to new cyclical or other movements.

Because of their absolutely random character, it is not possible to study them exclusively, nor we can forecast or estimate these precisely. The best that can be done about such variations is to obtain their rough estimates (from the past experience) and accordingly make provisions for such abnormalities during normal time in business.

12.6 METHODS OF MEASURING SECULAR TREND

There are four methods which are generally used for the study and measurement of secular trend in a time series. There are :

(i) Graphical Method

(ii) Semi-Average Method

(iii) Moving Average Method

(iv) Method of Least Square.

Note : Here we will restrict our discussion to the **Method of moving average** only.

12.7 ESTIMATION OF TREND BY METHOD OF MOVING AVERAGE

METHOD OF MOVING AVERAGES **is a very simple and flexible method of measuring trend. This method consists in taking averages (arithmetic mean) of the values for a certain time span and placing it at the centre of the time span.**

The moving average is characterioed by a constant known as **period** of the moving average. Thus the moving average of period 'm' is a series of successive averages (Arithmetic means) of 'm' overlapping values at a time starting with 1st, 2nd, 3rd value and so on. Thus, for the time series values $y_1, y_2, y_3, y_4, y_5, \ldots\ldots\ldots\ldots$, the moving average (M. A) values of period 'm' are given by :

1st M.A = $\frac{1}{m}(y_1 + y_2 + \ldots\ldots + y_m)$

2nd M.A = $\frac{1}{m}(y_2 + y_3 + \ldots\ldots + y_{m+1})$

3rd M.A = $\frac{1}{m}(y_3 + y_4 + \ldots\ldots + y_{m+2})$ and so on we shall discuss two cases :

Case (i) When period is odd : It the period 'm' of the moving average is odd, then the successive values of the moving averages are placed against the middle most observation of the corresponding time period. For example, if m = 5, the first moving average value is placed against the 3rd value of the series, the second moving average value is placed against the 4^{th} value of the time series and so on.

Case (ii) When period is even : It the period 'm' of the moving average is even, then there are two middle periods and the moving average values are placed in between the two middle periods of the time interval it covers. Obviously, in this case the moving average values will not coincide with a period of the given time series and an attempt is made to synchronise them with the original data by taking a two-period average of the moving averages and placing them in between the corresponding time periods. This technique is called **centering** and the corresponding moving average values are called **centred moving average.** For example, if m = 4, the first moving average

value is placed against the middle of 2nd and 3rd time intervals, the second moving average value is placed in between 3rd and 4th time period and so on. These values are given by–

$$\bar{y}_1 = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$\bar{y}_2 = \frac{1}{4}(y_2 + y_3 + y_4 + y_5)$$

$$\bar{y}_3 = \frac{1}{4}(y_3 + y_4 + y_5 + y_6) \text{ and}$$

The centered moving averages are obtained on taking 2 period moving average of $\bar{y}_1, \bar{y}_2, \bar{y}_3$ and so on.

Thus

First centered moving average $= \frac{1}{2}(\bar{y}_1 + \bar{y}_2)$

Second centered moving average $= \frac{1}{2}(\bar{y}_2 + \bar{y}_3)$ and so on.

These centered moving averages are placed against 3rd, 4th, 5th period and so on.

MERITS AND DEMERITS OF MOVING AVERAGE METHOD :

Merits :

(i) This method does not involve any mathematical complexities and is quite simple to understand and use.

(ii) There is some inherent flexibility in this method. Addition of few more observations to the existing data does not affect the trend values already obtained. This will simple result in some more values at the end.

(iii) If the period of moving averages coincides with the period of cyclic fluctuations in the data, then this method completely eliminates the effect of cyclical fluctuations.

(iv) This method is also used for measurement of seasonal, cyclical and irregular variations.

DEMERITS :

(i) One of the prime limitations of moving average method is that we can not obtain the trend values for all the given observations. Some

trend values in the beginning and at the end have to forego depending on the period of moving average.

(ii) This method is not suitable for the purpose of forecasting.

(iii) The selection of period of moving average is a difficult task. Therefore, great care has to be taken in selecting the period, particularly, when there is no business cycle is the time series.

(iv) If the time series reveals lineartrend only then this method is applicable.

Example 1 : Calculate 3 yearly moving average from the data given below:

Year:	1980	1981	1982	1983	1984
Production ('oootons):	112	117	120	123	129
Year :	1985	1986	1987	1988	1989
Production ('oootons):	134	140	146	149	152

Solution:

Table for calculation of 3 yearly moving averages.

Year (1)	Prodiction ('ooo tons) (2)	3 Yearly Moving Total (3)	3 Yearly Moving Average (4)
1980	112		
1981	117	349	116.3
1982	120	360	120
1983	123	372	124
1984	129	386	128.7
1985	134	403	134.3
1986	140	420	140
1987	146	435	14.5
1988	149	447	149
1989	152		

Explanation: In column (3) total of the first three entries namely 112, 117 and 120 is written against the middle value i.e. 117 of this set. Then total of next. three entries i.e. 117, 120 and 123 is written against 120 and so on. Figures of column (4) are obtained by dividing the columns of figures (3) by 3.

Note : These 3 yearly moving average values are the trend values for the given time series.

Example 2 : Calculate trend values by the method of 4 yearly moving average from the following data.

Year :	1991	1992	1993	1994	1995	1996
Value :	44	64	59	52	57	71
Year :	1997	1998	1999	2000	2001	2002
Value :	66	64	71	77	82	80

Solution :

Table for calculation of 3 yearly mouing averages.

Year	Value	4 yearly moving Total	4 yearly moving average	2-period moving column	Centred average (trend value)
(1)	(2)	(3)	(4)	(5)	(6) = (5) ÷ 2
1991	44				
1992	64				
		219	54.75		
1993	59			112.75	56.375
		232	58		
1994	52			117.75	58.875
		239	59.75		
1995	57			121.25	60.625
		246	61.5		
1996	71			126	63
		272	68		
1997	66			132.5	68.75
		272	69.5		
1998	64			137.5	68.75
		278	69.5		
1999	71			143	71.5
		294	73.5		
2000	77			151	75.5
		310	77.5		
2001	82				
2002	80				

Explanation: In column (3) total of the first four entries namely 44, 64, 59 and 52 is written against the middle of this set i.e. in between 64 and 59. Then the total of next four entries i.e. 64, 59, 52 and 57 is written in between 59 and 52 and so on.

Figures of column (4) are obtained by dividing the figures of column (3) by 4.

In column (5) the total of the first two entries of column (4) i.e. 54.76 and 58 is put in between them i.e. corresponding to the year 1993. Then the total of second and third entries of column (4) i.e. 58 and 59.75 is put corresponding to the year 1994 and so on.

The figures of the column (6) are obtained by dividing the entries of column (5) by 2.

Note : The 4 yearly centred moving averages are the trend values for the given time series.

Example 3: From the data given below, determine trend values by 5 yearly moving average method.

Solution :

Year :	1970	1971	1972	1973	1974	1975	1976	1977	1978
Sales (Rs. lakhs)	130	127	124	135	140	132	129	127	145

Table for calculation trend values by 5 yearly moving average method

Year (1)	Sales (Rs. Lakhs) (2)	5 Yearly Moving Total (3)	5 Yearly Moving Average (Trend value) (4)
1970	130		
1971	127		
1972	124	656	131.2
1973	135	658	131.6
1974	140	660	132.0
1975	132	663	132.6
1976	129	673	
1977	127		
1978	145		

Note: The first moving total 656 of column (3) is the sum of first five values 14, 130, 127, 124, 135 and 140. The second moving total is 127 + 124 + 135 + 140 + 132 = 658 which can also be obtained by adding (132 – 130) = 2 with first moving total 656. Similarly, 3rd moving Total is 658 + (129 – 127) = 660 and so on.



CHECK YOUR PROGRESS

Q 1: Fill in the blanks.

(i) A time series consists of data arranged .....

(ii) An overall rise or fall in the time series is called .....

(iii) ..... can be made with the help of time series.

(iv) Cyclical variations are caused by .....

(v) ..... is the overall tendency of the time series data ..... or ..... over a ..... period of time.

(vi) Short term variations are classified as (a) ..... (b) .....

(vii) The four phases of business cycle (in order) are ....

(viii) For the annual data ..... component is absent.

(ix) The most important factors causing seasonal variations are .....

Q 2: With which component of time series would you associate each of the following.

(i) An era of prosperity.

(ii) An after puja sale in a departmental store.

(iii) A strike in a factory delaying production by 4 weeks.

(iv) The sale of umberlla in rainy season.

(v) Decrease in death rate due to advancement of medical science.

(vi) A need for increased rice production due to constant increase in population.

Q 3: Select the correct alternatives out of the given one.

(a) A time series is a set of data recorded.

(i) at time intervals

(ii) at successive point of time

(iii) periodically

(iv) All the above.

(b) The decline in birth rate is attached to

(i) Secular Trend (ii) Irregular Variation

(iii) Seasonal Variation (iv) Cyclical Variation.

(c) The sale of cold drink in summer season is attached to

(i) Secular Trend (ii) Seasonal Variation

(iii) Cyclical Variation (iv) Irregular Variation.

(d) Seasonal variation repeats during a period of

(i) 5 years (ii) 4 years

(iii) 7 years (iv) None of the above.

(e) Salient factors responsible for seasonal variation are :

(i) Weather (ii) Festival (iii) Social Custom

(iv) All the alove.

(f) Increase in death rate due to an epidemic is attached to

(i) Secular Trend (ii) Cyclical Variation

(iii) Seasonal Variation (iv) None of the alove.



12.8 LET US SUM UP

- Time series is a set of observation recorded at successive periods of time.
- The factors which affect the components of time series.
- The component of time series responsible for long term movement in a time series is called secular trend or simply trend.
- The component of time series that shows a regular and periodic pattern of movement over a time span of less than 1 year is called seasonal component.

- Cyclical variation is an oscillatory movement in a time series with period of oscillation greater than 1 year.
- Irregular movement are such variations inherent in a time series.
- Analysis of time series helps us to understand past behaviour, forecasting, evaluation of current achievement and making comparison between two time period.
- There are four methods for studying and measuring secular trend.
- One of the methods of measuring trend is 'Moving average method' which involve smoothing a time series by averaging succesive groups of data for a certain time span.



12.9 FURTHER READINGS

1) Gupta S. C. *Fundamentals of Statistics*. Himalaya Publishing House.
2) Gupta S. C and Gupta Indira. *Business Statistics*. Himalaya Publishing House.
3) P. N. Arora, Sumeet Arora, S. Arora. *Comprehensive Statistical Methods*. S. Chand & Company.
4) R. K. Ghosh and S.Saha. *Business Mathematics and Statistics*. New Central Book Agency (P) Ltd.
5) Agarwal D. R. *Business Statistics*. Delhi. Vrinda Publication.
6) Sharma J. K. *Business Statistics*. Pearsor Education Ltd.



12.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (i) Chronologically (ii) Trend / Secular Trend (iii) Forecasts (iv) Business / Trade Cycles (v) Secular Trend, increase, decrease, long (vi) (a) Seasonal (b) Cyclical (vii) Boom, Recession, Depression, Recovery (viii) Seasonal (ix) Weather and Social Customs.

Ans to Q No 2: (i) Cyclical (ii) Seasonal (iii) Irregular (iv) Seasonal (v) Secular (vi) Secular.

Ans to Q No 3: (a) (iv) (b) (i) (c) (ii)
(d) (iv) (e) (iv) (f) (iv)



12.11 MODEL QUESTIONS

Q 1: Define time series.

Q 2: Explain various components of time series.

Q 3: What is meant by Time Series Analysis? Discuss its importance in business.

Q 4: What do you mean by Secular Trend? Explain by giving examples.

Q 5: Explain the meaning of seasonal variations, with illustrations.

Q 6: What are cyclical variation? How are they caused?

Q 7: How do cyclical variations differ from seasonal variation.

Q 8: Which component of time series is associated with the following:

(i) Recession

(ii) A fire in a factory delaying production for three weeks.

(iii) An era of depression.

(iv) Sale of ice-cream in summer.

(v) Increase in the production of rice in Assam due to application of modern technique.

Q 9: Obtain Trend Value by 5 yearly moving average method.

Year :	1990	1991	1992	1993	1994	1995	1996	1997	1998
Value:	242	250	252	249	253	255	251	257	260
Year :	1999	2000							
Value:	265	262							

Q 10: Calucate Trend Values by 3 yearly moving average method.

Year :	1981	1982	1983	1984	1985	1986	1987	1988
Sugar								

Production								
(lakh tons)	37.4	31.2	38.7	39.6	47.6	42.6	48.5	56.2
Year :	1989	1990						
Sugar								
Production								
(lakh tons)	60.6	58.2						

Q 11: Calculate trend values by 4 yearly moving average method.

Year :	1985	1986	1987	1988	1989	1990	1991	1992	1993
Value:	39	59	53	46	51	65	59	58	65
Year :	1994	1995	1996						
Value:	71	76	74						

Q 12: Determine the period of the moving average for the data given below and calculate moving average value for that period.

Year :	1	2	3	4	5	6	7	8	9
Value:	130	127	124	135	140	132	129	127	145
Year :	10	11	12	13	14	15			
Value:	158	153	146	145	164	170.			

[**Hint :** Since the peaks of the given data occur at the years 1, 5, 10 and 15, the data clearly exhibits a regular cyclical movement with period 5, hence the period of moving average for determining trend valus is also 5 i.e. period of cyclical variations]

*** ***** ***

UNIT 13: BASIC CONCEPT IN SAMPLING

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Defining population, parameter and sampling

13.3.1 Random and Non Random Sampling

13.3.2 Sampling with and without replacement

13.3.3 Large samples

13.4 Errors in Sampling

13.5 Let Us Sum Up

13.6 Further Readings

13.7 Answer to check your progress

13.8 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the concept of a parameter and a statistic
- explain the concepts of a sampling with and without replacement
- explain Random and Non Random Sampling
- discuss about standard error
- discuss about different methods of sampling
- explain merits and limitations of sampling.

13.2 INTRODUCTION

Often we are interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group, which may be difficult to do, we examine only a small part. We do this with the aim of inferring certain facts about the group of individuals.

The entire group of individuals is known as population. The word 'population' used in statistics has altogether a different meaning.

You all know the dictionary meaning of the word population. But in

mathematics we use the word with a bit different meaning. Mathematician use the word population to refer not only to people but to all items that have been chosen for study. When it is possible to examine every item in the population that we have taken for study, then it is called a Census. But complete enumeration is not always possible in all situations. Then we take another procedure, which we will refer as sampling method. Mathematician use the word sample to describe portion chosen from the population. For example we have population of heights of all female inhabitants of Guwahati or a population of incomes of families living in Guwahati etc. A population may be finite, infinite or hypothetical (for example, outcomes of tossing a coin infinite number of times).

Often you have seen in news paper that Government estimated the production of rice in a year. Do they take complete enumeration? Certainly not. When they do this they select some areas and measure the production there and after that they do some mathematical calculation. This is just an example. Every estimation is done like this. Care should be taken in selecting the areas; otherwise you will not get good result. Sample must be the representative of population.

13.3 DEFINING POPULATION, PARAMETER AND SAMPLING

In mathematics, we describe samples and population by using measures such as mean, median, mode and standard deviation (s.d). When these terms describe the characteristics of a sample then they are called ***Statistics***. A 'Statistic' is a function of the sample. Examples are arithmetic mean $(\bar{x})$ and sample variances (σ^2).

When they describe the characteristics of a population they are called ***Parameters***.

A ***population*** is the totality of any kind of units under consideration. It may be finite or infinite.

A sample is any portion of the population selected for study. The process of obtaining samples is called ***sampling***.

The fundamental assumption underlying the theory of sampling is random sampling, which consists in selecting the individuals from the populations, in such a way that each individuals of the population has the equal chance of being selected.

Two types of sampling will be discussed - (1) Random Sampling and (2) Non random Sampling.

Purpose of Sampling: A sample is not studied for its own sake. The main objective of its study is to draw inference about the population. In other words sampling is only a tool which helps us to know the characteristic of the universe or population by examining only a small part of it.

Simple sampling:

Sampling of attributes may be regarded as the drawing of samples from a population whose members possess the attribute A or not (-A). The choosing of an individual in a sampling is called an event or trial. Possession of a specified attribute A by the individual is called a success.

Simple sampling means random sampling in which each event has the same probability "p" of success and the probability of success is independent of the success or failure of the events in the preceding trials. Thus simple sampling is a special case of random sampling in which trials are independent and probability of success is constant.

Let us consider an example. In an urn there are 2 white and 3 black balls. Probability of drawing a white ball is 2/5. If a white ball is drawn on the first trial, probability of drawing a white ball will be 2/5. In the second trial if the ball is not replaced the probability of drawing a white ball will be ¼. Hence the sampling though random is not simple.

A random sampling from an infinite population is always a simple since drawing of one individual cannot affect the probability. But a sampling from finite population may or may not be simple.

Let us take a simple sample of 'n' members, which will be identical with a series of n independent trials with probability of success p (constant). The probabilities of 0, 1, 2, ... ,n success are the terms of binomial expansion of $(p+q)^n$ which can be written in sequence. The distribution so obtained is called sampling distribution of number successes in the sample.

The mean value and S.d (standard error S.E.) are (Missing) np and $\sqrt{npq}$ respectively.

Now proportion of successes in a sample is the number successes divided by the number of members in the sample. Hence

Mean of proportion of success = p

(S.E) of proportion of success = $\sqrt{\frac{pq}{n}}$

Already we have mentioned that care should be taken in selecting samples, otherwise inference made from the result get will mislead.

There are two types of sampling:

i) Random sampling

ii) Nonrandom sampling.

In random sampling all the items in the population have equal chance of being chosen in the sampling. Whereas in non random sampling personal knowledge and opinion are used to identify the items from the population that are to be included in the Sample. A Sample selected by this sampling is based on someone's expertise. The rigorous statistical analysis that can be done by random sample but cannot be done by non-random sample. Look to the following table for clear understanding the distinction between populations and samples.

	Population	Sample
Definition	Complete enumeration of items is considered	Part of the population chosen for study
Characteristics	Parameters	Statistics
Symbols	Population size = N Population mean = μ Population S.d = σ	Sample size = n Sample mean = $\bar{x}$ Sample S.d = s

In random sample, the probability of an element of the population will or will not be included in the Sample are known. As a result we can assess mathematically the estimates of the population characteristics from the result of the Sample.

13.3.1 Random and Non Random Sampling Methods

Sampling is a process of studying the characteristics of a population by selecting samples from it. Let us discuss some random and non-random methods–

(I) Methods of random Sampling:

When a sample is drawn from a population, it is necessary to decide which method should be applicable. The various methods of sampling can be grouped under the heads as **random sampling** and **non- random sampling.**

Some of the sampling methods are as follows:

A. Random Sampling Method:
   - (i) Simple Random Sampling
   - (ii) Systematic Sampling
   - (iii) Stratified Sampling
   - (iv) Cluster Sampling

B. Non-random Sampling Methods:
   - (i) Judgment Method
   - (ii) Quota Sampling
   - (iii) Convenience Sampling

1. Simple random sampling:

This method selects samples that allows

i) Each possible sample to have an equal probability of being selected.

ii) Each item in the population has equal probability of being included in the Sample.

Let n units of the sample be drawn from the population one by one, and after drawing the item selected is returned to the population before the next drawing is made. In this case at each stage of drawing, each of N items of the population gets the same probability $\frac{1}{N}$ of being selected. This sampling procedure is called Simple random Sampling with replacement. In this case, the same

unit of the population may occur more than once in the Sample. There will be n possible Samples and each has the probability $\frac{1}{N^n}$

In other case, the n members of the Sample are drawn one by one, but at any stage members drawn in the previous drawing is not returned to the population. This procedure is known as Simple random Sampling without replacement.

In this case no member of the population can occur more than once in the sample. There are N_{C_n} possible Samples and each Sample has the probability.

$$\frac{n}{N}\cdot\frac{n-1}{N-1}\cdot\frac{n-2}{N-2}\cdot\frac{n-3}{N-3}\text{ - - - - - }\frac{1}{N-n+1}=\frac{1}{N_{C_n}}$$

In this case also the probability that any particular population units will be selected of the drawing is

$$\frac{N-1}{N}\cdot\frac{N-2}{N-1}\cdot\frac{N-3}{N-2}\text{ - - - - - }\frac{N-r+1}{N-r+2}\cdot\frac{1}{N-r+1}=\frac{1}{N}$$

Technique of Simple random Sampling:

The easiest way to select a Sample randomly is to use random numbers. These numbers can be taken from a table of random digits or can be generated by a computer programming.

For example, suppose a company has 1000 employees and wish to interview a randomly chosen sample of 10. Write the name of each one on a slip of paper and put the papers on a cylindrical drum and rotate it so that they are thoroughly mixed. Now draw 10 slips at random.

2. **Systematic sampling:** In these sampling elements are selected from the population at a uniform interval that is measured in time, space, order or any others.Here each element in the population has the same chance of being selected from the sample. For example suppose we want to interview 20th students on a college campus. We choose a random starting point in the first 20 names in the student register book and then pick every 20th name thereafter. In this sampling also each element has an equal

chance of being selected; but each sample does not have an equal chance of being selected. In this point this sampling differs from Simple random Sampling. In this sampling there is problem of introducing error into the Sample process. The only advantage of this sampling is that it takes less time and lower cost.

3. **Stratified Sampling:** In this Sampling, population is divided into homogeneous groups. These groups are called strata. There are two approaches: i) From each strata, select specified number of elements at random. ii) Draw equal number of elements from each strata and give weight to the result according to the Stratum's proportion in the population. In Stratified sampling, every element in the population has chance being selected. In this sampling each group has small variation within itself. There are wide variations between the groups.
4. **Cluster Sampling:** In this sampling also population is divided into groups and the groups are called Clusters. The assumption is that individual clusters are representative of population. Then random Samples are selected from the clusters. In case of cluster sample there is considerable variation within each group, but the groups are essentially similar to each other. In cluster, stratified and systematic samples attempts are made to approximate simple random sample. These three methods are developed for economical and physical considerations. Simple random sampling is basis of all the samplings and it is fundamental one. Principles of simple random sampling are the foundations of statistical inferences.

(II) Non- Random Sampling Methods

Judgment Method: In this method the choice of sample items depends exclusively on the judgment of the investigator. For example, if a sample of 10 artist is to be selected from a group of 50 artists, then the investigator would select 10 artists who, in his opinion, represent the group.

Though this method is simple, it is not scientific because the result may be considerably affected by the personal biasness of the investigator. This method is generally applicable when the size of the sample is small. In general, for solving business problems, making public policy decisions, pilot studies etc we use this method.

(i) **Quota Sampling:** It is a type of judgment sampling where quotas are set up according to given criteria but within the quotas the selection of sample items depends on personal judgment. For example, if you want to take interview of 100 people, out of which there should be 50 Hindu, 50 Muslim, 50 Sikhs and 50 Christian. Within these quotas, the interviewer is free to select the people interviewed.

(ii) **Convenience Sampling:** In this method the sample is obtained from readily available lists, such as telephone directories, or automobile registration are example of convenience sampling.

Merits of Sampling Method:

1. It is less time consuming compared to census.
2. It is less costly either in terms of money or in terms of man powers, because only a small portion of the total population is studied.
3. More detailed information can be gathered as the size of the sample is small compared to total population.
4. A sample survey generally gives data of better quality than a complete census, because in a sample survey it may be possible to employ better trained personnel, effect better supervision or use better equipment than is possible or feasible in a complete census.
5. For infinite population, sampling is the only solution.

Limitations of Sampling:

1. Without proper planning, the samples result may be inaccurate and misleading.
2. In absence of qualified persons, the information obtained from samples surveys can not be relied upon.

3. If the information is required from whole population then sampling is meaningless.



CHECK YOUR PROGRESS

Q 1: Fill in the blanks with appropriate words:

(i) In ..... sampling all the items in the population have equal chance of being chosen in the sampling.

(ii) In ..... sampling personal knowledge and opinion are used to identify the items .

(iii) Convenience sampling is a method of ..... sampling.

(iv) In ..... Sampling population is divided into homogeneous groups.

(v) Sampling is less time consuming compared to .....

13.3.2 Sampling With And Without Replacement

Sampling where each member of the population may be chosen more than once is called a sample with replacement, while sampling where each member cannot be chosen more than once is called as sample without replacement.

If we draw an object from an urn, we have the choice of replacing or not replacing the object into the urn before we draw again. In the first case a particular object can come up again and again, whereas in the second case it can come up only once.

Random sampling:

If in a sampling each member of the population has the same chance of being selected then it is called a random sampling.

A random sample of size 'n' taken with replacement from a population with distribution p(x) = P (X=x) is a collection of random variables $x_1, x_2 \ldots\ldots x_n$ each having the same distribution as x and $x_1, x_2 \ldots\ldots x_n$ are independent.

Then μ = Mean of the population = $\frac{\sum x_i p_i}{n}$ = E(n)

Let us denote N numbers of the elements of the set be $x_1, x_2, \ldots\ldots, x_n$, taking a sample consisting of n elements from the set of N numbers. Sample vatiables are $x_1, x_2, \ldots\ldots, x_n$.

Mean of population $\mu = \frac{\sum_{i=1}^{N} xi}{N}$

Mean of sample $\bar{x} = \frac{\Sigma xi}{n}$

Now $E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$ ——(1)

$$E(x_i) = \frac{X_1}{N} + \frac{X_2}{N} + - - - - + \frac{X_n}{N}$$

[As x_i can take any element from $X_1, X_2, \ldots\ldots, X_N$ and its probability is $\frac{1}{N}$]

$\therefore E(x_i) = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu$ ——(2)

From (1) and (2)

$$E(\bar{x}) = \frac{1}{N}\sum_{i=1}^{n}\mu = \frac{1}{n}.n\mu = \mu$$

Variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$

If all probabilities are equal and equal to $\frac{1}{n}$ then,

Mean = $\frac{\sum x_i}{n}$ = E(x)

Variance = Var(x) = σ^2 = $E(x^2) - [E(x^2)]^2$

= σ^2

Where $[E(x)]^2 = \sum_{i=1}^{n} xi^2 p(xi)$



EXERCISE 1

A population consists of five numbers 2, 3, 6,8 and 11. Consider all possible samples of size 2 that can be drawn with replacement from this population. Find mean and s.d. of the population. Also find mean and s.d. of the samples.

Solution: Let μ = Mean of population

$$= \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6$$

Again $\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$

$$= \frac{16+9+0+4+25}{5} = 10.8$$

Therefore Standard Deviation of the population = σ

$$= \sqrt{10.8}$$
$$= 3.29$$

The samples of size of that can be drawn with replace mean are

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

They are 25 in numbers the corresponding sample means are

2.0	2.5	4.0	5.0	6.5
2.5	3	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

Mean of the sampling distribution is

$$\mu_{\bar{x}} = \frac{\text{Sum of the sample means}}{25}$$

$$= \frac{150}{25} = 6.0$$

So $\mu_{\bar{x}} = \mu$

For S.D of samples, subtract 6 from all the means of the sample and square them.

Add all squared members and divide the result by 25.

$$\sigma^{2}_{\bar{x}} = \frac{135}{25} = 5.4$$

S.D. of samples = $\sigma_{\bar{x}}$

$$\sigma_{\bar{x}} = \sqrt{5.4} = 3\sqrt{.6} = 2.32$$

Let us consider the same problem without replacement.

Here also, $\mu = 6$

And $\sigma = 3.29$

The samples in this case will be (2,3),(2,6), (2,8), (2,11), (3,6), (3,8),(3,11), (6,8), (6,11) and (8,11).

They are 10 in numbers.

The corresponding sample means are 2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5 and 9.5.

Mean of the sampling means are

$$\mu_{\bar{x}} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$$

$$= \frac{60}{10}$$

$$= 6$$

So, $\mu_{\bar{x}} = \mu$

Variance of the sample mean is

$$\sigma^{2}_{\bar{x}} = \frac{(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2}{10}$$

$$= \frac{40.5}{10} = 4.05$$

$$\sigma_{\bar{x}} = 2.01$$

13.3.3 Large samples

Suppose that a large number n of independents Bernoull's trials is performed and x successes are obtained. The hypothesis we want to test is that the probability of success in each trial is p. Assuming the hypothesis to be correct, the mean and variance of the sampling distribution of the number of success are np and npq.

Again we know for large n, Z= $\frac{x-np}{\sqrt{npq}}$ is distributed as a standard normal variate.

From table P ($|Z|$>3) = 0.0027. Therefore we conclude that the hypothesis is improbable if the difference between the observed number of successes x and the expected number of successes np, is highly significant.

If $|Z|$ <1.96 the data is consistent with the hypothesis.

If $|Z|$ >1.96 the difference is significant.

Let the observed proportion of success in a sample of n observation be p then,

Mean = p and variance = $\frac{pq}{n}$

Thus $\frac{x/n-p}{\sqrt{pq/n}}$ ~ N(0,1)

This test is valid for large n only.

Standard error: The s.d. of the Sampling distribution of statistic is known as Standard error (SE). If n be the sample size and σ^2 is the population variance then SE of the statistic $\bar{x}$ of the sample is $\frac{\sigma}{\sqrt{n}}$.

SE plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

If t be any statistic then for large samples

$$Z=\frac{t-E(t)}{S.E(t)} \sim N(0, 1)$$

Thus if the discrepancy between the observed and expected value of the statistic Z is greater than Z. S.E, then hypothesis is rejected at certain level of significance.

13.4 ERRORS IN SAMPLING PROCEDURE

There are two types of errors in sampling procedure. These are as

follows:

1. Sampling error and
2. Non sampling error

Sampling error:

A sample covers only a portion of the population from which it is drawn. The information gathered through sample is used for drawing certain inferences about the characteristic of the population. A discrepancy may arise between the population parameter and the sample statistic. The error thus occurs is called sampling error. There are two factors which will lead to sampling error. They are sampling bias and chance factor. When the method of selection is defective, sampling bias will be caused. This error can be minimized by use of proper sampling method. Even when sample bias is absent, discrepancy may occur due to chance.

Non- sampling error:

- **Response error:** It is due to response bias- viz vague answer, wrong answer and inaccurate answer given by a respondent.
- **Observational error:** It is due to bias observation.
- **Errors due to non response:** An interviewer may not be able to get information from a respondent.

Test of Significance

A very important aspect of Sampling theory is the study of the test of significance, which enables us on the basis of the Sample results, if the deviation between the observed sample statistic and the hypothetical parameter value.

The technique of randomization used for the selection of Sample units makes the test of significance valid. For applying the test of significance we first set up the hypothesis- a definite statement about the population parameter. This hypothesis is called null hypothesis.



EXERCISE 2

A dice is thrown 9,000 times and a throw of 3 or 4 observed 3,240 times. Show that the dice is biased.

Solution: Here, N= 9,000

X= number of successes = 3,240

Hypothesis: Dice is unbiased

P= probability of getting 3 or 4 = $\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$

Expected mean = np = 9,000 x $\frac{1}{3}=3{,}000$

n is large $\Rightarrow$ $|Z| = \frac{x-np}{\sqrt{npq}}$

Standard Error (S.E.) = $\sqrt{npq}$ = $\sqrt{9000\times\frac{1}{3}\times\frac{2}{3}}$

= $\sqrt{2000}$ = 20 $\sqrt{5}$

$Z = \frac{3240-3000}{20\sqrt{5}} = \frac{240}{20\sqrt{5}} = \frac{12\sqrt{5}}{5}$

= 24 $\sqrt{5}$

= 5.36

Since $|Z|$ > 3, the dice is biased.



EXERCISE 3

A random sample of 500 pineapples was taken from a large contingent and 65 were found to be bad. Show that Standard Error of the population of bad ones in a sample of this size is 0.015

Solution: Here, n= 500

x = number of bad pineapples in the sample

= 65

p = 65/500 = 0.13

q = 1- 0.13

= 0.87

S.E. = $\sqrt{\frac{pq}{n}}$ = $\sqrt{\frac{0.13\times0.87}{500}}$ = 0.015



CHECK YOUR PROGRESS

Q 2: In a sample of 1000 people in Guwahati, 840 are rice eaters and rest is wheat eaters. Calculate Z.

.....

.....



13.5 LET US SUM UP

- In this unit we have developed the concept of random sampling and discuss simple sample as a particular case of random sample.
- We have also discussed sampling with and without replacement. In all these discussion we have taken sample from the normal population.
- We have also discussed about different methods of sampling - simple random sampling, systematic sampling, stratified sampling and cluster sampling (all these are random sampling).
- We have discussed about some non random sampling like- judgment method, quota method and convenience sampling.
- We have also discussed about sampling and non sampling errors.



13.6 FURTHER READINGS

1) Agarwal, D. R. (2006). Business Statistics. Delhi, Vrinda Publications.

2) Gupta S.C. (1994). Fundamentals of statistics. New Delhi, Himalayan Publishing House.

3) Rajagopalan, S. P. & Sattanthan R. (2009).Business Statistics and

Operations Research. New Delhi, Tata McGraw-Hill.

4) Sharma, J. K. (2007). Business Statistics. New Delhi. Pearson Education Ltd.

5) Verma, A. P. (2007). Business Statistics. Guwahat, Asian Books Private Limited.



13.7 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (i) random sampling, (ii) non random sampling, (iii) non random sampling (iv) Stratified sampling,(v) census.

Ans to Q No 2: 24.64.



13.8 MODEL QUESTIONS

Q 1: A population is made up of groups that have wide variation within each group but little variation from group to group. What will be the appropriate type of sampling for this population?

Q 2: List the advantages of sampling over census.

Q 3: A population consists of four members 3, 7, 11 and 15. Find all possible samples of size 2 that can be drawn with replacement from this population. Find population mean and population s.d. Find also mean of sampling mean and s.d. of sampling s.d.

Q 4: Solve the above problem without replacement.

Q 5: The masses of 1500 ball bearings are normally distributed with mean 22.40 k.g. and s.d. of 0.048 k.g. If 3000 random samples of size 36 are drawn from this population, determine the mean and s.d. of the sampling means if sampling is done (a) with replacement (b) without replacement.

Q 6: In a random sample of 500 people from a particular district, 300 are found to be smokers. In one of 1000 people from another district, 550 are smokers. Calculate Z.

Q 7: A coin is tossed 10,000 times and it turns up head 5,195 times. Can the coin be regarded as unbiased?

Q 8: In some dice throwing experiments, 49,152 times of 75,145 yielded 4, 5, or 6. Is the dice unbiased?

Q 9: A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Calculate Standard Error.

Q 10: A sample of 900 members has mean 3.4 cm and s.d. 2.61 c.m Is the sample from a large population of mean 3.25 cm and s.d. 2.61 c.m.?

*** ***** ***



REFERENCES

1) Agarwal, D.R. (2006). *'Business Statistics'*, Delhi, Vrinda Publications.

2) Gupta S. C. (1994), *'Fundamentals of statistics'*, New Delhi, Himalayan Publishing House.

3) Rajagopalan, S.P. & Sattanthan R. (2009), *'Business Statistics and Operations Research'*. New Delhi, Tata McGraw-Hill.

4) Sharma, J. K. (2007), *'Business Statistics'*. New Delhi, Pearson Education Ltd.

5) Verma, A. P. (2007), *'Business Statistics'*. Guwahati, Asian Books Private Limited.