B.A. IInd Year
ECONOMICS

Course Code : ECONA 205
Course Code : 04 (SEC-3)

Statistical Methods – II

Units 1 to 14

By: Prof. Sanju Karol
Dr. Dharam Pal

Centre for Distance and Online Education

Himachal Pradesh University

Summer Hill, Shimla, 171005

Course No. **ECONA205**
Course title: **Statistical Methods – II**
Nature of Course: **SEC – 3**
Number of credits: **4**

COURSE OUTLINE

Block	Description
I.	**CORRELATION ANALYSIS** Meaning, Significance of the Study of Correlation. Correlation and Causation. Types of Correlation: Positive, Negative, Partial, Multiple, Linear and Non-Linear. Methods of Studying Correlation: Scatter Diagram Method, Graphic Method, Karl Pearson's Coefficient of Correlation, Spearman's Rank Correlation. Properties and Interpretation of Correlation Coefficient.
II.	**REGRESSION ANALYSIS** Meaning - Difference between Correlation and Regression - Regression Lines - Regression Equations of X on Y and Y on X Only. Regression Coefficients. Elementary application of regression in demand, supply, consumption and investment functions.
III.	**ANALYSIS OF TIME SERIES** Meaning and Importance of Time Series. Components of Time series, Measurement of Trend: Graphic Method, Semi Average Method, Moving Average Method, Least Square Method. Applications in Economics.
IV.	**INDEX NUMBERS** Meaning, Characteristics, Importance and Uses, Classification. Types of Index Numbers: Price, Quantity and Value Index Numbers. Special Purpose Indices: Cost of Living Index, Wholesale Price Index, Consumer Price Index. Problems in construction of Index Numbers.

CONTENTS

Unit No.	Title	Page No
1	Correlation Analysis: Meaning and Significance	3
2	Correlation Analysis: Graphic Methods	13
3	Correlation Analysis: Algebraic Methods	19
4	Regression: Meaning and Significance	34
5	Regression: Obtain Regression Equations	46
6	Regression: Numerical Problems	61
7	Elementary Applications of Regression	76
8	Time Series: Meaning and Significance	85
9	Time Series: Measurement of Trends-I	93
10	Time Series: Measurement of Trends-II	102
11	Time Series: Measurement of Trends-III	110
12	Index Number: An Introduction	124
13	Weighted Index Number	135
14	Special Purpose Index Numbers	150

Unit-01

CORRELATION ANALYSIS: MEANING AND SIGNIFICANCE

STRUCTURE

1.1 Introduction

1.2 Learning Objectives

1.3 Correlation

Self-Check Exercise 1.1

1.4 Utility of Correlation

Self-Check Exercise 1.2

1.5 Difference between Correlation and Causation

Self-Check Exercise 1.3

1.6 Types of Correlation

1.6.1 Positive Correlation and Negative correlation

1.6.2 Simple, Partial and Multiple correlations

1.6.3 Linear and Non-linear correlation

Self-Check Exercise 1.4

1.7 Correlation Coefficient

Self-Check Exercise 1.5

1.8 Significance of Correlation Analysis

Self-Check Exercise 1.6

1.9 Degrees of Correlation:

1.9.1. Perfect Positive Correlation

1.9.2 Perfect Negative Correlation

1.9.3 Limited Degree of Positive Correlation

1.9.4 Limited Degree of Negative Correlation

1.9.5 Zero Degree Correlation (No Correlation)

Self-Check Exercise 1.7

1.10 Methods of Measuring Correlation

1.10.1 Graphic Methods

1.10.1.1 Scatter Diagram Method

1.10.1.2 Correlation Graph Method

1.10.2 Algebraic Methods

1.10.2.1 Karl Pearson's Co-efficient of Correlation

1.10.2.2 Spearman's Rank Correlation Method

1.10.2.3 Concurrent Deviation Method

Self-Check Exercise 1.8

1.11 Summary

1.12 Glossary

1.13 Answers to Self-Check Exercise

1.14 References/Suggested Readings

1.15 Terminal Questions

1.1 INTRODUCTION

In real-world scenarios, we often encounter situations that require statistical analysis of one or more variables. When data involves only a single variable, it is referred to as univariate data. Examples include price, income, demand, production, weight, height, and academic scores. The examination of such data is known as univariate analysis. When data consists of two variables, it is termed bivariate data. Examples include relationships such as rainfall and agricultural output, income and consumption, price and demand, or height and weight. The study of these paired variables is called bivariate analysis. If the data involves three or more variables, it is categorized as multivariate data. For instance, agricultural production is influenced by factors like rainfall, soil quality, and fertilizer use. The statistical method used to analyze the relationships between two or more variables is known as correlation analysis.

1.2 LEARNING OBJECTIVES

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

- define correlation
- list the utility of correlation
- explain the different types of correlation
- explain different degrees of correlation

1.3 CORRELATION

Two or more variables are said to be correlated if the change in one variable results in a corresponding change in the other variable. According to Simpson and Kafka, "Correlation analysis deals with the association between two or more variables".

Lun Chou defines, "Correlation analysis attempts to determine the degree of relationship between variables".

Boddington states that "Whenever some definite connection exists between two or more groups or classes of series of data, there is said to be correlation."

Croxton and Cowden says, "When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation".

A.M. Tuttle says, "Correlation is an analysis of the co-variation between two or more variables."

W.A. Neiswanger says, "Correlation analysis contributes to the understanding of economic behaviour, aids in locating the critically important variables on which others depend, may reveal to the economist the connections by which disturbances spread and suggest to him the paths through which stabilizing forces may become effective.

L.R. Conner says, "If two or more quantities vary in sympathy so that the movement in one tends to be accompanied by corresponding movements in others than they are said to be correlated.

Self-Check Exercise 1.1

Q1. What do you mean by Correlation?

1.4 UTILITY OF CORRELATION

The study of correlation is very useful in practical life as revealed by these points.

1. With the help of correlation analysis, we can measure in one figure, the degree of relationship existing between variables like price, demand, supply, income, expenditure etc. Once we know that two variables are correlated then we can easily estimate the value of one variable, given the value of other.
2. Correlation analysis is of great use to economists and businessmen; it reveals to the economists the disturbing factors and suggests to him the stabilizing forces. In business, it enables the executive to estimate costs, sales etc. and plan accordingly.
3. Correlation analysis is helpful to scientists. Nature has been found to be a multiplicity of interrelated forces.

Self-Check Exercise 1.2

Q1. Write down the utility of correlation?

1.5 DIFFERENCE BETWEEN CORRELATION AND CAUSATION

Correlation and causation are often mistaken for one another, but they have distinct meanings. Correlation between two variables does not necessarily imply that one causes changes in the other. In simple terms, while two variables may change together, this relationship does not confirm that one variable is responsible for the change in the other.

When two variables exhibit a correlation without a direct cause-and-effect relationship, this is referred to as **spurious correlation** or **nonsense correlation**. Such correlations can arise for various reasons, including:

(i) **Pure Chance Correlation** – This occurs in small sample sizes. For example, there might be a correlation between the income levels and body weights of a group of four individuals, even though there is no causal link between income and weight. Such correlations may result from random sampling variations or biases in sample selection.

(ii) **Influence of a Third Variable** – Sometimes, two correlated variables are both affected by one or more external factors. For instance, a high correlation may exist between the yield per acre of rice and tea, not because one crop influences the other, but because both are affected by a common factor, such as rainfall.

(iii) **Mutual Influence Between Variables** – In some cases, two variables may influence each other simultaneously, making it difficult to determine which one is the cause and which is the effect. This reciprocal relationship can lead to correlation without a clear causation.

Self-Check Exercise 1.3

Q1. Distinguish between correlation and causation.

1.6 TYPES OF CORRELATION

Correlation can be categorized as one of the following:

1.6.1 Positive and Negative Correlation.

1.6.2 Simple, Partial and Multiple Correlation

1.6.3 Linear and Non-Linear Correlation

1.6.1 Positive and Negative Correlation

Positive Correlation: Positive correlation occurs when two variables move in the same direction. This means that an increase in one variable corresponds to an increase in the other, and similarly, a decrease in one variable is associated with a decrease in the other.

Example 1:

A	10	20	30	40	50
B	80	100	150	170	200

Example 2:

X	78	60	52	46	38
Y	20	18	14	10	5

Negative Correlation: Negative correlation occurs when two variables move in opposite directions. This means that as one variable increases, the other decreases, and vice versa. In simpler terms, if a rise in one variable's value corresponds to a decline in the other, or if a drop in one leads to a rise in the other, the relationship is considered a negative correlation.

Example 1:

A	5	10	15	20	25
B	16	10	8	6	2

Example 2:

X	40	32	25	20	10
Y	2	3	5	8	12

1.6.2 Simple, Partial and Multiple Correlation

Simple Correlation

Simple correlation refers to the analysis of the relationship between two variables. If only two variables are considered in a study, it is classified as a simple correlation. For instance, examining how the price of a product influences its demand or supply represents a case of simple correlation.

Multiple Correlation

Multiple correlation involves the simultaneous study of three or more variables to understand their interrelationship. For example, analyzing how rice yield is influenced by both rainfall and fertilizer application is an example of multiple correlation, as it considers multiple factors affecting the outcome.

Partial Correlation

Partial correlation is used when more than two variables are involved, but the focus is on the relationship between only two variables while keeping other influencing variables constant. For example, rice yield depends on both rainfall and fertilizer usage. However, if we assess the correlation between rice yield and rainfall while holding fertilizer use constant, we are applying the concept of partial correlation.

1.6.3 Linear and Non-linear correlation

Linear Correlation

Linear correlation occurs when the rate of change between two variables remains constant. This means that a proportional change in one variable results in a similar proportional change in the other. For instance, if a 10% increase in one variable leads to a 10% increase in another, the relationship is classified as linear correlation.

X	10	15	30	60
Y	50	75	150	300

In this case, the rate of change between X and Y remains consistent. When the data is plotted on graph paper, all the points align along a straight line.

Non-Linear Correlation: In correlation analysis, if a change in one variable does not result in a proportional change in the other variable, it is referred to as non-linear correlation.

X	2	4	6	10	15
Y	8	10	18	22	26

In this case, a change in the value of X does not result in a proportional change in the value of Y. This represents the issue of non-linear correlation. When the data is plotted on a graph, the points do not align in a straight line.

In summary, correlation analysis is a statistical method used to assess the strength and direction of the relationship between two or more variables.

Self-Check Exercise 1.4

Q1. Explain:

i) Positive and Negative Correlation

ii) Simple, Partial, and Multiple Correlation

iii). Linear and Non-linear Correlation

1.7 CORRELATION COEFFICIENT

Correlation analysis is a statistical method used to determine the strength and direction of the relationship between two or more variables. The numerical representation of this relationship is known as the correlation coefficient, which ranges between -1 and +1.

Self-Check Exercise 1.5

Q1. What is meant by Correlation Coefficient

1.8 SIGNIFICANCE OF CORRELATION ANALYSIS

Correlation analysis plays a crucial role in various fields due to the following reasons:

1. **Quantifying Relationships** – It provides a single numerical value to measure the degree of association between variables.
2. **Understanding Economic Behavior** – Correlation analysis helps analyze patterns and trends in economic activities.
3. **Business Decision-Making** – It assists business professionals in estimating costs, prices, and other relevant variables.
4. **Foundation for Regression Analysis** – If two variables exhibit a strong correlation, regression analysis can be used to predict the value of one variable based on the other.
5. **Reducing Uncertainty in Decision-Making** – Predictions based on correlation analysis tend to be more accurate and closer to real-world outcomes.
6. **Testing Significance of Correlation** – The significance of correlation can be determined by comparing the correlation coefficient with six times the probable error (6PE). If the correlation coefficient (r) exceeds 6PE, the correlation is considered significant.

Self-Check Exercise 1.6

Q1. What are the significance of correlation analysis?

1.9 DEGREES OF CORRELATION

Correlation exists in various degrees

1.9.1 Perfect Positive Correlation: A perfect positive correlation occurs when an increase in one variable leads to a proportional increase in another related variable, or when a decrease in one variable results in a proportional decrease in the other. For instance, if a 10% increase in the price of a commodity leads to a 10% increase in its supply, the correlation is perfectly positive. Likewise, if a 5% decrease in price causes a 5% reduction in supply, the correlation remains perfectly positive.

1.9.2 Perfect Negative Correlation: A perfect negative correlation is observed when an increase in one variable leads to a proportional decrease in another related variable, or vice versa. For example, if a 10% rise in the price of a product results in a 10% decline in demand, the correlation is perfectly negative. Similarly, if a 5% drop in price leads to a 5% rise in demand, it also signifies a perfect negative correlation.

1.9.3 Limited Degree of Positive Correlation: A limited degree of positive correlation exists when changes in one variable lead to a non-proportional change in another related variable. For instance, if a 10% increase in the price of a commodity results in only a 5% rise in its supply, the correlation is considered a limited degree of positive correlation. Similarly, if a 10% decrease in price leads to a 5% reduction in supply, it also falls under this category.

1.9.4 Limited Degree of Negative Correlation: A limited degree of negative correlation occurs when an increase in one variable leads to a non-proportional decrease in another variable, or vice versa. For example, if a 10% rise in price causes only a 5% decline in demand, it is classified as a limited degree of negative correlation. Similarly, if a 5% decrease in price results in a 10% increase in demand, the correlation remains limited and negative.

1.9.5 Zero Correlation (No Correlation): Zero correlation, or no correlation, exists when there is no discernible relationship between two variables. In other words, changes in one variable do not correspond to changes in the other. If the values of one variable show no association with the values of another, the correlation is considered nonexistent.

Self-Check Exercise 1.7

Q1. What are the different degrees of correlation?

1.10 METHODS OF MEASURING CORRELATION

Correlation between two variables can be assessed using both graphical and algebraic methods.

1.10.1 Graphic Methods

1.10.1.1 Scatter Diagram Method

1.10.1.2 Correlation Graph Method

1.10.2 Algebraic Methods (Mathematical methods or statistical methods or Co-efficient of correlation methods):

1.10.2.1 Karl Pearson's Co-efficient of Correlation

1.10.2.2 Spearman's Rank Correlation Method

1.10.2.3 Concurrent Deviation Method

Self-Check Exercise 1.8

Q1. What are the various methods of measuring correlation?

1.11 Summary

In this chapter we have gone through the correlation. Different types of correlations, different degrees of correlation and different methods of correlation.

1.12 Glossary

- **Correlation:** Correlation analysis examines the extent of the relationship between two or more variables.
- **Positive Correlation:** When two variables move in the same direction, it is termed a positive correlation. This means that an increase in one variable corresponds with an increase in the other, and a decrease in one results in a decrease in the other.
- **Negative Correlation:** A negative correlation occurs when two variables move in opposite directions. In other words, an increase in one variable is associated with a decrease in the other, and vice versa.
- **Simple Correlation:** When correlation analysis is conducted between only two variables, it is referred to as simple correlation.
- **Multiple Correlation:** When three or more variables are analyzed together to determine their interrelationships, it is called multiple correlation. For instance, studying the effect of both rainfall and fertilizer on rice yield is an example of multiple correlation.
- **Partial Correlation:** In partial correlation, more than two variables are considered, but the focus is on the relationship between one dependent variable and one independent variable while keeping other independent variables constant. For example, if the correlation between rice yield and rainfall is

analyzed while holding the amount of fertilizer constant, it is a case of partial correlation.

- **Linear Correlation:** If the change in one variable leads to a proportional and consistent change in another, the correlation is considered linear.
- **Non-Linear Correlation:** When the change in one variable does not result in a consistent proportional change in another, it is known as non-linear correlation.
- **Perfect Positive Correlation:** This occurs when a change in one variable leads to an exactly proportional change in the same direction in another variable.
- **Perfect Negative Correlation:** A perfect negative correlation exists when a proportional increase in one variable corresponds with an exactly proportional decrease in another variable, and vice versa.
- **Limited Degree of Positive Correlation:** When an increase in one variable results in an increase in another, but not in a proportional manner, it is called a limited degree of positive correlation.
- **Limited Degree of Negative Correlation:** This occurs when an increase in one variable leads to a decrease in another, but not in a strictly proportional way.
- **Zero Correlation (No Correlation):** When there is no identifiable relationship between two variables, it is termed zero correlation. In this case, the changes in one variable do not correspond with changes in the other.

1.13 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 1.1

Ans. Q1. Refer to Section 1.3

Self-Check Exercise 1.2

Ans. Q1. Refer to Section 1.4

Self-Check Exercise 1.3

Ans. Q1. Refer to Section 1.5

Self-Check Exercise 1.4

Ans. Q1. Refer to Section 1.6

Self-Check Exercise 1.5

Ans. Q1. Refer to Section 1.7

Self-Check Exercise 1.6

Ans. Q1. Refer to Section 1.8

Self-Check Exercise 1.7

Ans. Q1. Refer to Section 1.9

Self-Check Exercise 1.8

Ans. Q1. Refer to Section 1.10

1.14 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

1.15 TERMINAL QUESTIONS

Q1. What do you understand by correlation? Explain it significance and different degrees.

Unit-02

CORRELATION ANALYSIS: GRAPHIC METHODS

STRUCTURE

2.1 Introduction

2.2 Learning Objectives

2.3 Graphic Methods of Measuring Correlation

2.3.1 Scatter Diagram Method

2.3.1.1 Merits of Scatter Diagram Method

2.3.1.1 Demerits of Scatter Diagram Method

Self-Check Exercise 2.1

2.3.2 Correlation Graph Method

2.3.2.1 Merits of Correlation Graph Method

2.3.2.2 Demerits of Correlation Graph Method

Self-Check Exercise 2.2

2.4 Summary

2.5 Glossary

2.6 Answers to self-check exercises

2.7 Suggested Reading

2.8 Terminal Questions

2.1 INTRODUCTION

In this last Unit, we have learned about the meaning and significance of correlation, and its different types and degrees. In this unit, we will study about the different graphic methods of measuring correlation.

2.2 LEARNING OBJECTIVES

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

- Explain the various graphic methods of measuring correlation,
- Discuss the merits and demerits of these diagrams

2.3 GRAPHIC METHODS OF MEASURING CORRELATION

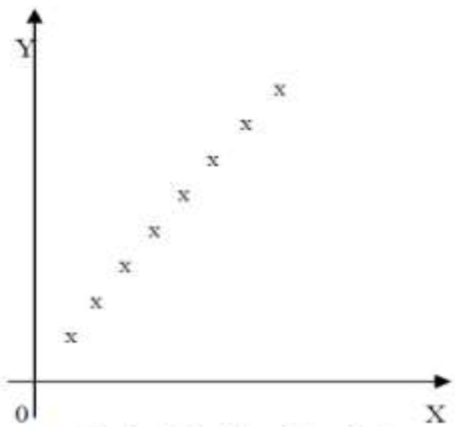
There are the following two graphic methods of measuring correlation:

2.3.1 Scatter Diagram Method

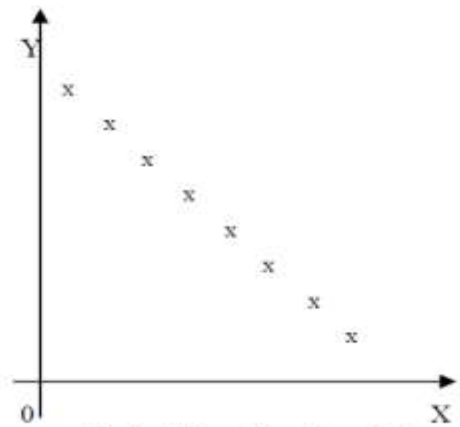
2.3.2 Correlation Graph Method

2.3.1 Scatter Diagram Method

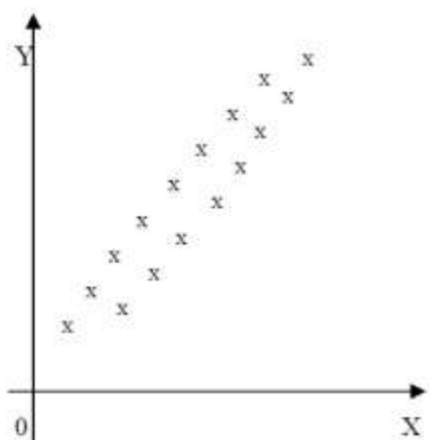
The scatter diagram is the easiest technique to determine the correlation between variables. In this method, all values of the two variables are represented as dots on a chart, which is why it is also referred to as a dot chart. By analyzing the distribution of these dots, we can assess whether there is a relationship between the variables. This diagram not only shows the direction of the correlation but also indicates the strength of the relationship. A wider spread of dots suggests a weaker correlation between the variables.



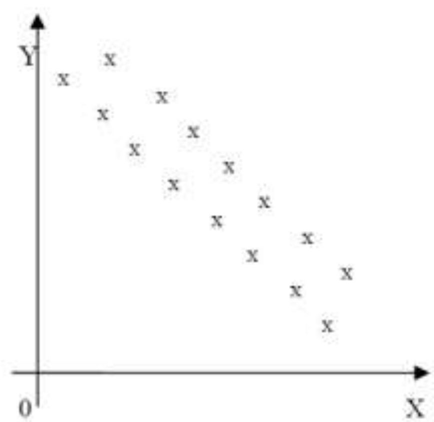
Perfect Positive Correlation



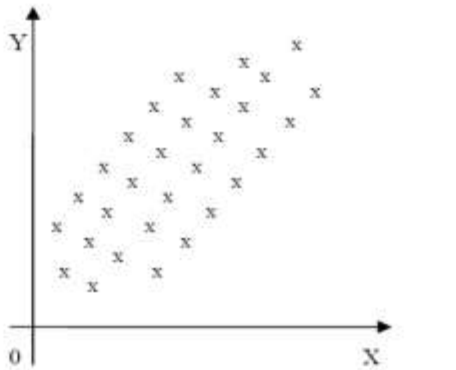
Perfect Negative Correlation



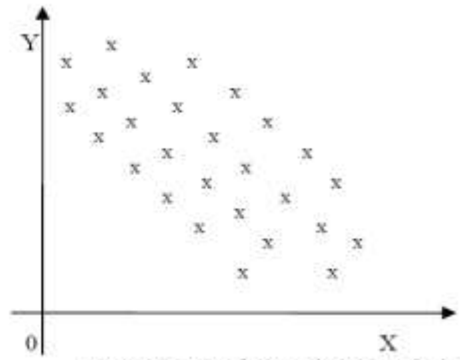
High Degree of Positive Correlation



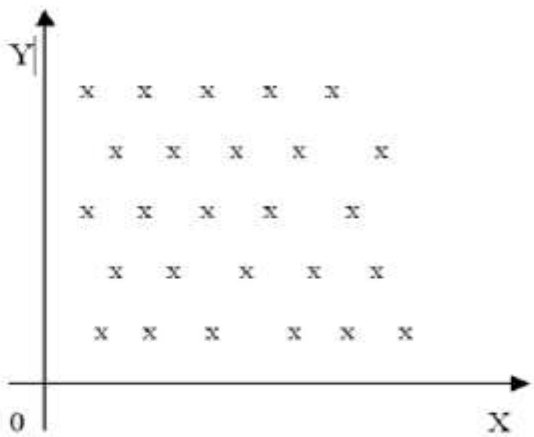
High Degree of Negative Correlation



Low Degree of Positive Correlation



Low Degree of Negative Correlation



No Correlation (r = 0)

2.3.1.1 Merits of Scatter Diagram Method

(i) This method provides a straightforward approach to analyzing the correlation between two variables.

(ii) It does not involve mathematical computations, making it a non-mathematical technique for studying correlations.

(iii) It is easy to interpret, offering even a non-expert a visual representation of the relationship between variables.

(iv) The presence of extreme values does not significantly affect the results.

(v) Creating a scatter diagram is often the initial step in examining the relationship between two variables.

2.3.1.2 Demerits of Scatter Diagram Method

(i) It provides only a general understanding of the correlation rather than precise results.

(ii) This method does not allow for the calculation of the correlation coefficient.

(iii) It does not determine the exact strength or degree of the relationship between the variables.

Self-Check Exercise 2.1

Q1 Describe the Scatter Diagram Method.

Q2 List the advantages and disadvantages of the Scatter Diagram Method.

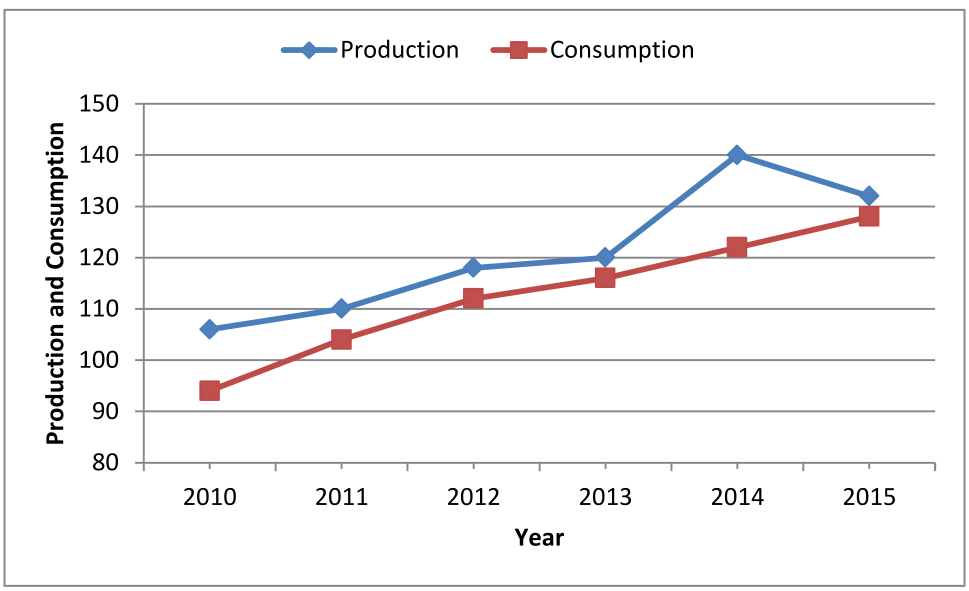
2.3.2 Correlation Graph Method

In the correlation graph method, individual values of two variables are plotted on graph paper. The data points for each variable are then connected separately to form two distinct curves. By observing the direction and proximity of these curves, we can assess whether the variables are related. If both curves move in the same direction (either increasing or decreasing), the correlation is considered positive. Conversely, if the curves move in opposite directions, the correlation is negative.

Example 1: Make a correlation graph on the basis of following data.

Year	2010	2011	2012	2013	2014	2015
Production	106	110	118	120	140	132
Consumption	94	104	112	116	122	128

Solution:



The graph illustrates a strong connection between the two variables—production and construction. Both curves follow the same directional trend, and the gap between them remains nearly constant. This indicates a high degree of positive correlation between the two.

2.3.2.1 Merits of Correlation Graph Method

(i) It is a straightforward approach to analyzing the relationship between variables.

(ii) No mathematical computations are required.

(iii) The method is easy to comprehend.

2.3.2.2 Demerits of Correlation Graph Method

(i) It does not provide a numerical measure of correlation.

(ii) The method is purely a visual representation of variable relationships.

(iii) Determining the precise strength of the relationship is not feasible.

Self-Check Exercise 2.2

Q1. Explain the Correlation Graphic Method.

Q2. Write the merits and demerits of Correlation Graphic Method

2.4 SUMMARY

In this Unit, we have studied about the various methods of measuring the correlation. We have studied in detail the graphic methods of measuring correlation with the help of some diagrams. In the next unit, we will discuss the mathematical methods of measuring correlation analysis.

2.5 GLOSSARY

- **Correlation:** Correlation analysis seeks to assess the extent of the relationship between two or more variables.
- **Perfect Positive Correlation:** A perfect positive correlation occurs when an increase in one variable leads to a proportional increase in the other, or similarly, a decrease in one variable results in a corresponding decrease in the related variable.
- **Perfect Negative Correlation:** A perfect negative correlation is observed when an increase in one variable leads to a proportional decrease in the other, or when a decrease in one variable causes a corresponding increase in the related variable.

2.6 ANSWERS TO SELF-CHECK EXERCISES

Self-Check Exercise 2.1

Ans. Q1. Refer to Section 2.3.1

Ans. Q2. Refer to Section 2.3.1

Self-Check Exercise 2.2

Ans. Q1. Refer to Section 2.3.2

Ans. Q2. Refer to Section 2.3.2

2.7 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.

- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

2.8 TERMINAL QUESTIONS

Q1. What are the different graph methods of measuring correlation? Also explain their merits and demerits.

Unit-03

CORRELATION ANALYSIS: ALGEBRAIC METHODS

STRUCTURE

3.1 Introduction

3.2 Learning Objectives

3.3 Algebraic Methods of Measuring Correlation

3.3.1 Karl Pearson's Co-efficient of Correlation

3.3.1.1 Interpretation of Co-efficient of Correlation

3.3.1.2 Properties of Pearson's Co-efficient of Correlation

3.3.1.3 Computation of Pearson's Co-efficient of Correlation

3.3.1.4 Probable Error and Coefficient of Correlation

3.3.1.5 Coefficient of Determination

3.3.1.6 Merits of Pearson's Coefficient of Correlation

3.3.1.7 Demerits of Pearson's Coefficient of Correlation

Self-check Exercise 3.1

3.3.2 Spearman's Rank Correlation Method

3.3.2.1 Merits of Rank Correlation Method

3.3.2.2 Demerits of Rank Correlation Method

Self-check Exercise 3.2

3.3.3 Concurrent Deviation Method

3.3.3.1 Merits of Concurrent Deviation Method

3.3.3.2 Demerits of Concurrent Deviation Method

Self-check Exercise 3.3

3.4 Summary

3.5 Glossary

3.6 Answers to Self-check Exercises

3.7 References/Suggested Readings

3.8 Terminal Questions

3.1 INTRODUCTION

In this last Unit, we have learned about the graphic methods of measuring correlation. In this unit, we will study about the different algebraic methods of measuring correlation.

3.2 LEARNING OBJECTIVES

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

- Explain the various algebraic methods of measuring correlation,
- Discuss the merits and demerits of these methods.

3.3 ALGEBRAIC METHODS OF MEASURING CORRELATION

Algebraic methods are also known as mathematical methods or statistical methods or co-efficient of correlation methods. The algebraic methods are as follow:

3.3.1 Karl Pearson's Co-efficient of Correlation

3.3.2 Spearman's Rank Correlation Method

3.3.3 Concurrent Deviation Method

3.3.1 Karl Pearson's Co-efficient of Correlation

Karl Pearson's Coefficient of Correlation is one of the most widely used algebraic techniques for assessing correlation. Introduced by Professor Karl Pearson in 1896, it is also known as the product-moment correlation coefficient. This statistical measure, represented by 'r' or r_{xy}, is calculated as the ratio of the covariance between two variables, X and Y, to the product of their standard deviations.

r = Covariance of X and Y

(SD of X) x (SD of Y)

3.3.1.1 Interpretation of Co-efficient of Correlation

The Pearson correlation coefficient always falls within the range of -1 to +1. The following guidelines can be used to interpret its values:

(i) If **r = +1**, there is a perfect positive correlation between the variables.

(ii) If **r = -1**, there is a perfect negative correlation between the variables.

(iii) If **r = 0**, there is no correlation between the variables.

(iv) When **r** is close to **+1**, it indicates a strong positive correlation.

(v) When **r** is close to **-1**, it signifies a strong negative correlation.

(vi) When **r** is near **0**, it suggests a weak or negligible correlation between the variables.

3.3.1.2 Properties of Pearson's Co-efficient of Correlation

(i) The correlation coefficient always falls within the range of -1 to +1 when a relationship exists between variables.

(ii) If there is no correlation, the coefficient is **0** (i.e., r = 0).

(iii) It indicates both the **strength** and **direction** of the relationship between variables.

(iv) It measures correlation but does not establish **causation**.

(v) The Pearson correlation coefficient is equivalent to the **geometric mean** of the two regression coefficients.

i.e. $r = \sqrt{bxy \times byx}$

3.3.1.3 Computation of Pearson's Co-efficient of Correlation:

Pearson's correlation co-efficient can be computed in different ways, which are as follow:

a) Arithmetic Mean Method
b) Assumed Mean Method
c) Direct Method

a) Arithmetic Mean Method:

In the arithmetic mean method, the coefficient of correlation is determined using the actual mean.

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\Sigma(y-\bar{y})^2}}$$

or

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \in y^2}} \text{ whereas x-x-}\bar{x} \text{ and y=y-}\bar{y}$$

Example 1: Calculate Pearson's co-efficient of correlation between age and playing habits of students:

Age	20	21	22	23	24	25
No. of students	500	400	300	240	200	160
Regular players	400	300	180	96	60	24

Let X = Age and Y = Percentage of regular players

Percentage of regular players can be calculated as follows:

$\frac{400}{500} \times 100 = 80$; $\frac{300}{400} \times 100 = 75$; $\frac{180}{300} \times 100 = 60$; $\frac{96}{240} \times 100 = 40$,

$\frac{60}{20} \times 100 = 30$; and $\frac{24}{160} \times 100 = 15$

$$\text{Coefficient of Correlation (r)} = \frac{\Sigma(x-\bar{x}).(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2.(y-\bar{y})^2}}$$

Computation of Pearson's Coefficient of correlation						
Age x	% of Regular Player y	$x-\bar{x}$ (x-22.5)	$(y-\bar{y})$ (y-50)	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
20	80	-2.5	30	-75.0	6.25	900
21	75	-1.5	25	-37.5	2.25	625
22	60	-0.5	10	- 5.0	0.25	100
23	40	0.5	-10	- 5.0	0.25	100
24	30	1.5	-20	-30.0	2.25	400
25	15	2.5	-35	-87.5	6.25	1225
135	300			-240	17.50	3350

$$\bar{x} = \frac{\Sigma x}{N} = \frac{135}{6} = 22.5 \qquad \bar{y} = \frac{\Sigma y}{N} = \frac{300}{6} = 50$$

$$r = \frac{240}{\sqrt{17.5 \times 3350}} = \frac{-240}{\sqrt{58,625}} = \frac{-240}{\sqrt{242.126}} = -0.9912$$

b) Assumed Mean Method:

Under assumed mean method, correlation coefficient is calculated by taking assumed mean only.

$$r = \frac{N\,\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{\sqrt{N\Sigma dx^{2} - (\Sigma dx)^{2}} \times \sqrt{N\Sigma dy^{2} - (\Sigma dy)^{2}}}$$

Where dx = deviations of X from its assumed mean; dy= deviations of y from its assumed mean

Example 2: Determine the correlation coefficient between shoe size and quality defects.

Size	15-16	16-17	17-18	18-19	19-20	20-21
No. of shoes produced:	200	270	340	360	400	300
No. of defectives	150	162	170	180	180	114

Let x = size (i.e. mid-values)

y = percentage of defectives

Therefore, x values are 15.5 , 16.5, 17.5, 18.5, 19.5 and 20.5

y values are 75, 60, 50, 50, 45 and 38

Take assumed mean: x = 17.5 and y = 50

Computation of Pearson's Coefficient of Correlation

x	y	dx	dy	dxdy	dx^2	dy^2
15.5	75	-2	25	-50	4	625
16.5	60	-1	10	-10	1	100
17.5	50	0	0	0	0	0
18.5	50	1	0	0	1	0
19.5	45	2	-5	-10	4	25
20.5	38	3	-12	-36	9	144
		$\Sigma dx = 3$	$\Sigma dy = 18$	$\Sigma dxdy = -106$	$\Sigma dx^2 = 19$	$\Sigma dy^2 = 894$

$$r = \frac{N\,\Sigma dxdy - (\Sigma dx)(\Sigma dy)}{\sqrt{N\Sigma dx^{2-}(\Sigma dx)^2} \times \sqrt{N\Sigma dy^{2-}(\Sigma dy)^2}}$$

$$r = \frac{(6x-106) - (3x18)}{\sqrt{(6\times19)-3^2} \times \sqrt{(6\times894)-18^2}}$$

$$= \frac{-636-54}{\sqrt{114-9} \times \sqrt{5364-324}}$$

$$= \frac{-690}{\sqrt{105} \times \sqrt{5040}} = \frac{-690}{727.46} = -0.9485$$

c) Direct Method:

In the direct method, the coefficient of correlation is determined without using either the actual mean or the assumed mean.

$$r = \frac{N\,\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^{2-}(\Sigma x)^2} \times \sqrt{N\Sigma y^{2-}(\Sigma y)^2}}$$

Example 3:

From the following data, compute Pearson's correlation coefficient:

Price	10	12	14	15	19
Demand (Qty)	40	41	48	60	50

Let us take x as price and y as demand

Computation of Pearson's Coefficient of Correlation				
Price (x)	Demand (y)	xy	x^2	y^2
10	40	400	100	1600
12	41	492	144	1681
14	48	672	196	2304
15	60	900	225	3600
19	50	950	361	2500
$\Sigma x = 70$	$\Sigma y = 239$	$\Sigma xy = 3414$	$\Sigma x^2 = 1026$	$\Sigma y^2 = 11685$

$$r = \frac{N\,\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{N\Sigma x^2 - (\Sigma x)^2} \times \sqrt{N\Sigma y^2 - (\Sigma y)^2}}$$

$$r = \frac{(5\times 3414) - (70 \times 239)}{\sqrt{(5x1026) - 70^2} \times \sqrt{(5\times 11685) - 239^2}}$$

$$r = \frac{17{,}070 - 16{,}730}{\sqrt{230} \times \sqrt{1304}} = \frac{340}{547.65} = {}^{+}0.621$$

3.3.1.4 Probable Error and Coefficient of Correlation

The Probable Error (PE) of the correlation coefficient is a statistical tool used to assess the reliability and consistency of the correlation coefficient's value.

$$\text{Probable Error} = \frac{2}{3}\ \text{standard error}$$

$$= 0.6745 \times \text{standard error}$$

$$\text{Standard Error (SE)} = \frac{1-r^2}{\sqrt{n}}$$

$$\therefore PE = 0.6745 \times \frac{1-r^2}{\sqrt{n}}$$

Example 4:

If the value of co-efficient of correlation is 0.6. find the Probable Error.

$$PE = 0.6745 \times SE$$

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

$$= \frac{1-(0.6)^2}{\sqrt{64}} = \frac{1-0.36}{8} = \frac{0.64}{8} = \underline{0.08}$$

$$P.E = 0.6745 \times 0.08$$

= 0.05396

$$\text{Limits of population Correlation coefficient} = r \pm PE$$

$$= 0.6 \pm 0.05396$$

3.3.1.5 Coefficient of Determination

The coefficient of determination provides a convenient and insightful way to interpret the correlation coefficient. It is derived by squaring the correlation coefficient (r) and is represented as:

Coefficient of Determination = r^2

This value represents the proportion of total variance in the dependent variable that is explained by the independent variable. For instance, if r=0.9 then r^2 = 0.81. This indicates that 81% of the variation in the dependent variable is accounted for by the independent variable, while the remaining 19% remains unexplained. The unexplained portion is referred to as the **coefficient of non-determination**, which is calculated as:

Coefficient of Non-determination (K^2) = 1 – r^2

K^2 = 1- Coefficient of Determination

Example 5: Calculate coefficient of determination and non-determination if coefficient of correlation is 0.8

Solution: r =0.8

Coefficient of Determination = r^2

= 0.82 = 0.64 = 64%

Coefficient of Non-determination = $1 - r^2$

= 1- 0.64 = 0.36 or 36%

Thus, the coefficient of determination is **0.64 (64%)**, meaning 64% of the variation in the dependent variable is explained by the independent variable. The coefficient of non-determination is **0.36 (36%)**, representing the unexplained variation.

3.3.1.6 Merits of Pearson's Coefficient of Correlation

(i) It is one of the most commonly used algebraic methods for measuring the correlation coefficient.

(ii) It quantifies the relationship between variables by providing a numerical value.

(iii) It determines both the strength and direction of the relationship between two variables.

(iv) It allows for further algebraic analysis, including the coefficient of determination and the coefficient of non-determination.

(v) It provides a single value that accurately represents the degree of correlation between two variables.

3.3.1.7 Demerits of Pearson's Coefficient of Correlation

(i) Calculating the correlation coefficient can be complex and time-consuming.

(ii) Understanding the concept and interpretation of the coefficient can be challenging.

(iii) The method involves intricate mathematical computations.

(iv) The process requires a significant amount of time.

(v) The results can be highly sensitive to extreme values (outliers).

(vi) It assumes a linear relationship between variables, which may not always hold in real-world scenarios.

Self-check Exercise 3.1

Q1. Explain Karl Pearson's Co-efficient of Correlation?

Q2. What are merits and demerits of Pearson's Coefficient of Correlation?

Q3. Explain the properties of Pearson's Co-efficient of Correlation.

Q4. How can Probable Error be calculated

Q5. What is Coefficient of Determination?

3.3.2 Spearman's Rank Correlation Method

Pearson's correlation coefficient is used when variables are expressed in numerical form. However, in many instances, measuring variables quantitatively is not feasible due to their qualitative nature. For example, attributes such as beauty, morality, intelligence, and honesty cannot be quantified directly. Nevertheless, these qualitative traits can be arranged in a specific order based on ranking.

When correlation is determined using the ranks of variables rather than their numerical values, it is referred to as rank correlation. This method was introduced by Charles Edward Spearman in 1904.

$$\text{Spearman's coefficient correlation (R)} = 1 - \frac{6\Sigma D^2}{N^{3}-N}$$

Where D = difference of ranks between the two variables

N = number of pairs

Example 6: Determine the rank correlation coefficient between poverty and overcrowding using the data provided below.

Town	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Poverty	17	13	15	16	6	11	14	9	7	12
Over crowing	36	46	35	24	12	18	27	22	2	8

Solution: Since the ranks are not provided, we need to assign them.

Computation of rank correlation Co-efficient						
Town	Poverty	Over crowding	R_1	R_2	D	D^2
A	17	36	1	2	1	1
B	13	46	5	1	4	16
C	15	35	3	3	0	0
D	16	24	2	5	3	9
E	6	12	10	8	2	4
F	11	18	7	7	0	0
G	14	27	4	4	0	0
H	9	22	8	6	2	4
I	7	2	9	10	1	1
J	12	8	6	9	3	9
					ΣD^2	44

$$R = 1 - \frac{6\Sigma D^2}{N^{3-N}}$$

$$R = 1 - \frac{6 \times 44}{10^3 - 10}$$

$$= 1 - \frac{264}{990}$$

$$= 1 - 0.2667$$

$$= +0.7333$$

Computation of Rank Correlation Coefficient when Ranks are Equal

In cases where two or more items share the same rank, the average rank must be assigned to all such items. For instance, if two observations receive the 4th rank, each should be given a rank of 4.5, calculated as (4 + 5) / 2 = 4.5. Similarly, if four observations obtain the 6th rank, they should each be assigned a rank of 7.5, computed as (6 + 7 + 8 + 9) / 4 = 7.5. When dealing with tied ranks, the following formula should be used to determine the rank correlation coefficient:

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \ldots\ldots\ldots\ldots\right]}{N^3 - N}$$

Where D = Difference of rank in the two series

N = Total number of pairs

m = Number of times each rank repeats

Example 7: Find coefficient of rank correlation for the folloing data:

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Since the ranks are not provided, we need to assign them. Further, this is the case of equal ranks.

$$\therefore R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \ldots\ldots\ldots\ldots\right]}{N^3 - N}$$

$$R = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\ldots\ldots\ldots\ldots\right]}{N^3 - N}$$

Computation of rank correlation coefficient					
x	y	R_1	R_2	$D=(R_1-R_2)$	D^2
68	62	4	5	1	1
64	58	6	7	1	1
75	68	2.5	3.5	1	1
50	45	9	10	1	1
54	81	6	1	5	25
80	60	1	6	5	25
75	68	2.5	3.5	1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				ΣD^2	72

$$R=1-\frac{6\left[72+\frac{1}{12}(2^3-2)+\frac{1}{12}(3^3-3)+\frac{1}{12}(2^3-2)\right]}{N^3-N}$$

$$=1-\frac{6\left[72+\frac{1}{12}+2+\frac{1}{12}\right]}{10^3-10}$$

$$=1-\frac{6\times[72+3]}{990}$$

$$=1-\frac{6\times 75}{990}$$

$$=1-\frac{450}{990}=1-0.4545$$

$$=\underline{\underline{0.5455}}$$

3.3.2.1 Merits of Rank Correlation Method

(i) The rank correlation coefficient provides an approximate measure of correlation since it does not rely on actual numerical values for calculation.

(ii) The method is straightforward and easy to understand.

(iii) It can be applied to both quantitative and qualitative data.

(iv) It is particularly useful for analyzing correlations between qualitative attributes such as honesty or beauty.

(v) Since the sum of rank differences between two qualitative datasets always equals zero, this method allows for a quick verification of calculations.

3.3.2.2 Demerits of Rank Correlation Method

(i) The rank correlation coefficient is only an estimate, as it does not involve actual numerical values.

(ii) It becomes cumbersome to use when dealing with a large number of data pairs (N).

(iii) Further algebraic manipulation of the results is not feasible.

(iv) Unlike mean and standard deviation, the rank correlation method does not allow for the calculation of a combined correlation coefficient across multiple datasets.

Self-check Exercise 3.2

Q1. What is Rank Correlation

Q2. Explain the merits and demerits of Spearman’s rank correlation

3.3.3 Concurrent Deviation Method:

The concurrent deviation method is a straightforward approach to measuring correlation. This technique focuses solely on the direction of deviations while disregarding the actual magnitudes of values. As a result, it is particularly useful for analyzing the relationship between two variables in a general sense rather than with precise measurement.

In this method, the correlation's nature is determined by observing the direction of deviations in the variables. If both variables exhibit concurrent deviations, they move in the same direction; otherwise, they move in opposite directions. The coefficient of concurrent deviation is calculated using the following formula:

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

Where N = No. of pairs of symbol

C = No. of concurrent deviations (i.e., No. of + signs in ‘dx dy’ column)

Steps:

1. Each value in the 'X' series is compared with the next value in the sequence. An increase is represented by a '+' sign, while a decrease is denoted by a '-'.

2. The same procedure is applied to the 'Y' series to obtain 'dy'.

3. The values of 'dx' and 'dy' are then multiplied, and the resulting product is recorded in a separate column labeled 'dx dy'.

4. The total count of '+' signs in the 'dx dy' column is determined. These '+' signs represent concurrent deviations and are denoted by 'C'.

5. Apply the formula:

$$r = \pm\sqrt{\pm\left(\frac{2c-N}{N}\right)}$$

Example 8:

Compute coefficient of correlation by concurrent deviation method:

Year:	2003	2004	2005	2006	2007	2008	2009	2010	2011
Supply:	160	164	172	182	166	170	178	192	186
Price:	292	280	260	234	266	254	230	190	200

Solution: Computation of coefficient of correlation by concurrent deviation method:

Supply (x)	Price (y)	dx	dy	dxdy
160	292	+	-	-
164	280	+	-	-
172	260	+	-	-
182	234	+	-	-
166	266	-	+	-
170	254	+	-	-
178	230	+	-	-
192	190	+	-	-
186	200	-	+	-
				C = 0

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

$$= \pm\sqrt{\pm\frac{(2\times 0)-8}{8}}$$

$$= \pm\sqrt{\frac{0-8}{8}} = = \pm\sqrt{\frac{-8}{8}} = \underline{-1}$$

3.3.3.1 Merits of Concurrent Deviation Method:

(i) The calculation of the correlation coefficient is straightforward and requires minimal effort.

(ii) The method is easy to comprehend and apply.

(iii) It is particularly useful for large datasets, allowing for a quick assessment of the relationship between variables.

(iv) This method effectively identifies the nature of correlation, whether positive or negative.

3.3.3.2 Demerits of Concurrent Deviation Method:

(i) The method does not account for the magnitude of changes, treating both small and large variations equally.

(ii) The results derived from this approach provide only a general indication of the presence or absence of correlation.

(iii) Further algebraic manipulations cannot be performed using this method.

(iv) Unlike arithmetic mean and standard deviation, it is not possible to compute a combined coefficient of concurrent deviation for multiple series.

Self-check Exercise 3.3

Q1. What is Concurrent Deviation Method

Q2. Explain the merits and demerits of Concurrent Deviation Method

3.4 SUMMARY

In this chapter we have studied about the spearman's correlation and the method of the computation of correlation by spearman's formula. We have also gone through the merits and demerits of the method.

3.5 GLOSSARY

- **Correlation:** Correlation analysis is used to measure the strength and direction of the relationship between two variables.
- **Karl Pearson's Coefficient of Correlation:** This coefficient, represented as 'r' or rxy, is calculated as the ratio of the covariance between two variables (X and Y) to the product of their standard deviations.
- **Coefficient of Determination:** The coefficient of determination is obtained by squaring the correlation coefficient and indicates the proportion of variance in one variable explained by the other.
- **Concurrent Deviation Method:** This method identifies the nature of correlation based on the direction of deviations in variable values. If the deviations of two variables occur in the same direction, they are said to be concurrent; otherwise, they move in opposite directions.

3.6 ANSWERS TO SELF-CHECK EXERCISES

Self-check Exercise 3.1

Ans. Q1. Refer to Section 3.3.1

Ans. Q2. Refer to Sections 3.3.1.6 and 3.3.1.7

Ans. Q3. Refer to Section 3.3.1.2

Ans. Q4. Refer to Section 3.3.1.4

Ans. Q5. Refer to Section 3.3.1.5

Self-check Exercise 3.2

Ans. Q1. Refer to Section 3.3.2

Ans. Q2. Refer to Section 3.3.2.1 and 3.3.2.2

Self-check Exercise 3.3

Ans. Q1. Refer to Section 3.3.3

Ans. Q2. Refer to Section 3.3.3.1 and 3.3.3.2

3.7 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

3.8 TERMINAL QUESTIONS

Q1. Explain Karl Pearson's Co-efficient of Correlation? Interpret r when r = 1, -1, 0.

Q2. Define rank correlation of coefficient. How is it measured? When is it preferred to Karl Pearson's coefficient of Correlation.

Q3. What is Concurrent Deviation Method. Also explain its merits and demerits.

Unit-04

REGRESSION: MEANING AND SIGNIFICANCE

STRUCTURE

4.1 Introduction
4.2 Learning Objectives
4.3 Regression
    Self-Check Exercise 4.1
4.4 Utility of Regression
    Self-Check Exercise 4.2
4.5 Difference between Regression and Correlation
    Self-Check Exercise 4.3
4.6 Types of Regression
  4.6.1 Simple and Multiple Regression
  4.6.2 Linear and Non- Linear Regression
  4.6.3 Partial and Total Regression
    Self-Check Exercise 4.4
4.7 Simple Linear Regression
  4.7.1 Regression Lines
  4.7.1.1 Nature of Regression Lines
  4.7.1.2 Methods of Obtaining Regression Lines
    Self-Check Exercise 4.5
4.8 Regression Equations
  4.8.1 Regression Equations of X on Y
  4.8.2 Regression Equation of Y on X
    Self-Check Exercise 4.6
4.9 Regression Co-efficient
  4.9.1 Regression Coefficients of X on Y
  4.9.2 Regression Coefficient of Y on X
  4.9.3 Properties of Regression Coefficients
    Self-Check Exercise 4.7
4.10 Summary
4.11 Glossary

4.12 Answers to Self-check Exercises

4.13 References/Suggested Readings

4.14 Terminal Questions

4.1 INTRODUCTION

Decision making is based upon understanding of the relationship between two or more variables. For example, a sales manager might be interested in knowing the impact of advertising on sales. Here advertisement can be considered as an independent variable and sales can be considered as dependent variable. This is an example of a simple linear regression where a single independent variable is used to predict a single numerical dependent variable.

4.2 LEARNING OBJECTIVES

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

- Define regression
- List the utility of regression
- Explain the different types of regression
- Elucidate the difference between Regression & Correlation
- Define the terms like regression line, regression equations
- Explicate the properties of Regression Coefficients

4.3 REGRESSION

Regression, in its original sense, refers to the concept of "returning to the average." The term was first introduced by Sir Francis Galton in 1877 during his study of the heights of fathers and their sons. His research revealed that while taller fathers generally had taller sons and shorter fathers had shorter sons, the sons' average height tended to move closer to the population average. Specifically, the sons of very tall fathers were, on average, shorter than their fathers, while the sons of very short fathers were slightly taller. From this, Galton concluded that extreme traits tend to "regress" toward the mean over generations.

However, in modern statistics, regression has evolved beyond its initial biological context. Today, regression analysis is a widely used statistical method for predicting or estimating the value of one variable based on the known value of another. It is an essential tool across various fields, including natural sciences, social sciences, and physical sciences. In business and economics, regression is particularly valuable for analyzing causal relationships between variables and is extensively used for estimating demand and supply curves, cost functions, production patterns, and consumption trends.

Morris Hambung defines, "The term 'regression analysis refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process".

In the words of M.M. Blair, "Regression is the measure of the average relationship between two or more variables."

According to Taro Yamane, “One of the most frequent used techniques in economics and business research, to find a relation between two or more variables that are related casually, in regression analysis.”

Ya-Lun Chou defines, “Regression analysis attempt to establish the ‘nature of the relationship’ between variables- that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting.”

Self-Check Exercise 4.1

Q1. What do you mean by Regression?

4.4 UTILITY OF REGRESSION

The study if regression is very useful and important in statistical analysis, its importance and utility cannot be ignored in statistics. The following points would make it clear.

(1) Nature of Relationship: Regression analysis explains the nature of relationship between two variables.

(2) Estimation of Relationship: Regression provides estimates of values of the dependent variable from the value of independent variable.

(3) Prediction and Forecasting: By regression analysis, the value of a dependent variable can be predicted on the basis of the value of an independent variable .For example when advertisement expenditure increases ,what will be the probable rise in sales of the firm, this can be predicted by regression.

(4) Useful in Economic and Business Research: Regression analysis is very useful in business and economic research. With the help of regression, business and economic policies can be formulated,

(5) Regression coefficients help to calculate Correlation Coefficients: With the help of Regression Coefficients we can calculate Correlation Coefficients.

Self-Check Exercise 4.2

Q1. What are the utility of Regression?

4.5 DIFFERENCE BETWEEN REGRESSION AND CORRELATION

The main difference between Correlation and Regression is as follows:

(i) **Nature of Relationship**: Correlation examines the relationship between two or more variables, whereas regression analyzes the nature of the relationship between two variables.

(ii) **Cause and Effect**: Correlation does not establish a cause-and-effect relationship between variables; it only measures the degree of association. Regression, however, involves dependent and independent variables, allowing for an analysis of causality. However, the presence of an association does not necessarily indicate causation.

(iii) **Prediction Ability**: Correlation does not facilitate predictions, while regression enables forecasting. Using a regression equation, such as predicting Y based on X, the estimated values of Y can be determined.

(iv) **Key Measure**: The primary measure in correlation is the correlation coefficient, whereas in regression, the key measure is the regression coefficient.

(v) **Effect of Scale and Origin**: The correlation coefficient remains unchanged with alterations in scale and origin. In contrast, regression coefficients are unaffected by changes in origin but are influenced by changes in scale.

(vi) **Symmetric**: In correlation analysis, correlation coefficient (rxy) is the measure of direction and degree of linear relationship between the two variables X and Y. rxy and ryx are symmetrical, i.e., rxy= ryx.In regression analysis, the regression coefficients bxy and byx are not symmetric, i.e., bxy≠byx. Thus, correlation coefficients rxy and ryx are symmetric whereas regression coefficients byx and bxy are not symmetric.

Self-Check Exercise 4.3

Q1. Distinguish between Correlation and Regression

4.6 TYPES OF REGRESSION

Regression can be categorized as one of the following:

4.6.1 Simple and Multiple Regression

4.6.2 Linear and Non- Linear Regression

4.6.3 Partial and Total Regression

4.6.1 Simple and Multiple Regression:

Simple regression is one in which we study only two variables at a time. In it one variable is dependent and the other is independent. The functional relationship between income and expenditure is an example of simple regression. Multiple regression is one in which we study more than two variables at a time in multiple regression one is dependent variable and the others are independent variables. The study of effect of effect of rain and irrigation on yield of wheat is an example of multiple regressions.

4.6.2 Linear and Non- Linear Regression:

When one variable changes with other variable in some fixed ratio, this is called linear regression. Such type of relationship is depicted on the graph by means of a straight line or first degree equation in variables x and y. On the contrary, when one variable varies with other variable in a changing ratio, then it is referred to as non-linear regression. In non- linear regression the curve is not a straight line .The regression equation will be a functional relation between x and y of degree higher than one. i.e., involving terms of the type x2 ,y2,xy,etc.

4.6.3 Partial and Total Regression:

When two or more variables are studied for functional relationship but at a time, relationship between only two variables is studied and other variables are held constant, then it is known as partial regression. On the other hand, in total regression all variables are studied simultaneously for the relationship among them.

Self-Check Exercise 4.4

Q1. Explain the following

(i) Simple and Multiple Regression

(ii) Linear and Non- Linear Regression

(iii) Partial and Total Regression

4.7 SIMPLE LINEAR REGRESSION

Simple linear regression is the most commonly used technique for determining how one variable of interest (the response variable) is affected by changes in another variable (the explanatory variable).Under simple linear regression, concepts like Regression Lines, Regression Coefficients ,Regression Equations are very important we discuss each of them separately below:

4.7.1 Regression Lines

The regression line is a graphical device to describe the average relationship between two variables. It indicates the law of changes in the mean of one variable corresponding to the mean value of the other is shown by the regression line. This is also known as the Line of Best Fit. On the basis of regression line, we can predict the value of a dependent variable on the basis of the given value of the independent variable. If two variables X and Y are given, then there are two regression lines related to them as follows.

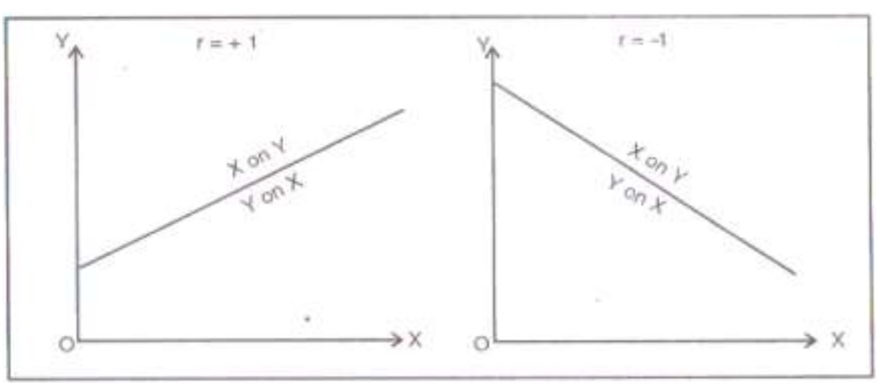
Regression Line of X on Y: The regression line of X on Y gives the best estimate for the value of X for any given value of Y.

Regression Line Y on X: The regression line of Y on X gives the best estimate for the value of Y for any given value of X.

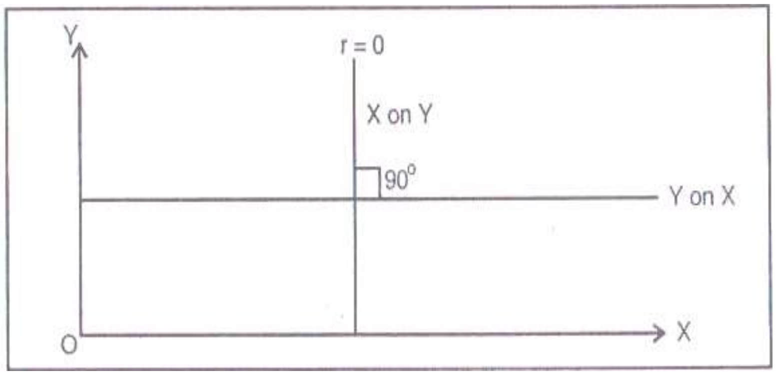
4.7.1.1 Nature of Regression Lines

With the help of the direction and magnitude of correlation, the nature of regression lines can be known. The main points are as follows:

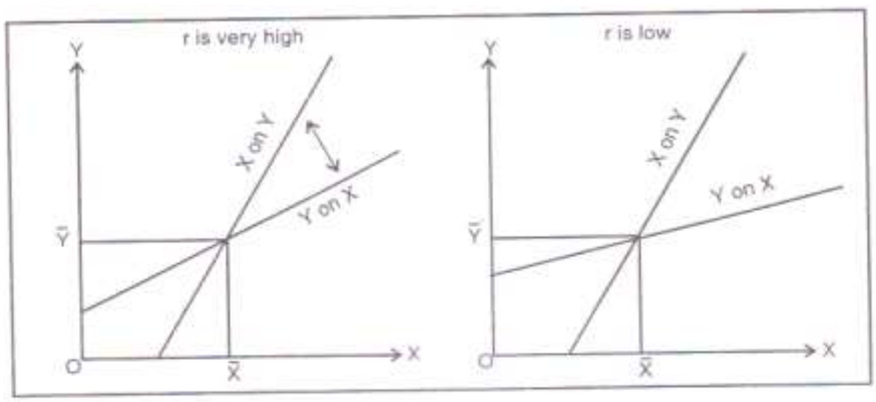
1. Two regression lines are coincident or there will be only one regression line if r =$\pm$ 1, i.e., there is perfect correlation. This is clear from the following diagrams:



2. The two regression lines intersect each other at 900 if r=0.This is clear from the diagram given below.

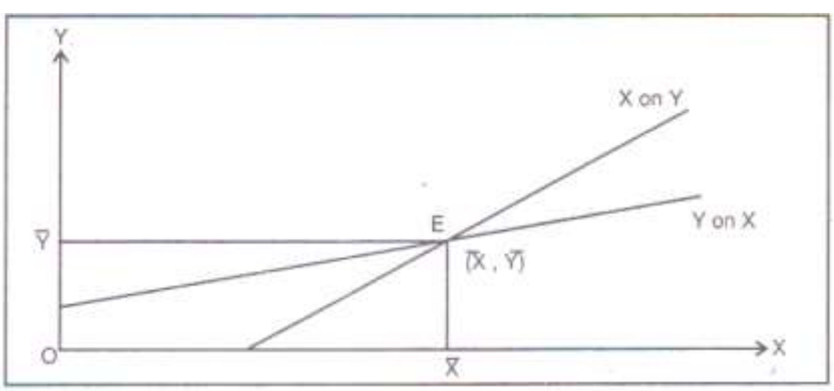


3. The nearer the regression lines are to each other, the greater will be the degree of correlation. On the contrary, the greater the distance between the two regression lines, the lesser will be the degree of correlation. This is clear from the diagram given below.



4. If regression lines rise from left to right upward, then the correlation is positive. On the other side, if these line move from right to left, then correlation is negative.

5. The regression lines cut each other at the point of intersection of $\overline{X}$ and $\overline{Y}$. This is clear from the diagram given below.



.

4.7.1.2 Methods of Obtaining Regression Lines

There are two important methods of obtaining regression lines:

(a) Scatter Diagram Method

(b) Least Square Method

Both the methods are discussed below:

(a) Scatter Diagram Method:

In scatter diagram method, values of the related variables are plotted on a graph. A straight line is drawn passing through the plotted points with free hands. The shape of the regression line can be linear or non-linear, this depends upon the location of points. This method is used very rarely in practice because in this method the decision of the person who draws the regression lines affects the results.

(b) Least Square Method:

The least squares method is a widely used technique for obtaining approximate solutions to over determined systems, where the number of equations exceeds the number of unknowns. This approach aims to minimize the sum of the squared differences between observed and estimated values. In regression analysis, a line is fitted through data points in such a way that the total squared deviations between the actual values and the predicted values on the line are minimized. This fitted line is known as the Line of Best Fit. The optimal fit, as determined by the least squares criterion, ensures that the sum of squared residuals—representing the discrepancies between observed and predicted values—is as small as possible.

Self-Check Exercise 4.5

Q1. Why are there two lines of regression?

Q2. Explain the nature of regression lines.

4.8 REGRESSION EQUATIONS

Regression equations represent the mathematical expressions of regression lines. These equations serve as tools for predicting the average value of one variable based on a given value of another. As there are two regression lines, there are also two corresponding regression equations, which are as follows:

4.8.1 Regression Equations of X on Y: This equation is used to estimate the probable values of X with the given values of Y. The equation is expressed as:

$$X = a + bY$$

Here, *a* and *b* are constants

Regression equations of X on Y can also be presented as:

$$X - \bar{X} = r.\frac{\sigma x}{\sigma y}(Y - \bar{Y})$$

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

Here, b_{xy} =Regression coefficient of X on Y

4.8.2 Regression Equation of Y on X: This equation is used to estimate the probable values of Y on the basis of the given values of X. The equation is expressed as:

Y= a+bX

Here, a and *b* are constants

Regression equations of Y on X can also be presented as:

$$Y - \bar{Y} = r.\frac{\sigma y}{\sigma x}(X - \bar{X})$$

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

Here, byx= Regression coefficient of Y on X

Self-Check Exercise 4.6

Q1. What are regression equations

4.9 REGRESSION COEFFICIENTS

A regression coefficient indicates the value by which one variable changes for a unit change in the other variable. Regression coefficient, in fact represents the slope of a regression line. There are two regression coefficients for two correlated series X and Y.

4.9.1 Regression Coefficients of Y on X: This coefficient shows that with a unit change in the value of X variable, what will be the average change in the value of Y variable. This is represented by b_{yx}. Its formula is as follows.

$$b_{yx} = r.\frac{\sigma y}{\sigma x}$$

4.9.2 Regression Coefficient of X on Y: This coefficient shows that with a unit change in the value of Y variable, what will be the average change in the value of X variable. This is represented by b_{xy}. Its formula is as follows.

$$b_{xy} = r.\frac{\sigma x}{\sigma y}$$

4.9.3 Properties of Regression Coefficients

The main properties of regression coefficients are as follows:

(1) **Coefficient of correlation is the geometric mean of the regression coefficients**, i.e.

$$\mathbf{r} = \sqrt{\boldsymbol{bxy} \times \boldsymbol{byx}}$$

This property can be proved in the following manner:

Regression coefficient of X on Y

b_{xy} = r. $\frac{\sigma x}{\sigma y}$ ……………(i)

Regression coefficient of X on Y

b_{yx} = r. $\frac{\sigma y}{\sigma x}$ ……………..(ii)

Multiplying (i) and (ii)

$$b_{xy} \times b_{yx} = r.\frac{\sigma x}{\sigma y} \times r.\frac{\sigma y}{\sigma x}$$

$$r^2 = b_{xy} \times b_{yx}$$

Hence, r = $\pm \sqrt{bxy \times byx}$

(**2) Both the regression coefficients must have the same algebraic signs:** Both regression coefficients will either have positive signs or negative signs. In other words, when one regression coefficient is negative, the other would also be negative .It is never possible that one regression coefficient is negative while the other is positive.

(3) The coefficient of correlation will have the same sign as that of regression coefficients: If both regression coefficients are negative, then the correlation coefficient would be negative. And if b_{yx} and b_{xy} have positive signs, then r will also take positive sign.

(4) Both the regression coefficient cannot be greater than unity: If one regression coefficient of y on x is greater than unity, then the regression coefficient of x on y must be less than unity. If both regression coefficients happen to be greater than one then their geometric mean will exceed one which will not give the correlation coefficients whose value never exceeds one.

r = $\sqrt{bxy \times byx}$ r = ± 1

(5) Arithmetic mean of two regression coefficients is either equal to or greater than the correlation coefficient: In terms of the formula:

$$\frac{byx+bxy}{2} \geq r$$

(6) Shift of origin does not affect regression coefficients but shift in scale does affect regression coefficients: Regression coefficients are independent of the change of origin but not of scale. This means if some common factor is taken out from the items of the series, then in that case, we will have to make adjustments in the regression coefficient formula which is as follows:

$$byx = bvu \cdot \frac{i_y}{i_x} \quad \text{and} \quad bxy = buv \cdot \frac{i_x}{i_y}$$

Where, $u = \frac{X-a}{h}$ and $v = \frac{Y-b}{k}$ and

i_y and i_x are common factors of Y and X series respectively.

Self-Check Exercise 4.7

Q1. What is meant by Regression Coefficients

Q2. Explain the properties of Regression Coefficients

4.10 SUMMARY

In this unit, we have confer the concept of regression in detail, starting with the meaning of regression, Its utility we have moved on to its various types, and have dealt in detail with regression lines , regression equations .We have also talked about regression coefficients and their properties.

4.11 GLOSSARY

- **Regression:** the statistical method which helps us to estimate or predict the unknown value of one variable from the known value of the related variable is called regression.
- **Regression Equations:** Regression equations are the algebraic formulation of regression lines. The regression equations may be regarded as expressions for estimating from a given value of one variable the average corresponding value of the other.
- **Regression Coefficient :** A regression coefficient indicates the value by which one variable changes for a unit change in the other variable. Regression coefficient, in fact represents the slope of a regression line
- **Simple Regression:** Simple regression is one in we study only two variables at a time. In it one variable is dependent and the another is independent. The functional relationship between income and expenditure is an example of simple regression.
- **Multiple Regression:** Multiple regression is one in which we study more than two variables at a time in multiple regression one is dependent variable and the others are independent variables. The study of effect of effect of rain and irrigation on yield of wheat is an example of multiple regressions.
- **Linear Regression:** When one variable changes with other variable in some fixed ratio, this is called linear regression. Such type of relationship is depicted on the graph by means of a straight line or first degree equation in variables x and y.
- **Non-Linear Regression:** On the contrary, when one variable varies with other variable in a changing ratio, then it is referred to as non-linear regression. In non-linear regression the curve is not a straight line .The regression equation will be a functional relation between x and y of degree higher than one i.e., involving terms of the type x^2 ,y^2,xy,etc.
- **Partial Regression:** When two or more variables are studied for functional relationship but at a time, relationship between only two variables is studied and other variables are held constant ,then it is known as partial regression.
- **Total Regression**: On the other hand, in total regression all variables are studied simultaneously for the relationship among them.

4.12 ANSWERS TO SELF-CHECK EXERCISES

Self-Check Exercise 4.1

Ans. Q1. Refer to Section 4.3

Self-Check Exercise 4.2

Ans. Q1. Refer to Section 4.4

Self-Check Exercise 4.3

Ans. Q1. Refer to Section 4.5

Self-Check Exercise 4.4

Ans. Q1 (i). Refer to Section 4.6.1

Ans. Q1(ii). Refer to Section 4.6.2

Ans. Q1(iii). Refer to Section 4.6.3

Self-Check Exercise 4.5

Ans. Q1. Refer to Section 4.7.1

Ans. Q2. Refer to Section 4.7.1.1

Self-Check Exercise 4.6

Ans. Q1. Refer to Section 4.8

Self-Check Exercise 4.7

Ans. Q1. Refer to Section 4.9

Ans. Q2. Refer to Section 4.9.3

4.13 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

4.14 TERMINAL QUESTIONS

Q1. Distinguish between Correlation and Regression. Why are there two lines of regression?

Q2. What are regression coefficients? Explain the properties of regression coefficients.

Q3. From the following data obtain the two regression equations.

Sales	91	97	108	121	67	124	51	73	111	57
Purchases	71	75	69	97	70	91	39	61	80	47

(Ans: X=1.36Y-52, Y=0.61X+15.1)

Q4. What would be the lines if r= +1, r= -1, r=0. Give interpretation in each case.

Unit-05

REGRESSION: OBTAIN REGRESSION EQUATIONS

STRUCTURE

5.1 Introduction

5.2 Learning Objectives

5.3 Obtain Regression Equations

5.3.1 Regression Equations using Normal Equations

5.3.2 Regression Equations using Regression Coefficients

5.4 Summary

5.5 Glossary

5.6 Answers to Self-check Exercises

5.7 References/Suggested Readings

5.8 Terminal Questions

5.1 INTRODUCTION

In this last unit, we have learnt about regression. Different types of regression and utility of regression. We talked about regression coefficients, its properties and dealt with terms like regression lines, regression equations etc. In this unit, we will continue with our discussion on regression analysis and will study the different method of measuring regression equations.

5.2 Learning Objectives

After going through this unit you will be able to:

- Explicate the method of obtaining regression equations
- Use simple linear regression equation

5.3 Obtain Regression Equations

5.3.1 Regression Equations using Normal Equations

5.3.2 Regression Equations using Regression Coefficients

5.3.1 Regression Equations using Normal Equations:

This approach is known as the Least Squares Method. In this method, regression equations are derived by solving two normal equations. These equations include the regression of Y on X and the regression of X on Y.

Regression Equation of Yon X

Regression Equation of Y on X is expressed as follows:

Y=a+bx

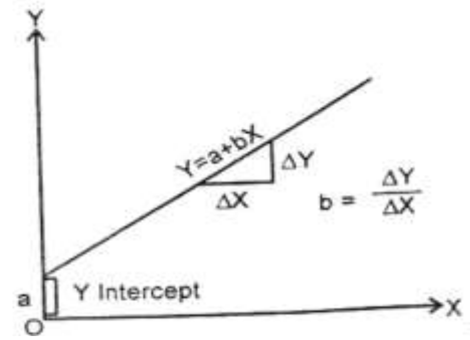
Where, Y=Dependent variable, X= Independent variable

a=y-Intercept, *b*=Slope of the line.

Under lest square method, the values of a and b are obtained by using the following two normal equations:

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

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



Solving these equations, we get the following value of a and b.

$$b_{yx} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2}$$

$$a = \overline{Y} - b\overline{X}$$

Finally the calculated value of a and b is put in the equation Y=a+bX. The regression equation of *Y* on *X* will be used to estimate the value of *Y* when the value of *X* is given.

Note: a is the *Y* –intercept, which indicates the minimum value of *Y* for *X*=0 and b is the slope of the line or called regression coefficient of *Y* on *X*, which indicates the absolute increase in *Y* for a unit increase in *X*.

Regression Equation of X on Y

Regression Equation of X on y is expressed as follows:

$$X = a_0 + b_0 Y$$

Where, *X*=Dependent Variable, *Y*= Independent Variable, a_0=X-intercept,b_0= Slope of the line.

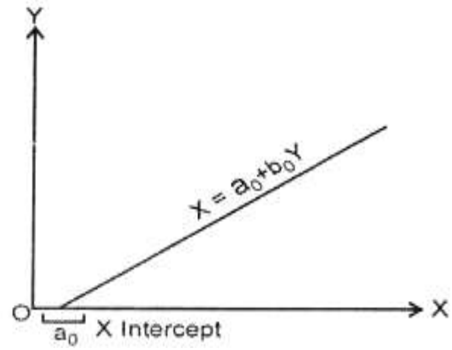
Under least square method, the value of a_0 and b_0 are obtained by using the following two normal equations:

$$\sum X = Na_0 + b_0\sum Y$$

$$\sum XY = a_0\sum Y + b_0\sum Y^2$$

Solving these equations, we get the following value of a_0 and b_0

$$b_0 = b_{xy} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2}$$
$$a_0 = \overline{X} - b_0 \overline{Y}$$



Finally, the calculated value of a_0 and b_0 are put in the equation $X=a_0+b_0Y$. The regression equation of *X* on *Y* will be used to estimate the value of *X* when the value of *Y* is given.

Note: a_0 is the X-intercept, which indicates the minimum value of X for Y=0 and b_0 is the slope of the line or called regression coefficient of X on Y.

The following examples make the above said method more clear:

Example 1:

Compute the appropriate regression equation for the following data:

Independent Variable (X)	2	4	5	6	8	11
Dependent Variable (Y)	18	12	10	8	7	5

Solution : The appropriate regression equation in the given case is that of Y on X, viz., Y=a+ bX, since Y is the dependent variable and X is the independent variable. For finding values of a and b, we use the two normal equations:

$$\sum Y = Na + b\sum X \qquad \ldots(i)$$
$$\sum XY = a\sum X + b\sum X^2 \qquad \ldots(ii)$$

Computation of Regression Equation Y on X

X	Y	X^2	XY	$(X-\overline{X})$	$(Y-\overline{Y})$	$(X-\overline{X})^2$	$(X-\overline{X})(Y-\overline{Y})$
2	18	4	36	−4	1	16	−32
4	12	16	48	−2	2	4	−4
5	10	25	50	−1	0	1	0
6	8	36	48	0	−2	0	0
8	7	64	56	2	−3	4	−6
11	5	121	55	5	−5	25	−25
$\Sigma X = 36$	$\Sigma Y = 60$	$\Sigma X^2 = 266$	$\Sigma XY = 293$			$\Sigma(X-\overline{X})^2 = 50$	$\Sigma(X-\overline{X})(Y-\overline{Y}) = -67$

Substituting the values in **(i)** and **(ii)** we have

$$60 = 6a + 36b$$

$$293 = 36a + 266b$$

Solving the two equations for a for b, we get

$$a = 18.04 \qquad \text{and} \qquad b = -1.34$$

Hence the required equation is :

$$Y = 18.04 - 1.34\,X$$

Aliter. The regression equation of Y on X is :

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

where $\bar{X} = \frac{\Sigma X}{N} = \frac{36}{6} = 6, \quad \bar{Y} = \frac{\Sigma Y}{N} = \frac{60}{6} = 10,$

$$b_{yx} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} = \frac{-67}{50} = -1.34$$

∴ The required regression equation is :

$Y - 10 = -1.34\,(X - 6)$ Hence $Y = 18.04 - 1.34\,X$.

Example 2:

By using the following data, find out the two lines of regression and from them compute the Karl Pearson's coefficient of correlation.

$\sum X = 250$; $\sum Y = 300$; $\sum XY = 7{,}900$; $\sum X^2 = 6500$; $\sum Y^2 = 10{,}000$; and $N = 10$

Solution: We have :

$\bar{X} = \frac{\sum X}{N} = \frac{250}{10} = 25$; $\bar{Y} = \frac{\sum Y}{N} = \frac{300}{10} = 30$

byx = Coefficient of regression of Y on X $= \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$

$$= \frac{10 \times 7900 - 250 \times 300}{10 \times 6500 - (250)^2}$$

$$= \frac{79000 - 75000}{65000 - 62500}$$

$$= \frac{4000}{2500}$$

$$= 1.6$$

bxy = Coefficient of regression of X on Y $= \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum Y^2 - (\sum Y)^2}$

$$= \frac{10 \times 7900 - 250 \times 300}{10 \times 10000 - (300)^2}$$

$$= \frac{79000 - 75000}{100000 - 90000} = \frac{4000}{10000} = 0.4$$

Hence, correlation coefficient rxy between X and Y is given by:

r_{xy}^2 = byx . bxy

= 1.6 x0.4 = 0.64

rxy =$\pm\sqrt{0.64}$

= $\pm$ 0.8

Since the regression coefficients are positive , we take r =$\pm$ 0.8

Regression Equation

Regression Equation Y on X	*Regression Equation X on Y*
$Y - \bar{Y} = byx\ (X - \bar{X})$	$X - \bar{X} = bxy\ (Y - \bar{Y})$
Y – 30 = 1.6 (X – 25)	*X – 25 = 0.4 (Y – 30)*
Y-1.6X – 40 +30	*X = 0.4 Y – 12 + 25*
Y = 1.6 X – 10	*X = 0.4 Y + 13*

5.3.2 Regression Equations using Regression Coefficients

Regression equations can also be computed with the help of regression coefficients. For this, we will have to find out, byx and bxy from the given data. Regression equations can be computed from the regression coefficients by using any of the following methods:

(a) Using the actual values of X and Y series

(b) Using deviations from Actual Means

(c) Using Deviations from Assumed Mean

(d) Using r,σ_x. σ_y and ………

(a) Using the actual values of X and Y series:

In this method, actual values of X and Y are used to determine regression equations. With regard to regression coefficients, regression equations are put in the following ways:

Regression Equation Y on X

$$Y - \overline{Y} = byx\,(X - \overline{X})$$

$$\text{or} \quad Y = \overline{Y} + byx\,(X - \overline{X})$$

Here, $\overline{X}$ = Arithmetic mean of X series $= \frac{\Sigma X}{N}$

$\overline{Y}$ = Arithmetic mean of Y series $= \frac{\Sigma Y}{N}$

byx = Regression coefficient of Y on X

Using actual values, the value of *byx* can be calculated as:

$$\mathbf{byx} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma X^2 - (\Sigma X)^2} \quad \text{or} \quad \mathbf{byx} = \frac{\Sigma XY / N - \overline{X}.\overline{Y}}{\sigma_x^2}$$

Regression Equation X on Y

$$X - \overline{X} = bxy\,(Y - \overline{Y})$$

$$\text{or} \quad X = \overline{X} + bxy(Y - \overline{Y})$$

Where *bxy*= Regression coefficient of X on Y.

Using actual values, the value of *bxy* can be calculated as:

$$\mathbf{bxy} = \frac{N.\Sigma XY - \Sigma X.\Sigma Y}{N.\Sigma Y^2 - (\Sigma Y)^2} \quad \text{or} \quad \mathbf{bxy} = \frac{\Sigma XY / N - \overline{X}.\overline{Y}}{\sigma_y^2} = \frac{Cov\,(X_1 Y)}{\sigma_y^2}$$

Example 3:

Given that:$\sum X=120$, $\sum Y=432$, $\sum XY=4992$, $\sum X^2=1392$, $\sum y^2=18252$,n=12.

Find out :(i) the two regression coefficients; (ii) the two regression equations, and (iii) the coefficient of correlation between X and Y.

Solution: Given : $n = 12, \Sigma X = 120, \Sigma X^2 = 1392, \Sigma Y = 432, \Sigma Y^2 = 18252, \Sigma XY = 4992.$

(*i*) Regression coefficient of Y on X is:

$$b_{yx} = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = \frac{12(4992) - (120)(432)}{12(1392) - (120)^2} = \frac{12[4992 - 4320]}{12[1392 - 1200]} = \frac{672}{192} = 3.5$$

Regression coefficient of X on Y is:

$$b_{xy} = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma Y^2 - (\Sigma Y)^2} = \frac{12(4992) - (120)(432)}{12(18252) - (432)^2} = \frac{12[4992 - 4320]}{12[18252 - 15552]} = \frac{672}{2698} = 0.249$$

(*ii*) $\bar{X} = \frac{\Sigma X}{n} = \frac{120}{12} = 10$ and $\bar{Y} = \frac{\Sigma Y}{n} = \frac{432}{12} = 36$

Regression Equations

Regression Equation of Y on X is:

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

$$\Rightarrow Y - 36 = 3.5(X - 10)$$

$$\Rightarrow Y - 36 = 3.5X - 35$$

$$\Rightarrow Y = 3.5X + 1$$

Regression Equation of X on Y is:

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

$$\Rightarrow X - 10 = 0.249(Y - 36)$$

$$\Rightarrow X - 10 = 0.249X - 8.964$$

$$\Rightarrow X = 0.249X + 1.036$$

(*iii*) $r^2 = b_{yx} \cdot b_{xy} = (3.5)(0.249) = 0.8715 \Rightarrow r = \sqrt{0.8715} = 0.934$

(b) Using Deviations taken from Actual Means:

When the size of the values of X and Y is very large, then the method using actual values becomes very difficult to use. In such cases, in place of actual values, deviations taken from arithmetic means are used. In such a case, regression equations are expressed as follows:

Regression Equation Y on X

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

$$\text{Or } Y=\bar{Y} + byx\,(X-\bar{X})$$

Here, $\bar{X}$=Arithmetic Mean of X

$\bar{Y}$= Arithmetic Mean of Y

byx=Regression coefficient Y on X

The value of *byx* can be calculated as:

$$\boldsymbol{byx} = \frac{\sum xy}{\sum x^2}$$

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

Regression Equation X on Y

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

$$\text{Or } X=\bar{X}+bxy\,(Y-\bar{Y})$$

Where, *bxy* = Regression coefficient X on Y

The value of *bxy* can be calculated as:

$$\boldsymbol{bxy} = \frac{\sum yx}{\sum y^2}$$

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

The following example will make it clear:

Example 4:

In the estimation of regression equation of two variable X and Y, the following results were obtained:

$\sum X=900$, $\sum Y=700$, $n=10$, $\sum x^2=6360$, $\sum y^2=2860$ and $\sum xy=3900$.

Obtain two regression equations.

Solution: Regression Coefficients

Regression Coefficient of Y on X is:

$$byx = \frac{\sum xy}{\sum x^2} = \frac{3900}{6360} = 0.6132$$

Regression Coefficient of X on Y is:

$$bxy = \frac{\sum yx}{\sum y^2} = \frac{3900}{2860} = 1.3636$$

$$\text{Also } \bar{X}=\frac{\sum X}{n}=\frac{900}{10}=90 \text{ and } \bar{Y}=\frac{\sum Y}{n}=\frac{700}{10}=70$$

Regression Equations:

Regression Equation Y on X is:

$$Y-\bar{Y} = byx\ (X-\bar{X})$$

$$=Y-70=0.6132(X-90)$$

$$=Y=70+0.6132X-55.188$$

$$=Y_c=0.6132X-14.812$$

Regression Equation X on Y is:

$$X-\bar{X} = bxy\ (Y-\bar{Y})$$

$$=X-90=1.3636(Y-70)$$

$$=X=90+1.3636Y-95.452$$

$$=X_c=1.3636Y-5.452$$

(c) Using Deviations from Assumed Means:

When actual means turn out to be in fractions rather than whole numbers like 31.65, 72.21, etc., then it becomes difficult to take deviations from actual means and squaring them up. To avoid such difficulty, deviations from assumed means rather than actual mean are used. In such cases, regression equations are expressed as follows:

Regression Equation Y on X

$$Y-\bar{Y} = byx\ (X-\bar{X})$$

Here, *byx*= Regression coefficient of Y on X.

Using deviations from assumed means, the value of *byx* can be calculated as:

$$byx = \frac{N \times \Sigma dxdy - \Sigma dx\ .\ \Sigma dy}{N\ .\ \Sigma dx^2 - (\Sigma dx)^2}$$

or

$$byx = \frac{\Sigma dxdy - \frac{\Sigma dx\ .\ \Sigma dy}{N}}{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}}$$

Where, $dx=X-A_x$, $dy=Y-A_y$

Regression Equation X on Y

$$X-\bar{X} = bxy\ (Y-\bar{Y})$$

Where, *bxy*=Regression coefficient X on Y.

Using deviations from assumed means, the value of bxy can be calculated as:

$$bxy = \frac{N \times \Sigma dxdy - \Sigma dx \, . \, \Sigma dy}{N \, . \, \Sigma dy^2 - (\Sigma dy)^2}$$

or

$$bxy = \frac{\Sigma dxdy - \frac{\Sigma dx \, . \, \Sigma dy}{N}}{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}$$

Where, $dx=X-A_x$, $dy=Y-A_y$

The following examples will clarify this method:

Example 5:

Obtain the lines of regression from the following data

X	4	5	6	8	11
Y	12	10	8	7	5

Verify that the coefficient of correlation is the geometric mean of the two regression coefficients.

Solution:

X	Y	dx (X-6)	dx^2	dy (Y-8)	dy^2	dxdy
4	12	-2	4	4	16	-8
5	10	-1	1	2	4	-2
6	8	0	0	0	0	0
8	7	2	4	-1	1	-2
11	5	5	25	-3	9	-15
∑X=34	**∑Y=42**	**∑dx=4**	**$\sum dx^2=34$**	**∑dy=2**	**$\sum dy^2=30$**	**∑dxdy=-27**

$\bar{X}=\frac{\sum X}{N}=\frac{35}{5}=6.8$ $\bar{Y}=\frac{\sum Y}{N}=\frac{42}{5}=8.4$

Regression equation of X on Y

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

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

$$= \frac{5(-27)-(4)(2)}{5(30)-(2)^2} = \frac{-135-8}{150-4} = -0.979$$

$$X - 6.8 = -0.979(Y - 8.4)$$

$$X - 6.8 = -0.979Y + 8.4$$

$$X = -0.979Y + 15.2$$

Regression equation of Y on X

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

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

$$= \frac{5(-27)-4(2)}{5(34)-(4)^2} = \frac{-135-8}{170-16} = -0.929$$

$$Y - 8.4 = -0.929(X - 6.8)$$

$$Y - 8.4 = -0.929X + 6.317$$

$$Y = -0.929X + 14.717$$

Correlation coefficient

$$r = \frac{N\,\Sigma dxdy - \Sigma dx\,\Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\,\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{5(-27) - 4\times 2}{\sqrt{5\times 34 - (4)^2}\,\sqrt{5\times 30 - (2)^2}}$$

$$= \frac{-135-8}{\sqrt{170-16}\sqrt{150-4}} = \frac{-143}{\sqrt{154\times 146}}$$

$$= \frac{-143}{149.947} = -0.954$$

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

$$= -\sqrt{0.979\times 0.929} \qquad \text{(Since } b_{xy} \text{ and } b_{yx} \text{ are negative)}$$

$$= -0.954$$

Thus, the coefficient of correlation is the geometric mean of two coefficients of regression.

(c) To find regression Equations from Coefficient of Correlation. Standard Deviations and Arithmetic Mean of X and Y:

When the values of ….. and….. r,σ_x. σ_y are given, then regression equations are expressed as following :

(1) Regression Equation of Y on X

$Y-\bar{Y}$ =byx ($X-\bar{X}$) *where, byx=r. $\frac{\sigma_y}{\sigma_x}$*

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

(2)Regression Equation of X on Y

$X-\bar{X}$ =bxy ($Y-\bar{Y}$) *where, bxy=r. $\frac{\sigma_x}{\sigma_y}$*

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

Example 6:

You are given below the following information about advertising and sales

	Adv. Exp (X) (Rs. Lakhs)	**Sales (Y) (Rs. Lakhs)**
Mean	10	90
S.D.	3	12
Correlation Coefficient	0.8	

(i) Determine the equations of the two regression lines.

(ii) Estimate the expected sales when the advertisement expenditure is Rs.15 lakhs.

(iii) Calculate the required advertisement expenditure to achieve a sales target of Rs. 120 lakhs.

Solution

(i) **Regression equation of *X* on *Y***

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

$$b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$r = 0.8, \bar{X} = 10, \bar{Y} = 90, \sigma_x = 3, \sigma_y = 12$$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} = 0.8\times\frac{3}{12} = 0.2$$

$$X-10 = 0.2(Y-90)$$

$$X = 0.2Y-18+10$$

$$X = 0.2Y-8$$

Regression equation of Y on X

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

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.8 \times \frac{12}{3} = 3.2$$

$$Y - 90 = 3.2(X - 10)$$

$$Y = 3.2X - 32 + 90$$

$$Y = 3.2X + 58$$

(ii) When advertisement expenditure is Rs. 15 lakhs, the sales is

$$Y = 3.2(15) + 58 = 46 + 58 = 106$$

Sales = 106 lakhs.

(iii) When sales target is 120 lakhs, the advertisement expenditure is

$$X = 0.2(120) - 8 = 24 - 8 = 16$$

Advertisement expenditure is 16 lakhs.

Self-Check Exercise 5.1

Q1. How would you identify regression equation on X on Y and Y on X.

Q2. Obtain the two regression equations from the following data:

Independent Variable (X)	2	4	6	8	10	12
Dependent Variable (Y)	4	2	5	10	3	6

5.4 SUMMARY

Regression analysis is the process of developing a statistical model which is used to predict the value of a dependent variable by at least one independent variable. In simple linear regression analysis, there are two types of variables. The variable whose value is influenced or is to be predicted is called dependent variable and the variable which influences the value or is used for prediction is called independent variable. Simple Linear regression is based on the slope – intercept equation of the line. In regression analysis, sample regression model can be used to make predictions about the population parameters. So, β_0 and β_1 (population parameters) are estimated on the basis of sample statisticsb$_0$ and b$_1$.For this purpose Least Square Method is used. Regression is an important tool in economics it is widely used to measure the relationship between dependent and independent variables such as price and demand or price or supply etc.

5.5 GLOSSARY

- **Regression Equations-** Regression equations are the algebraic formulation of regression lines. The regression equations may be regarded as expressions for estimating from a given value of one variable the average corresponding value of the other.
- **Demand Function:** The demand function represents the relationship between the quantity of a commodity demanded (denoted as *x*) and its price (denoted as *p*). It follows an inverse relationship, meaning that when the price increases, the quantity demanded decreases, and when the price decreases, the quantity demanded increases.
- **Supply Function:** The supply function describes the connection between the quantity of a commodity supplied (denoted as *x*) and its price (denoted as *p*). Unlike the demand function, it exhibits a direct relationship, where an increase in price leads to a rise in the quantity supplied, while a decrease in price results in a lower quantity supplied.
- **Linear Regression:** When one variable changes with other variable in some fixed ratio, this is called linear regression. Such type of relationship is depicted on the graph by means of a straight line or first degree equation in variables x and y.

5.6 ANSWERS TO SELF-CHECK EXERCISES

Self-Check Exercise 5.1

Ans. Q1. Refer to Section 5.3

Ans. Q2. Refer to Section 5.3 (Example 5) **$Y = 0.257X + 3.201$, $X = 0.45\ Y + 4.75$**

5.7 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

5.8 TERMINAL QUESTIONS

Q1 What is the conceptual framework of simple linear regression?

Q2 You are given below the following information about advertising and sales

	Advertise Expenditure (X) (Rs. Lakhs)	**Sales (Y) (Rs. Lakhs)**
Mean	10	90
S.D.	3	12
Correlation Coefficient	0.8	

(i) Obtain the two regression lines.

(ii) Find the likely sales when advertisement expenditure is Rs.15 lakhs.

(iii) What should be advertisement expenditure if the Company wants to attain sales target of Rs.120 lakhs?.

REGRESSION: NUMERICAL PROBLEMS

STRUCTURE

6.1 Introduction
6.2 Learning Objectives
6.3 Numerical Problems
Self-check Exercise 6.1
6.4 Summary
6.5 Glossary
6.6 Answers to Self-check Exercises
6.7 References/Suggested Readings
6.8 Terminal Questions

6.1 INTRODUCTION

In this last unit, we have studied the different method of measuring regression. In this unit, we will solve some of the numerical problems relating to regression analysis.

6.2 LEARNING OBJECTIVES

After going through this unit you will be able to solve numerical problems relating to regression analysis.

6.3 NUMERICAL PROBLEMS

Example 1:
Obtain two regression equations from the following data. Also find X if Y = 10 , Y if X = 15

X	2	4	6	8	10	12
Y	4	2	5	10	3	6

Solution:

		A = 7		**A = 5**		
X	**Y**	**dx = (x-7)**	**dx^2**	**dy = (Y-5)**	**dy^2**	**dxdy**
2	4	-5	25	-1	1	5
4	2	-3	9	-3	9	9
6	5	-1	1	0	0	0
8	10	1	1	5	25	5
10	3	3	9	-2	4	-6
12	6	5	25	1	1	5
42	**30**	**0**	**70**	**0**	**40**	**18**

X on Y	**Y on X**
Mean	**Mean**
$\bar{X} = \frac{\sum X}{N} = \frac{42}{6} = 7$	$\bar{Y} = \frac{\sum Y}{N} = \frac{30}{6} = 5$
Regression Coefficient of X on Y	**Regression Coefficient of Y on X**
$b_{xy} = \frac{N\sum dxdy - (\sum dx)(\sum dy)}{N\sum dy2 - (\sum dy)2}$	$b_{yx} = \frac{N\sum dxdy - (\sum dx)(\sum dy)}{N\sum dx2 - (\sum dx)2}$
$= \frac{6(18)-(0)(0)}{6(40)-(0)(0)} = 0.45$	$= \frac{6(18)-(0)(0)}{6(70)-(0)(0)} = 0.257$
Regression Equation of X on Y	**Regression Equation of Y on X**
$X-\bar{X} = bxy\ (Y-\bar{Y})$	$Y-\bar{Y} = byx\ (X-\bar{X})$
$X – 7 = 0.45\ (Y – 5)$	$Y – 5 = 0.257\ (X – 7)$
$X – 7 = 0.45\ Y – 2.25$	$Y – 5 = 0.257\ X – 1.799$
$X = 0.45\ Y – 2.25 + 7$	$Y = 0.257X – 1.799 + 5$
$X = 4.75 + 0.45\ Y$	$X = 3.201 + 0.257\ X$
if Y = 10	***if X = 15***
$X = 4.75 + 0.45(10)$	$Y = 3.201 + 0.257\ (15)$
$X = 9.25$	$Y = 7.056$

Example 2:
Find two regression lines from the following data

X	10	12	18	16	15	19	18	17
Y	30	34	45	44	45	48	44	46

Solution :

	A=16			**A=42**		
X	**dx= (X-16)**	**dx^2**	**Y**	**dy=(Y-42)**	**dy^2**	**dxdy**
10	-6	36	30	-12	144	72
12	-4	16	34	-8	64	32
18	2	4	45	3	9	6
16	0	0	44	2	4	0
15	-1	1	45	3	9	-3
19	3	9	48	6	36	18
18	2	4	44	2	4	4
17	1	1	46	4	16	4
125	**-3**	**71**	**336**	**0**	**286**	**133**

X on Y

Mean

$$\bar{X} = \frac{\sum X}{N} = \frac{125}{8} = 15.265$$

Regression Coefficient of X on Y

$$b_{xy} = \frac{N\sum dxdy - (\sum dx)(\sum dy)}{N\sum dy2 - (\sum dy)2}$$

$$= \frac{8(133)-(-3)(0)}{8(286)-(0)(0)} = \frac{1064-0}{2288-0} = 0.465$$

Regression Equation of X on Y

X-$\bar{X}$ = bxy (Y-$\bar{Y}$)

X – 15.265 = 0.465 (Y – 42)

X – 15.265 = 0.465 Y – 19.53

X = 0.465 Y – 19.53 + 15.265

X = - 4.265 + 0.465Y

if Y = 50

X = - 4.265 + 0.465(50)

X = 18.985

Y on X

Mean

$$\bar{Y} = \frac{\sum Y}{N} = \frac{336}{8} = 42$$

Regression Coefficient of Y on X

$$b_{yx} = \frac{N\sum dxdy - (\sum dx)(\sum dy)}{N\sum dx2 - (\sum dx)2}$$

$$= \frac{8(133)-(-3)(0)}{8(71)-(-3)(-3)} = \frac{1064-0}{568-9} = 1.903$$

Regression Equation of Y on X

Y-$\bar{Y}$ =byx (X-$\bar{X}$)

Y – 42 = 1.903 (X – 15.265)

Y– 42 = 1.903 X – 29.049

Y = 1.903 X – 29.049 +42

Y = 12.95 + 1.903X

if X = 20

X = 12.95 + 1.903 (20)

X = 51.01

Example 3:

From 10 observations on Price *(X*) and supply (*Y)* of a commodity, the following figures were obtained.

∑X=130, ∑Y=220, ∑X2=2288, ∑Y2=5506 and ∑Xy=3467

Compute a line of regression of Y on X and estimate the supply when the price is 16

Solution:

Regression line of Y on X is given by

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

$$\overline{X}=\frac{\Sigma X}{N}=\frac{130}{10}=13$$

$$\overline{Y}=\frac{\Sigma Y}{N}=\frac{220}{10}=22$$

$$b_{yx}=r\frac{\sigma_y}{\sigma_x}=\frac{N\,\Sigma XY-\Sigma X\,\Sigma Y}{N\Sigma X^2-(\Sigma X)^2}$$

$$=\frac{10\times 3467-130\times 220}{10\times 2288-(130)^2}$$

$$=\frac{34670-28600}{22880-16900}$$

$$=\frac{6070}{5980}=1.015$$

$$Y-22=1.015(X-13)$$

$$Y=1.015X-13.195+22$$

$$Y=1.015X+8.805$$

When X = 16, Y is given by

$$Y=1.015(16)+8.805$$

$$=16.24+8.805$$

$$Y=25.045$$

Thus the estimated supply when price is 16 units is 25.045.

Example 4:

For 25 observations on Price(X) and Supply (Y) , the following data were obtained:

$\bar{X}$ = 5, $\bar{Y}$=4, $\sum X^2$=680, $\sum Y^2$= 480. $\sum XY$=508

It was later discovered that two pairs of observations (X,Y) were noted (8,10),(5, 8) while correct values were (10, 12) and (3, 6) respectively. Determine the correct values and find:

(i) The two regression coefficients;
(ii) the price (X) when supply (Y)=80 units and Supply (Y) when price(X)=70
(iii) correlation coefficient; and

Solution:

$\sum X = n\bar{X} = 25(5) = 125$

Corrected $\sum x = 125 - (8 + 5) + (10+3) = 125$

$\sum Y = n\bar{Y} = 25(4) = 110$

Corrected $\sum Y = 100 - (10+8) + (12+6) = 100$

Corrected $\sum X^2 = 680 - (8)^2 - (5)^2 + (10)^2 + (3)^2 = 680 - 64 - 25 + 100 + 9 = 700$

Corrected $\sum Y^2 = 480 - (10)^2 - (8)^2 + (12)^2 + (6)^2 = 480 - 100 - 64 + 144 + 36 = 496$

Corrected $\sum XY = 508 - (8\times10) - (5\times8) + (10\times12) + (3\times6)$

$= 508 - 80 - 40 + 120 + 18 = 526$

Corrected $\bar{X} = \frac{Corrected \sum X}{n} = \frac{125}{25} = 5$ Corrected $\bar{Y} = \frac{Corrected \sum Y}{n} = \frac{100}{25} = 4$

X on Y

Regression Coefficient of X on Y

$$b_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum Y2 - (\sum Y)2}$$

$$= \frac{25(256)-(125)(100)}{25(496)-(100)(100)} = \frac{526-500}{496-400}$$

$$= \frac{26}{96} = 0.271$$

Regression Equation Coefficient of X on Y

$X - \bar{X} = bxy(Y - \bar{Y})$

$X - 5 = 0.271(Y - 4)$

$X - 5 = 0.271Y - 1.084$

$X = 0.271Y - 1.084 + 5$

$X = 0.271Y + 3.916$

When Supply (Y) = 80 units , then

$X = 0.271(80) + 3.916$

$= 25.596$

Y on X

Regression Coefficient of Y on X

$$b_{yx} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X2 - (\sum X)2}$$

$$= \frac{25(256)-(125)(100)}{25(700)-(125)(125)} = \frac{526-500}{700-625}$$

$$= \frac{26}{75} = 0.347$$

Regression Equation of Y on X

$Y - \bar{Y} = byx(X - \bar{X})$

$Y - 4 = 0.347(X - 5)$

$Y - 4 = 0.347X - 1.735$

$Y = 0.347X - 1.735 + 4$

$Y = 0.347X + 2.265$

When Price (X) = 70 then

$Y = 0.347(70) + 2.265$

$= 26.555$

Correlation Coefficient

$$r = \pm\sqrt{bxy \times byx}$$

$$= \pm\sqrt{0.347 \times 0.271}$$

$= 0.3066$

Example 5:

A building contractor seeks to determine whether there is a relationship between the number of building permits issued and the corresponding sales volume of such buildings over the past years. To analyze this, he gathers data on sales (Y, measured in thousands of rupees) and the number of building permits issued (X, measured in hundreds) over the last ten years. The computed results are as follows:

$\sum X=117$, $\sum Y=78$, $\sum XY=981$, $\sum X^2=1491$, $\sum Y^2=662$

(i) What is the projected sales volume for the upcoming year if the issuance of 2,000 building permits is anticipated?

(ii) How much of a change in sales can be expected with an increase of 100 building permits?

Solution:

(i) Given that sales are denoted by Y and the number of building permits issued is represented by X, we will determine the regression equation of Y on X to estimate the expected sales level. The regression equation is expressed as:

$$Y-\bar{Y} = byx\ (X-\bar{X}) \quad \ldots\ldots\ldots\ldots\ldots\ldots(i)$$

$$\text{Where } \bar{X} = \frac{\sum X}{N} = \frac{117}{10} = 11.7, \quad \bar{Y} = \frac{\sum Y}{N} = \frac{78}{10} = 7.8$$

$$b_{yx} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X2 - (\sum X)2}$$

$$= \frac{10(981)-(117)(78)}{10(1491)-(117)(117)} = \frac{68.40}{122.10} = 0.56$$

Substitute these values in (i), we get

$$Y - 7.8 = 0.56\ (X - 11.7)$$

$$\text{Or } Y = 0.56\ X + 1.25$$

(i) Since the building permits expressed in hundreds, 2000 permits will mean X = 20 and the level of sales that can be expected on the issue of 2000 permits next year will be:

$$Y = 0.56\ (10) + 1.25 = \text{Rs. } 12{,}450/-$$

(ii) 100 building permits will mean an increase of one unit in X-value and therefore there will be a change of 0.56 thousand rupees in Y value, I.e., sales will change by Rs. 560 with an increase of 100 building permits.

Example 6:

The following table gives age (X) in years of cars and annual maintenance cost (Y) (in hundred rupees):

X	1	3	5	7	9
Y	15	18	21	23	22

Estimate the maintenance cost for a 4 year old car after finding the regression equation.

Solution : Since here we are to estimate the maintenance cost for a 4 year old car, we have to find out the regression equation of cost to age which means the regression equation of Y on X.

Age	Cost	A = 5		A = 20		
X	Y	dx = (x-5)	dx2	dy = (Y-20)	dy2	dxdy
1	15	-4	16	-5	25	20
3	18	-2	4	-2	4	4
5	21	0	0	1	1	0
7	23	2	4	3	9	6
9	22	4	16	2	4	8
25	**99**	**0**	**40**	**-1**	**43**	**38**

Regression equation of Y on X is

$$Y-\bar{Y} = byx\,(X-\bar{X}) \quad \ldots\ldots\ldots\ldots\ldots\ldots(i)$$

$$\text{Where } \bar{X} = \frac{\sum X}{N} = \frac{25}{5} = 5, \quad \bar{Y} = \frac{99}{5} = 19.8$$

$$b_{yx} = \frac{N\sum dxdy - (\sum dx)(\sum dy)}{N\sum dx2 - (\sum dx)2}$$

$$= \frac{5(38)-(0)(-1)}{5(40)-(0)(0)} = \frac{38}{40} = 0.95$$

Substitute these values in (i)

$\therefore$ $Y-19.8=0.95\,(X-5)$ or $Y= 0.95X+15.05$

When $X= 4$, $\hat{Y} = 0.95 \times 4 + 15.05 = 18.85.$

Example 7:

A panel of judges A and B graded seven debaters and independently awarded the following marks:

Debater	1	2	3	4	5	6	7
Marks by A	40	34	28	30	44	38	31
Marks by B	32	39	26	30	38	34	28

An eighth debater was awarded 36 marks by Judge A while Judge B was not present. If Judge B was also present, how many marks would you expect him to award to eighth debater assuming same degree of relationship exists in judgment?

Solution: Let the marks by the judge A be denoted by the variable X and the marks by judge B by the variable Y. We are given that for the eight performance the score by judge A is 36, i.e., X=36 and we want the corresponding score by judge B, i.e., Y. Hence we have to compute the line of regression of Y on X .

COMPUTATION OF REGRESSION LINE OF Y ON X

Marks by A	Marks by B	$d_x = x-35$	d_x^2	$d_y = y-32$	d_y^2	dx dy
40	32	5	25	0	0	0
34	39	−1	1	7	49	−7
28	26	−7	49	−6	36	42
30	30	−5	25	−2	4	10
44	38	9	81	6	36	54
38	34	3	9	2	4	6
31	28	−4	16	−4	16	16
$\Sigma X = 245$	$\Sigma Y = 227$	$\Sigma d_x = 0$	$\Sigma d_x^2 = 206$	$\Sigma d_y = 3$	$\Sigma d_y^2 = 145$	$\Sigma d_x d_y = 121$

$\overline{X} = \frac{1}{N}\Sigma X = \frac{1}{7} \times 245 = 35,$ $\quad \overline{Y} = \frac{1}{N}\Sigma Y = \frac{1}{7} \times 227 = 32.42$

Regression coefficient of Y on X is :

$$b_{yx} = \frac{\Sigma d_x d_y - \frac{\Sigma d_x \times \Sigma d_y}{N}}{\left\{\Sigma d_x^2 - \frac{(\Sigma dx)^2}{N}\right\}} = \frac{121-0}{206-0} = 0.49$$

Hence, the equation of the line of regression of Y on X is :

$Y - \overline{Y} = b_{YX}(X - \overline{X})$

or $Y - 32.42 = 0.49(X - 35)$

or $Y = 0.49X - 17.15 + 32.42$

$\therefore$ $Y = 0.49X + 15.27$

When X=36, the estimate of Y is given by :

$\hat{Y} = 0.49 \times 36 + 15.27$

$= 17.64 + 15.27 = 32.91$

Hence if the judge B was present at the eighth performance, he would have assigned the score 32.91, i.e., 33.

Example 8:

The following data relate to the advertisment expenditure and sales:

	Advertisment Expenditure (x) Rs Lakhs	**Sales (Y) Rs Crores**
Arithmetic Mean	20	90
Standard Deviation	4	12
Correlation Coefficient	+0.9	

(i) Calculate the two regression equations.
(ii) Estimate the likely sales when advertisment expenditure is rupees 30 lakhs
(iii) What would be the likely advertisment expenditure for attaining sales of rupees 120 crores?

Solution. In the usual notations, we have

$\overline{X} = 20, \quad \overline{Y} = 90, \quad \sigma_x = 4, \quad \sigma_y = 12, \quad r = 0.9$

Equation of line of regression of Y on X is given by :

$$Y - \overline{Y} = r\frac{\sigma_Y}{\sigma_X}(X - \overline{X})$$

or $$Y - 90 = \frac{0.9 \times .12}{4}(X - 20) = 2.7X - 54$$

$\therefore$ $$Y = 2.7X + 36$$

The most likely sales $(\hat{Y})$ when the advertisement expenditure (X) is Rs.30 lakhs is given by :

$$\hat{Y} = 2.7 \times 30 + 36 = \text{Rs. } 117 \text{ crores}$$

Equation of line of regression of X on Y is : $X - \overline{X} = r\frac{\sigma_X}{\sigma_Y}(Y - \overline{Y})$

or $$X - 20 = \frac{0.9 \times 4}{12}(Y - 90) = 0.3\ Y - 27$$

$\therefore$ $$X = 0.3Y - 7$$

The most likely advertisement expenditure (X) for a sale target (Y) of Rs. 120 crores is given by :

$$\hat{X} = 0.3 \times 120 - 7 = 29.$$

Example 9:

The table below presents the average price and standard deviation of two stocks traded on the stock exchange.

Share	*Mean (in Rs.)*	*Standard deviation (in Rs.)*
A Co. Ltd.	*39.5*	*10.8*
B Co. Ltd.	*47.5*	*16.8*

If the coefficient of correlation between the prices of two shares is 0.42, find the most likely price of share A corresponding to a price of Rs. 55 observed in the case of share B.

Solution. Let the share A be denoted by Y and share B by X. We have to find equation of line of regression of Y on X, i.e.,

$$Y-\overline{Y} = r\frac{\sigma_Y}{\sigma_X}(X-\overline{X}) \text{ or } Y-39.5 = \frac{0.42 \times 10.8}{16.8}(X-47.5)$$

or $Y-39.5 = 0.27(X-47.5) \quad \therefore Y = 0.27X + 26.675.$

The most likely price of share A corresponding to a price of Rs. 55 observed in case of share B is given by :

$$\hat{Y} = 0.27(55)+26.675=41.5$$

Example 10:

Find out the regression equation showing the regression of capacity utilization on production from the following data:

	Average	Standard Deviation
Production (in lakh units)	35.6	10.5
Capacity utilization (in percentage)	84.8	8.5

Estimate the production when the capacity utilization is 70 %

Solution. Let the production and capacity utilization be denoted by X and Y respectively. Then we are given :

$$\overline{X} = 35.6,\ \overline{Y} = 84.8,\quad \sigma_x = 10.5,\quad \sigma_y = 8.5 \quad \text{and} \quad r = 0.62$$

Now the regression equation showing the regression of capacity utilization (Y) on production (X) is given by the formula :

$$Y - \overline{Y} = r\frac{\sigma_Y}{\sigma_X}(X - \overline{X})$$

or $$Y - 84.8 = \frac{0.62 \times 8.5}{10.5}(X - 35.6) = 0.5019\ X - 17.8676$$

$\therefore$ $$Y = 0.5019\ X + 66.9324$$

To estimate production, we shall have to find the regression equation of X on Y, i.e., $$X - \overline{X} = r\frac{\sigma_X}{\sigma_Y}(Y - \overline{Y})$$

or $$X - 35.6 = \frac{0.62 \times 10.5}{8.5}(Y - 84.8) = 0.766Y - 64.95$$

$\therefore$ $$X = 0.766\ Y - 29.35$$

When $$Y = 70,\ \hat{X} = 0.766 \times 70 - 29.35 = 24.27$$

Hence the estimate production is 2,42,700 units when the capacity utilization is 70%.

Example 11

The following calculations have been made for prices of twelve stocks(X) on the Calcutta Stock Exchange on a certain day along with the volume of sales. In thousands of shares (Y). From these calculations find the regression equation of prices on stocks, on the volume of sales of shares.

$\sum X=580$, $\sum Y=370$, $\sum XY=111494$, $\sum X^2=41658$, $\sum Y^2=17206$

Solution. *Regression equation of X on Y*

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

$$\bar{X} = \frac{\Sigma X}{N} = \frac{580}{12} = 48{\cdot}333 \; : \; \bar{Y} = \frac{\Sigma Y}{N} = \frac{370}{12} = 30{\cdot}833$$

$$r\frac{\sigma_x}{\sigma_y} \text{ or } b_{xy} = \frac{\Sigma XY - N\bar{X}\bar{Y}}{\Sigma Y^2 - N(\bar{Y})^2}$$ [when figures are given in original values]

$$= \frac{11494 - 12(48{\cdot}333 \times 30{\cdot}833)}{17206 - 12(30{\cdot}833)^2}$$

$$= \frac{11494 - 17883{\cdot}015}{17206 - 11408{\cdot}09} = \frac{-6389{\cdot}015}{5797{\cdot}91} = -1{\cdot}102$$

$$X - 48{\cdot}333 = -1{\cdot}102\,(Y - 30{\cdot}833)$$

$$X - 48{\cdot}333 = -1{\cdot}102\,Y + 33{\cdot}978 \text{ or } X = 82{\cdot}31 - 1.102\,Y.$$

Example 12:

A survey was carried out to examine the correlation between accommodation expenses (X) and spending on food and entertainment (Y). The following statistical data was recorded:

	Mean	Standard Deviation
Expenditure on Accommodation	Rs.173	63.15
Expenditure on Food & Entertainment	Rs.47.8	22.98
Co-efficient of Correlation	+0.57	

Based on this information:

(i) Derive the regression equation.
(ii) Estimate the expenditure on food and entertainment when accommodation expenses amount to Rs. 200.

Solution: To solve this problem we have to fit regression equation of Y on X

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

Given $\bar{X}$ =173, σ_x=63.15, $\bar{Y}$ =47.8, σ_y=22.98

And r=0.57. Substituting these values

$$Y\text{-}47.8 = 0.57\,\frac{22.98}{63.15} = (X - 173)$$

$$Y\text{-}47.8 = .207\,(X\text{-}173)$$

$$Y\text{-}47.8 = .207X\text{-}35.81$$

$$Y = .207X + 11.99$$

Putting X=200 in the equation

$$Y = .207(200) + 11.99 = 53.39$$

Example 13:

	Rainfall (in cm)	Production (tones)
Mean	35	50
Standard deviation	5	8
Coefficient of correlation equation	0.8	

From the following data, find the most likely production corresponding to the rainfall of 40 cm.

Solution : Let X denotes rainfall and Y denotes productions. To find the likely production (Y) when the rainfall (X) is 40cm, we have to find the regression equation of Y on X.

$\overline{X} = 35, \overline{Y} = 50, \sigma_x = 5, \sigma_y = 8, r = 0.8.$

Regression equation of Y on X

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

$$Y - 50 = (0.8) \times \left(\frac{8}{5}\right)(X - 35).$$

$$Y - 50 = 1.28(X - 35)$$

$$= 1.28X - 44.8$$

$$Y = 1.28X - 44.8 + 50$$

$$= 1.28X + 5.2.$$

When the rainfall is 40 cm ($X = 40$), the likely production is

$Y = 1.28(40) + 5.2 = 56.4$ tonnes.

Self-Check Exercise 6.1

Q1. Obtain two regression equations from the given table

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

Q2. Find b_{xy} , if $\sigma_x^2 = 0.75$, $\sigma_y^2 = 1.2$, $r_{xy} = 0.65$

Q3. Find r, if $\sigma_x^2 = 25$, $\sigma_y^2 = 625$, $b_{xy} = 0.16$

6.4 SUMMARY

In this unit, we have solved the numerical problems related to the regression analysis. Different types of numerical problems have been solved in this unit.

6.5 GLOSSARY

- **Regression Equations-** Regression equations are the algebraic formulation of regression lines. The regression equations may be regarded as expressions for estimating from a given value of one variable the average corresponding value of the other.
- **Linear Regression:** When one variable changes with other variable in some fixed ratio, this is called linear regression. Such type of relationship is depicted on the graph by means of a straight line or first degree equation in variables x and y.

6.5 ANSWERS TO SELF-CHECK EXERCISES

Self-Check Exercise 6.1

Ans. Q1. Refer to Example 2 (Y = 0.95 X + 7.25, X = 0.95Y – 6.4)

Ans. Q2. $b_{xy} = r.\frac{\sigma_y}{\sigma_x}$

$= 0.65\,\frac{\sqrt{0.75}}{\sqrt{\sigma y}} = 0.52$

Ans. Q3. $b_{xy} = r.\frac{\sigma_y}{\sigma_x}$

$0.16 = r\,\frac{\sqrt{25}}{\sqrt{625}}$

$0.16 = r\,\frac{5}{25}$

$r = 0.16 \times 5 = 0.8$

6.6 REFERENCES/SUGGESTED READING

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum’s Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

6.7 TERMINAL QUESTIONS

Q1. You are the following given data about sale and advertisement expenditure of a firm.

	Sale (Rs. Crore)	Adv. Exp. (Rs. Crore)
Mean	50	10
S.D.	10	2
Correlation Coefficient	+0.9	

(i) Obtain two regression equations

(ii) Estimate the sale when expenditure on advertisement is Rs 13.5 crore.

(iii) What should be advertisement budget if company wants to achieve the sale target of Rs. 70 crore.

Q2. The following table gives ages and blood pressure of 10 women

Age	56	42	36	47	49	42	60	72	63	55
Blood Pressure	147	125	118	128	145	140	155	160	149	150

Estimate the blood pressure of a woman whose age is 45 year.

Unit-07

ELEMENTARY APPLICATIONS OF REGRESSION

STRUCTURE

7.1 Introduction

7.2 Learning Objectives

7.3 Application of Regression

7.3.1 Application of Regression in Demand Function

7.3.2 Application of Regression in Supply Function

7.3.3 Application of Regression in Consumption

7.3.4 Application of Regression in Investment

7.4 Summary

7.5 Glossary

7.6 Answers to Self-Check Exercises

7.7 References/Suggested Readings

7.8 Terminal Questions

7.1 INTRODUCTION

Economist and business people are generally interested in knowing how changes in some variables affect the changes in other variables in other words how change in independent variable bring about change in dependent variable for example changes in parameters like inventory, production, supply, advertising and price affect other parameters like profits, inflation, demand, supply .Regression is an empirical tool used by economists to study these changes.

Regression refers to the concept of returning to the average. In statistics, regression analysis is used in various fields where two or more related variables tend to revert to their mean values. The primary goal of regression analysis is to understand the relationship between variables and utilize this understanding to predict the probable value of the dependent variable based on a given value of the independent variable.

7.2 LEARNING OBJECTIVES

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

- Elucidate the elementary application of regression in demand,
- Explain the elementary application of regression in supply, consumption, and investment
- Expound the purpose of regression in consumption
- Illuminate the relevance of regression in investment

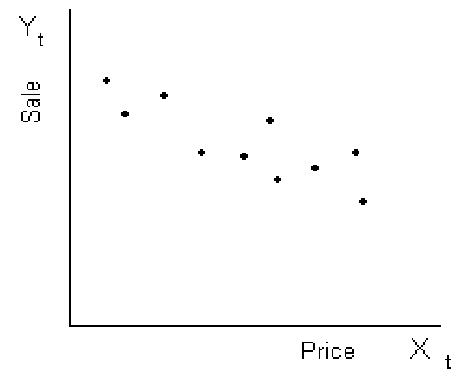
7.3 APPLICATION OF REGRESSION

The regression analysis is an integral part of every empirical investigation, given below are the applications of regression in demand, supply, consumption and investment.

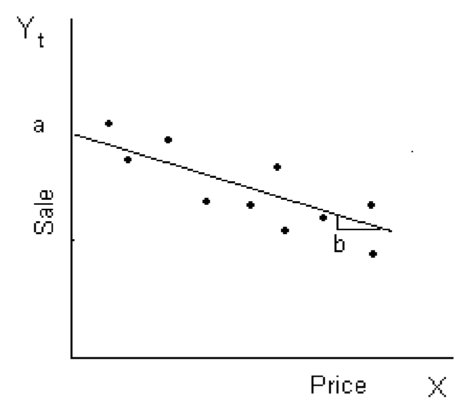
7.3.1 Application of Regression in Demand Function

Given data on relevant variables, the empirical method of Regression Analysis can be used to estimate demand functions. We will begin by considering a simplified version of a demand function showing quantity demanded as a function of only price. Such a demand function would clearly be inadequate to successfully estimate a demand function. We use this simplified example of a demand function merely as a means of illustrating the technique of regression analysis.

Let Y_t be quantity demanded (i.e. Sales) in period t. Let X_t be the price charged in period t. Plotting a firm's historical data for these two variables we may have the following graph. As the graph reveals, there is apparently a negative relationship between Sales and Price, i.e. as Price increases, Sales also *tends* to decrease.



The relationship between Y_t and X_t can be represented by a line that "best fits" the scatter of points. The line shown is defined by its intercept, *a*, and its slope, *b*. Notice, however, that the actual points do not fall exactly on the line. This is due to the fact that Sales may be partly determined by price, but other factors such as advertising, which are not taken into account, also affect Sales.



A business analyst might use univariate regression method to estimate gasoline consumption using the price of oil as predictor. The resulting regression would estimate how much gasoline production would change in response to change in the price of per barrel oil.

Multivariate or multiple regression, is the study of relationship between dependent variable, say X_1 and the joint effect of all other independent variables (X_2 X_3…….X_n)on X_1 . Such as the sales(X_1) of departmental store may depend on the population size(X_2) of the area, the size of the store inventory(X_3), the amount of money spend on advertisement(X_4) and the number of competing stores in the sales area (X_5)

Suppose that a firm's demand function is

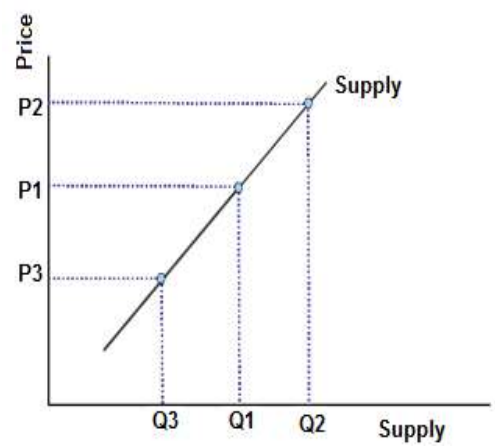
$$D = A + B_1X + B_2P + B_3I + B_4Pr$$

where D is the quantity demanded of the firm's product, X is the selling expense (such as advertising) of the firm, P is the price of its product, I is the disposable income of consumers, and Pr is the price of the competing product sold by its rival. What we want are estimates of the values of A, B_1, B_2, B_3, and B_4.Regression analysis enables us to obtain them from historical data concerning Y, X, P, I, and Pr.

7.3.2 Application of Regression in Supply

Consider an economic situation in which the variables are the price and the quantity of the commodity supplied. Let p be the price of one unit of the commodity in rupees and x be the number of units of the commodity. Upon reflection it should seem reasonable that the amount of the commodity supplied in the market place by the producers depends on the price of the commodity. According to the law of supply there is a direct relationship between price (p) and the quantity supplied(x).Stated another way, as price increases, quantity supplied increases, and as price decreases, quantity supplied decreases.

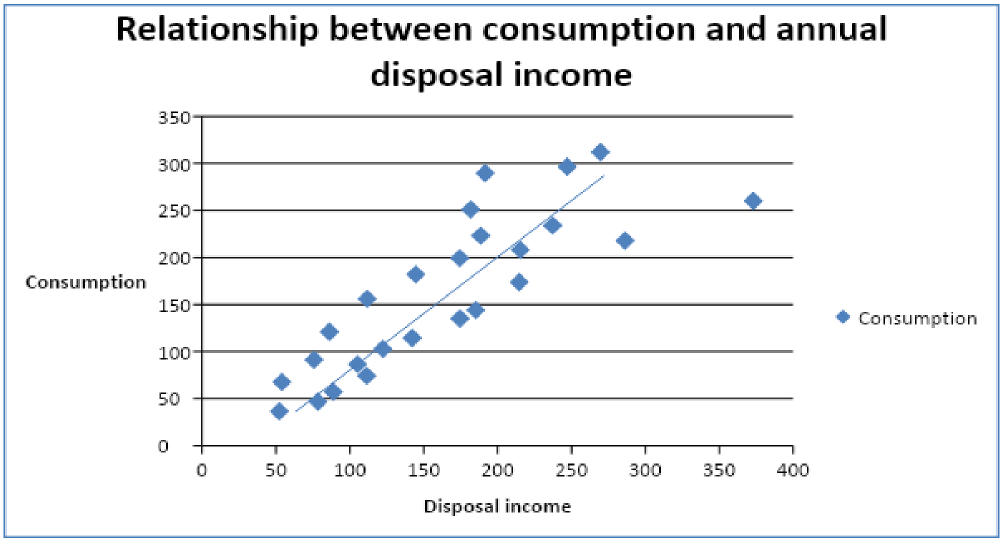
Supply Function (Definition): the relationship between quantity (given by x) of a commodity supplied and the price (given by p) is called a supply function.



The graph above depicts that the slope of a supply curve is usually positive or upward, that is, all these curves are sloping upward from left to right indicating that, as the price increases, quantity supplied increases, quantity supplied increases and as price decreases, quantity supplied decreases. Given data on relevant variables, the empirical method of *Regression Analysis* can be used to estimate supply functions.

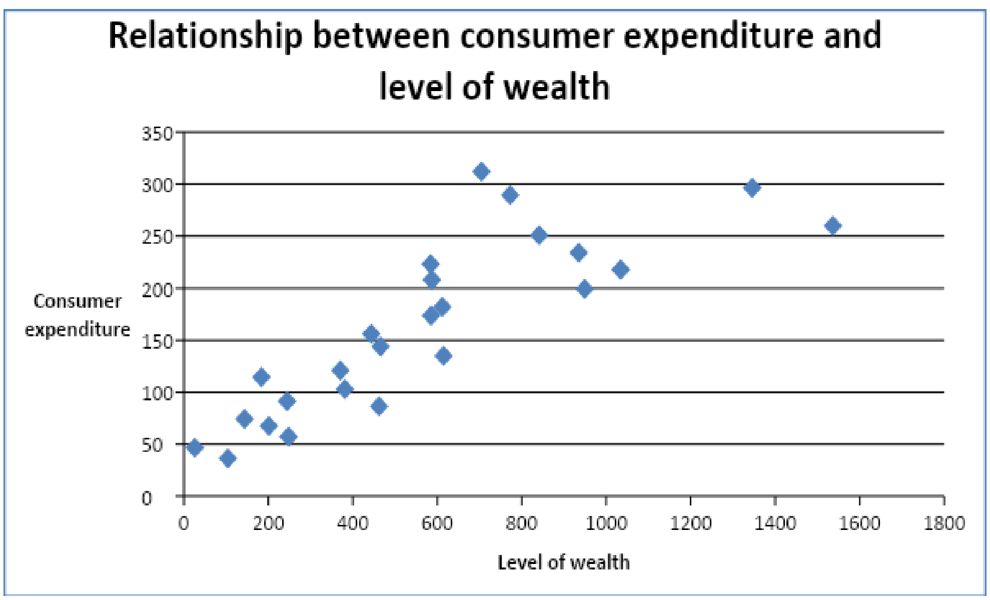
7.3.3 Application of Regression in Consumption

Consumption is a function of income and wealth. The amount of consumption is plotted on the y – axis while the disposable income is plotted on the x-axis. Different graphs will be plotted for each explanatory variable. A Scatter diagram tries to establish if there exists a linear relationship between two variables plotted on the diagram. This can be observed by looking at the trend of the scatter plots. The graph below shows the scatter diagram of consumer expenditure (Y) against disposable income (X_1).



The scatter plots on the diagram above tend to slope upwards. The points on the diagram tend to concentrate along the line. This indicates a strong positive relationship between the variable. The correlation coefficient shows the value of correlation between consumption expenditure and annual disposable income.

The graph below shows the scatter diagram of consumer expenditure (Y) against the level of wealth (X_2).



The scatter plots on the diagram above tend to slope upwards. This indicates a strong positive relationship between the variable. Thus, it is clear that both variables have a positive relationship with consumer expenditure.

Regression of total consumer expenditure on annual disposal income

The dependent variable is the total consumer expenditure while the independent variable is the amount of annual disposable income. A sample of twenty-five households is used to estimate the regression equation. The regression line will take the form.

$$Y_i = E(Y_i) + \varepsilon_i = \alpha_1 + \beta_1 X_{1i} + \varepsilon_i$$

when the ordinary least squares method is used. The regression line can be simplified as shown below. Simplified regression equation $Y = b_0 + b_1X_1$

Y = Consumer expenditure

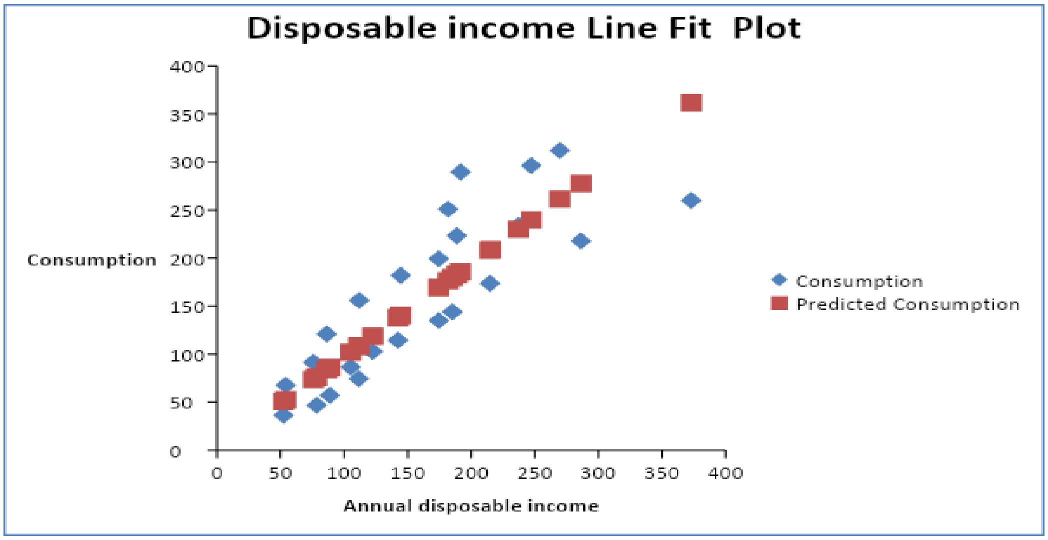
X_1 = annual disposable income

The theoretical expectations are b_0 can take any value and $b_1 > 0$.

Regression Results

Variable		**Coefficients of the variable**
b_0	Y-intercept	0
b_1	Coefficient of annual disposable income	0.969973

From the above table, the regression equation can be written as $Y = 0.969973X_1$. The coefficient value of 0.969973 implies that as the annual disposal income increases by one unit, the consumer expenditure will increase by 0.969973 units. The positive value of the coefficient implies a positive relationship between the consumer expenditure and annual disposal income as was evident in the scatter diagram. The regression line can be drawn on the scatter diagram as shown below.



In the above diagram, the line of best fit is shown by the plots of predicted consumption.

7.3.4 Application of Regression in Investment

In general, linear regression can be used to identify trends in the behavior of one asset relative to another asset or a relevant index. The most common uses of regression analysis by investors are to establish the relative variability of an asset with its underlying index and to estimate required rates of return based on relative volatility. The capital asset pricing model was developed using linear regression analysis, and a common measure of the volatility of a stock or investment is its beta- which is determined using linear regression .Linear regression is the key in assessing the risk associated with most investment opportunities.

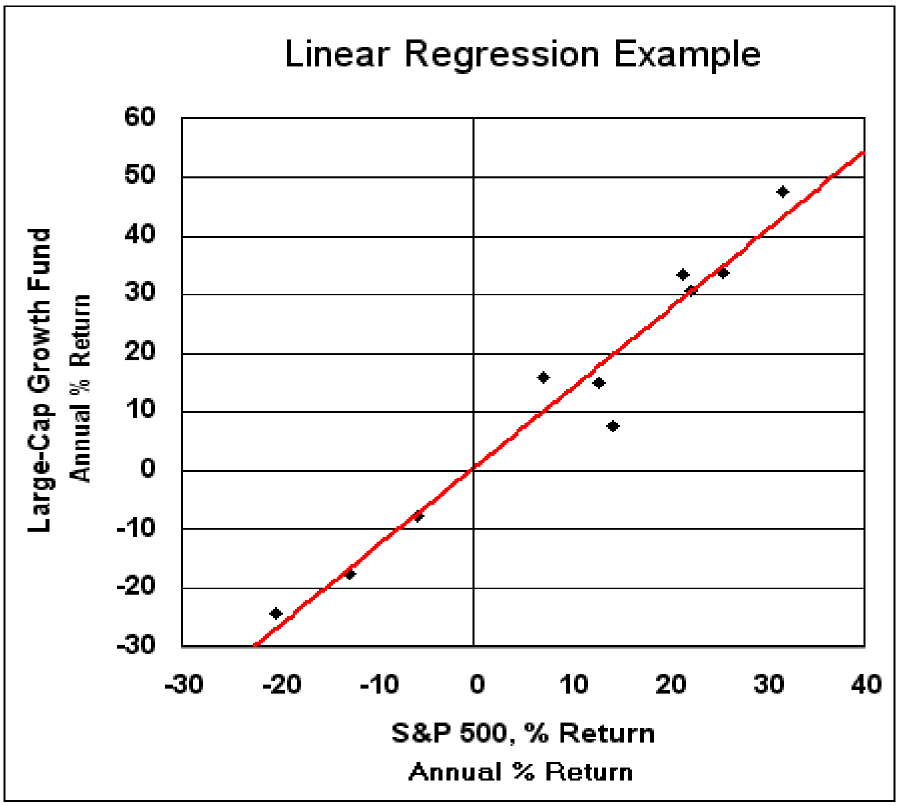
When you input two series (lists) of numbers, identify one as independent and one as dependent, and run a linear regression analysis, the output will provide the information you need to construct an equation, i.e., the regression model, in the form Y = a + bX, which you should recognize as the equation of a straight line. Y is the dependent variable, X is the independent variable, **a** is the Y intercept and **b** is the slope of the line. In linear regression there is actually a third term, e, which is the error of the estimate. But, being that we don't know what the error is, it's dropped from the equation. Now, the probable error of the estimate is of concern to us it is addressed below. If there is little or no correlation between the variables, the error will be too high and the model will be unreliable. The regression equation allows us to compute estimated values of Y for any value of X.

Application of regression in investment can be explained with the help of the following illustration.

Year	S&P 500 Total return	Large-Cap Growth Fund
1997	31.7	47.5
1998	25.5	33.6
1999	21.5	33.4
2000	-5.8	-7.6
2001	-12.8	-17.6
2002	-20.2	-24.4
2003	22.2	30.8
2004	12.8	15.1
2005	7.1	15.9
2006	14.3	7.5
Mean	9.6	13.4
Std Dev	17.3	23.9

Regression Output:

Constant	3.2513963 = a
Std Err of Y Est	6.7994303
R Squared	0.9281605
No. of Observations	10
Degrees of Freedom	8
X Coefficient(s)	1.2295772 = b
Std Err of Coef.	0.120943
t	10.166583



Here we have regressed the returns of a large-cap growth fund on the returns of the S&P 500. From the regression output we can assemble the formula that is the regression model.

Y = a + bX = 3.25 + 1.23(X)

The slope of the regression line, b, is a measure of the expected change in Y per unit of change in X. This is used in investing as a relative measure of variability,

Now we can use this formula to compute values of Y for any value of X, where X is the return on the S&P 500 index and Y is the estimated return on the large-cap growth fund. For example: If there's a consensus among analysts that the return on the

S&P 500 will be 15% next year, we can estimate the expected return on the growth fund as [3.25 + 1.23(15)] = 21.70%. The error term, e, noted above would be the difference between the estimated value of 21.70% and what the fund actually returned next year.

We can get an idea of how accurate the regression model is by examining the graph and the regression output. The straight line in the graph is the regression line, i.e., the straight line described by the formula Y = 3.25 + 1.23(X). By examination, it appears that the line fits the data points well. If the points were more dispersed it would be hard to make an evaluation on this basis. So we look at the R-squared of the regression output, 0.928. This is known as the coefficient of determination and it is used as a measure of the "goodness of fit." 0.928 indicates an excellent fit. With linear regression, 0.5 and greater is pretty darned good. R-squared is a measure of the dispersion of the data points around the regression line, and it tells us what percent of the variation in the dependent variable is explained by its correlation with the independent variable. In the example, 92.8% of the variation in X is explained by its correlation with Y. But remember, correlation does not necessarily mean causation.

R-squared is the square of the correlation coefficient, R. But remember, squaring is a one way street. The square root of a number is an absolute value, i.e., we don't know whether it's positive or negative. In this case R could be +0.963 or -0.963, and, as you know from the discussion of correlation, there is a vast difference between a correlation of +0.963 and -0.963. But, looking at the graph and seeing how well the line fits the data, it seems safe to assume that it is +0.963. Indeed, the sign of the slope of the regression line will always match the sign of the correlation coefficient.

Self-Check Exercise 7.1

Q1. Elucidate the application of regression in demand?

Q2. Illuminate the relevance of regression in Investment?

7.4 SUMMARY

In this unit, we have learnt about elementary application of regression in demand, supply, consumption and investment. After going through the lesson we have learnt the importance of regression in our daily lives, in our daily lives we often come across situations in which some variables depend on other variables. Taking a simple example of demand and price in which price is an independent variable and demand is a dependent variable. Change in the price level changes the demand for a commodity; similarly change in advertisement expenditure change the sales of the firm.

7.5 GLOSSARY

- **Regression:** the statistical method which helps us to estimate or predict the unknown value of one variable from the known value of the related variable is called regression.
- **Demand:** An economic principle that describes a consumer's desire and willingness to pay a price for a specific good or service.
- **Investment:** An asset or item that is purchased with the hope that it will generate income or appreciate in the future. In finance, an investment is a monetary asset

purchased with the idea that the asset will provide income in the future or appreciate and be sold at a higher price.

- **Supply:** Supply can relate to the amount available at a specific price or the amount available across a range of prices.

7.6 ANSWERS TO SELF-CHECK EXERCISES

Self-Check Exercise 7.1

Ans. Q1: Refer to Section 7.3.1.

Ans. Q2: Refer to Section 7.3.4

7.7 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.

7.8 TERMINAL QUESTIONS

Q1. Explicate the application of regression in Consumption?

Q2. How do we fit a simple linear regression model?

Unit-08

TIME SERIES: MEANING AND SIGNIFICANCE

STRUCTURE

8.1 Introduction

8.2 Learning Objectives

8.3 Meaning of Time Series

8.3.1 Definition of Time Series

8.4 Utility of Time Series

8.5 Components of Time Series

8.5.1 Secular Trend

8.5.2 Seasonal Variations

8.5.3 Cyclical Variation

8.5.4 Irregular Variation

8.6 Measurement of Trend

8.6.1Graphic Method,

8.6.2 Semi Average method

8.6.3 Moving Average Method

8.6.4 Least Square Method

8.7 Summary

8.8 Glossary

8.9 Answers to Self-Check Exercise

8.10 References/Suggested Readings

8.11 Terminal Questions

8.1 INTRODUCTION

In today's rapidly evolving landscape, time series analysis plays a crucial role in both Economics and Statistics. A time series is a collection of data points recorded at regular intervals over time, capturing the changing patterns of a phenomenon. Many economic, business, and commercial datasets—such as those related to prices, production, consumption, national income, foreign exchange reserves, investment, sales, and stock market performance—are structured as time series. These datasets span extended periods and provide valuable insights into trends and fluctuations. Consequently, time series analysis holds significant importance in the fields of Business and Economics.

8.2 LEARNING OBJECTIVES

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

- Understand the meaning of Time Series
- List the utility of Time Series
- Elucidate Components of Time Series
- Construct a trend line for time series data
- Compute trend using different methods

8.3 MEANING OF TIME SERIES

A time series is a series of statistical data recorded in accordance with their time of occurrence. It is a set of observations taken at specified times, usually(but not always)at equal intervals. Thus a set of data depending on the time is called time series.

8.3.1 Definition of Time Series

Ya-Lun Chou: “A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables”.

Kenny and Keeping: “A set of data depending on time is called time series”.

Croxton and Cowden: “A time series consists of data arranged chronological”.

Spiegel: “A time series is a set of observation taken at specified times, usually at equal intervals”.

Morris Hamburg, ”A Time Series is a set of statistical observations arranged in chronological order”.

Self-Check Exercise-8.1

Q1. Define Time Series.

8.4 UTILITY OF TIME SERIES

(i) **Effective Tool for Predicting Behavioral Changes**: Time series analysis enables forecasting future trends by examining patterns and variations observed in the past.

(ii) **Understanding Historical Trends**: It aids in analyzing past behaviors of a given phenomenon, providing insights into its progression over time.

(iii) **Assessing Business Cycles**: Cyclical fluctuations within a time series offer valuable insights into economic phases such as boom, recession, depression, and recovery, helping businesses make informed decisions.

(iv) **Enhancing Comparisons**: Time series facilitates comparative analysis by evaluating variations in different phenomena across various time periods or locations.

(v) **Analyzing Current Trends**: It assists in assessing present fluctuations in economic indicators such as national income, prices, production, exports, and imports.

Self-Check Exercise-8.2

Q1. State the utility of Time Series.

8.5 COMPONENTS OF TIME SERIES

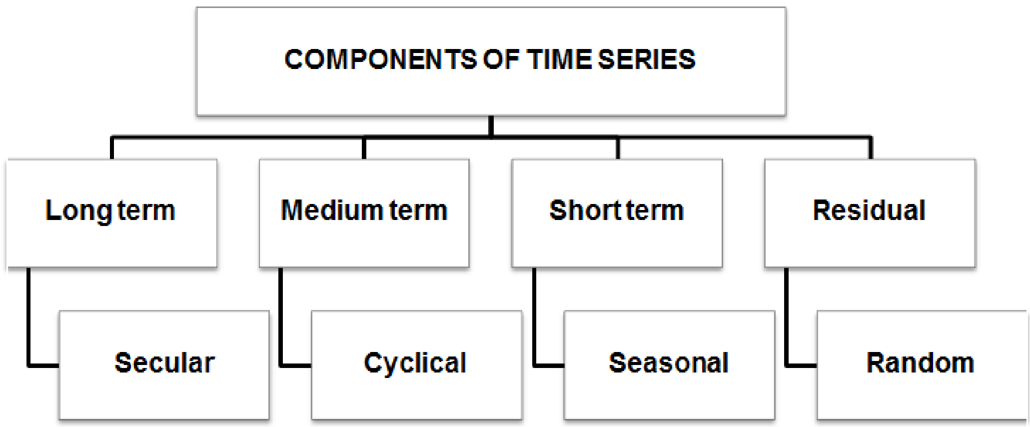
The main components of time series are:

8.5.1 Secular Trend

8.5.2 Seasonal Variations

8.5.3 Cyclical Variations

8.5.4 Irregular Variations



8.5.1 Secular Trend

The long-term pattern of time series data, which shows a consistent increase, decrease, or stability over an extended period, is known as the secular trend or simply the trend. This pattern is commonly observed in various economic and business-related data. For instance, time series related to population growth, production levels, product sales, prices, income, money supply, and bank deposits typically exhibit an upward trend. Conversely, factors such as mortality rates, epidemic occurrences, and cases of diseases like tuberculosis tend to show a declining trend due to advancements in medical technology, improved healthcare services, better sanitation, and enhanced nutrition. Secular trends are generally classified into two types:

(a) **Linear Trend:** When long-term rise or fall in a time series takes place by a constant amount, then that is called a linear trend. This is also known as straight line trend. This is represented by the following equations.

$$Y=a+bX$$

(b) **Parabolic Trend:** The trend is said to be parabolic when long-term rise or fall in a time series is not taking place at a definite rate. It has many forms but most prominent of them is second Degree Parabolic or Quadratic trend. Its equation is as follows:

$$Y=a+bX+cX^2$$

8.5.2 Seasonal Variations

Seasonal variations in a time series occur due to recurring patterns that follow a regular and periodic cycle of less than a year, typically within a span of 12 months. These variations repeat in a consistent manner year after year. Seasonal fluctuations are observed in data recorded at quarterly, monthly, weekly, daily, or even hourly intervals. While the magnitude of these variations may differ across different timeframes, they all share a common periodicity of one year. Several factors contribute to seasonal variations, including climatic conditions, production and consumption trends, fluctuations in sales and profits, as well as banking activities such as deposits and clearings.

The primary causes of seasonal variations can be classified into two categories:

1. **Natural Influences** – These result from environmental and climatic conditions. For example, the demand for umbrellas rises significantly during the rainy season, ice cream sales increase in summer, and woollen clothing experiences a surge in demand during winter. These variations are directly influenced by seasonal weather patterns.
2. **Social and Cultural Practices** – These variations are driven by human habits, traditions, and social conventions. For instance, jewellery and ornament sales increase during wedding seasons, while retail sales experience a significant boost during festivals such as Diwali and Dussehra.

8.5.3 Cyclical Variations

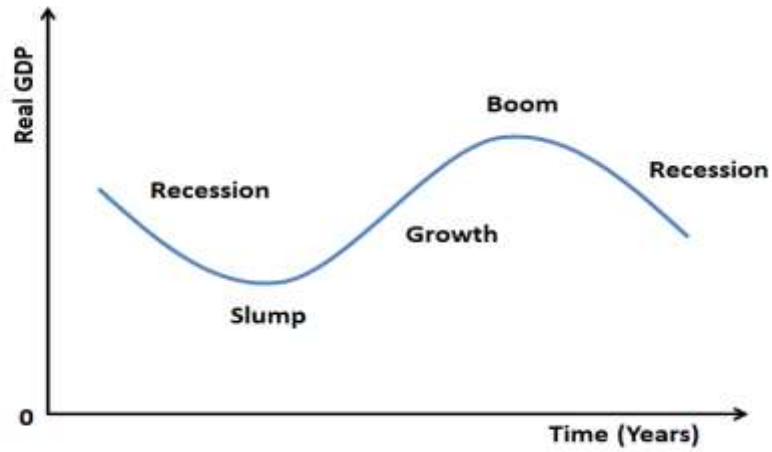
Cyclical variations in a time series refer to fluctuations that occur over a period longer than one year. These movements are characterized by recurrent ups and downs, although they do not necessarily follow a perfectly uniform pattern over fixed time intervals. The duration of one complete cycle generally ranges between 7 to 9 years.

In business and economic contexts, these fluctuations are associated with business cycles, which consist of four key phases:

- **Boom (Prosperity)** – A period of rapid economic growth and high business activity.
- **Recession** – A phase where economic activity starts to decline.
- **Depression** – The lowest point in the cycle, marked by reduced production and economic stagnation.
- **Recovery** – A phase where economic activity begins to improve, leading to the next boom.

These cyclical patterns affect various economic indicators such as production levels, prices, wages, and investment trends. The recurrence of booms and downturns follows a relatively predictable sequence, with the complete cycle lasting between 7 and 9 years.

The graph clearly illustrates that business activities reach their peak during periods of prosperity, followed by a downturn marked by recession. This decline continues until it reaches the lowest point, known as the depression stage. However, after a phase of depression, the economy gradually recovers.



Understanding cyclical variations is crucial for business executives when formulating strategies to stabilize business activities. Awareness of economic cycles allows entrepreneurs to anticipate periods of boom and downturn, enabling them to take proactive measures to maintain a stable market for their products.

8.5.4 Irregular Variations or Random movements

Irregular variations, also referred to as random movements, are unpredictable in nature. These fluctuations are often triggered by unforeseeable factors such as wars, natural disasters (floods), strikes, or lockouts. At times, they may also result from the cumulative impact of numerous minor factors, which individually seem insignificant but collectively have a notable effect.

Since random movements do not follow any discernible pattern or repetition, they are considered residual variations. Due to their unpredictable nature, it is challenging to isolate and study them separately or forecast them with precision. The most practical approach is to analyze past occurrences to estimate their potential impact and incorporate contingency measures during stable business conditions.

Self-Check Exercise 8.3

Q1. What is Secular Trend?

Q2. What is meant by Seasonal Variations?

8.6 MODELS OF TIME SERIES ANALYSIS

The following are the two models which we generally use for the decomposition of time series into its four components. The objective is to estimate and separate the four types of variations and to bring out the relative effect of each on the overall behavior of the time series.

8.6.1 Additive Model, and

8.6.2 Multiplicative Model

8.6.1 Additive Model

In the additive model, we represent a particular observation in a time series as the sum of these four components.

i.e. $O = T + S + C + I$

where O represents the original data, T represents the trend. S represents the seasonal variations, C represents the cyclical variations and I represents the irregular variations.

In another way, we can write $Y(t) = T(t) + S(t) + C(t) + I(t)$

8.6.2 Multiplicative Model

In this model, four components have a multiplicative relationship. So, we represent a particular observation in a time series as the product of these four components:

i.e. $O = T \times S \times C \times I$

where O, T, S, C and I represents the terms as in additive model.

In another way, we can write $Y(t) = T(t) \times S(t) \times C(t) \times I(t)$

This model is the most used model in the decomposition of time series. To remove any doubt between the two models, it should be made clear that in Multiplicative model S, C, and I are indices expressed as decimal percentages whereas, in Additive model S, C and I are quantitative deviations about a trend that can be expressed as seasonal, cyclical and irregular in nature.

Self-Check Exercise 8.4

Q1. What is Additive Model of Time Series Analysis

Q2. What is Multiplicative Model of Time Series Analysis

8.7 MEASUREMENT OF TREND

The trend component in a time series is typically analyzed and measured using four commonly applied methods.

8.7.1 Graphic Method

8.7.2 Semi Average Method

8.7.3 Moving Average Method

8.7.4 Least Square Method

Self-Check Exercise 8.5

Q1. State the various methods of measuring Trend.

8.8 SUMMARY

In this chapter we have gone through time series .As already discussed Time series is a set of statistical observations on variables arranged in chronological order. The change in the variable concerned can be explained to same extent on the basis of components of time series. These components are secular trend, seasonal variations, cyclical fluctuations and irregular variations. Trend can be measured by the method of Graphic Method, Semi Average method Moving Average Method, and fitting equations by the method of least squares. Once we estimate the trend, we can predict the future values and also estimate the monthly or quarterly values from the annual trend.

8.9 GLOSSARY

- **Time Series**: A time series refers to a sequence of statistical data points recorded in chronological order. These observations are typically (but not necessarily) collected at consistent intervals. Any dataset that depends on time is classified as a time series.
- **Cyclical Variations**: Cyclical variations represent periodic fluctuations in a time series, where the oscillation cycle extends beyond one year.
- **Irregular Variations**: Also known as random movements, irregular variations are unpredictable fluctuations in a time series. These variations occur due to unforeseen factors such as natural disasters, strikes, wars, or other unexpected events.
- **Seasonal Variation**: These are repetitive and predictable patterns in a time series that occur within a period of less than a year, typically following the same trend annually. Seasonal variations are evident when data is collected at intervals such as daily, weekly, monthly, or quarterly.
- **Secular Trend**: The long-term movement of time series data, which may show an increasing, decreasing, or stable pattern over an extended period, is referred to as the secular trend or simply the trend.

8.10 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 8.1

Ans. Q1. Refer to Section 8.3

Self-Check Exercise 8.2

Ans. Q1. Refer to Section 8.4

Self-Check Exercise 8.3

Ans. Q1. Refer to Section 8.5.1

Ans. Q2. Refer to Section 8.5.2

Self-Check Exercise 8.4

Ans. Q1. Refer to Section 8.6.1

Ans. Q2. Refer to Section 8.6.2

Self-Check Exercise 8.5

Ans. Q1. Refer to Section 8.7

8.11 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.

8.12 TERMINAL QUESTIONS

Q1. What is Time Series? What is the need for analysis of Time Series

Q2. Explain the components of Time Series?

Unit-09

TIME SERIES: MEASUREMENT OF TRENDS-I

STRUCTURE

9.1 Introduction

9.2 Learning Objectives

9.3 Measurement of Trend

9.3.1 Free Hand Curve Method

9.3.1.1 Merits of Free Hand Curve Method

9.3.1.2 Demerits of Free Hand Curve Method

Self-Check Exercise 9.1

9.3.2 Semi Average Method

9.3.2.1 Merits of Semi Average Method

9.3.2.2 Demerits of Semi Average Method

Self-Check Exercise 9.2

9.3.3 Moving Average Method

9.3.4 Least Square Method

9.4 Summary

9.5 Glossary

9.6 Answers to Self-Check Exercise

9.7 References/Suggested Readings

9.8 Terminal Questions

9.1 INTRODUCTION

In the last unit, we have studied about the meaning of Time Series. Its utility and components have been also studied. In this unit, we will be able to understand the calculation of Trend values under Free hand Curve Method and Semi Averages Method with suitable graphs and examples with solutions.

9.2 LEARNING OBJECTIVES

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

- Compute trend using Free Hand Curve Method and Semi Averages Method, and
- Discuss the merits and demerits of these methods.

9.3 MEASUREMENT OF TREND

A time series refers to a collection of statistical data recorded sequentially based on the time of occurrence. It consists of observations taken at specific time intervals, which are typically, but not necessarily, equal. In essence, any dataset that varies over time is known as a time series. One of its key components is the secular trend, which represents the long-term direction of the data, whether it exhibits growth, decline, or remains stable over an extended period. The trend component in a time series is commonly analyzed and measured using four primary methods.

9.3.1 Free Hand Curve Method

9.3.2 Semi Average Method

9.3.3 Moving Average Method

9.3.4 Least Square Method

We are going to discuss the graphic methods of measuring trends, i.e. Free Hand Curve and Semi Average Methods in this unit. Moving Average Method and Least Square Method will be discussed in the next unit.

9.3.1 Free Hand Curve Method

The Free-Hand Curve Fitting Method is one of the simplest and most adaptable techniques for estimating secular trends. This method involves plotting time series data on graph paper to create a histogram and then drawing a smooth, free-hand curve through the points. The goal is to accurately capture the long-term trend while eliminating seasonal, cyclical, and random variations. To ensure an appropriate trend line or curve, the following guidelines should be followed:

(i) The curve should be smooth.

(ii) The number of points above and below the trend line should be approximately equal.

(iii) The total vertical deviations of points above the trend line should roughly balance those below it.

(iv) The sum of the squared vertical deviations from the trend line should be minimized.

(v) If cyclical patterns are present in the data:

  a. The trend line should have an equal number of cycles above and below it.

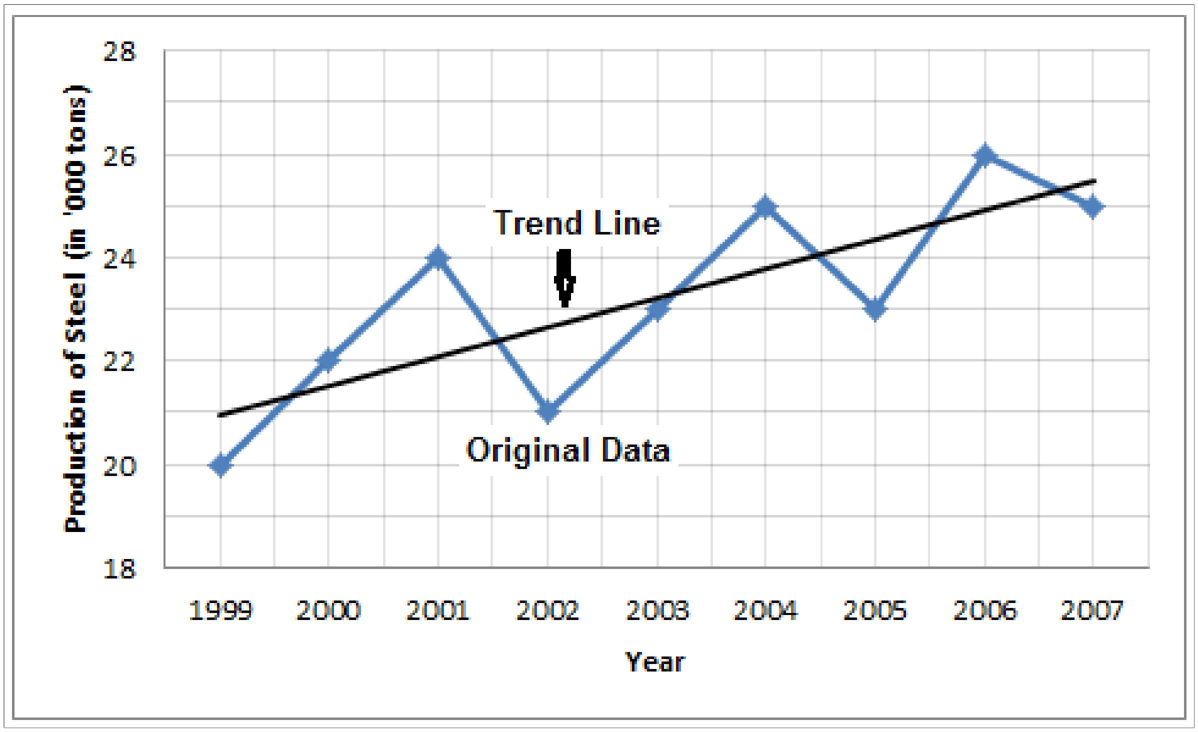
  b. It should bisect the cycles so that the areas above and below the trend line are nearly equal.

(vi) Short-term fluctuations and abrupt variations should be disregarded.

This method can be better understood through an illustrative example.

Example 1. Fit a trend line to the following data by the free hand method:

Years	Production of Steel in '000 tons	Years	Production of Steel in '000 tons
1999	20	2004	25
2000	22	2005	23
2001	24	2006	26
2002	21	2007	25
2003	23		



The upward straight line in the figure shows the required trend line and other line shows the original data.

9.3.1.1 Merits of Free Hand Curve Method

(i) This method is straightforward, time-efficient, and does not require complex mathematical computations.

(ii) It is highly adaptable, allowing for the representation of both linear and non-linear trends.

(iii) It can be useful in making future projections.

9.3.1.2 Demerits of Free Hand Curve Method

(i) The method is largely subjective, as the trend curve depends on the analyst's personal judgment and bias, leading to variations in results when used by different individuals.

(ii) It does not provide a precise measurement of trends.

(iii) Due to its subjective nature, relying on the free-hand curve for forecasting or predictions may lead to inaccurate conclusions.

Self-Check Exercise 9.1

Q1. Explain Free Hand Curve Methods.

Q2. Write down merits and demerits of Free Hand Curve Method.

Q3. Fit a trend line of the following data by Free Hand Curve Method

Year	2015	2016	2017	2018	2019	2020	2021	2022
Profit (Rs. in Lakhs)	80	82	27	86	90	72	75	84

9.3.2 Semi Average Method

Compared to the graphic method, this approach is more objective. It involves dividing the data into two parts, preferably equal, and calculating the average for each segment. These averages are then plotted as points on graph paper, corresponding to the midpoint of the respective time intervals. A straight line connecting these points represents the trend line. As in previous cases, the vertical distance of the trend line from the horizontal axis indicates the trend values. The rationale behind this method is that if the actual trend follows a straight line, the results will be reasonably accurate. While this method is straightforward, indiscriminate use may lead to unreliable outcomes. If the ratio of consecutive time series values is approximately constant, applying this method to the logarithms of the values is recommended. In such cases, trend values can be derived by taking the antilogarithm of the trend line's vertical distance from the horizontal axis.

The semi average method can be applied in case of two situations:

(i) **When the number of years is even**

(ii) When the number of years is odd

(i) When the number of years is even

When the number of years in the series are even like 4, 6, 8, etc., then the whole series can be divided into two equal parts. The following example would make it clear.

Example 2 Fit a trend by the method of semi average to the data given below

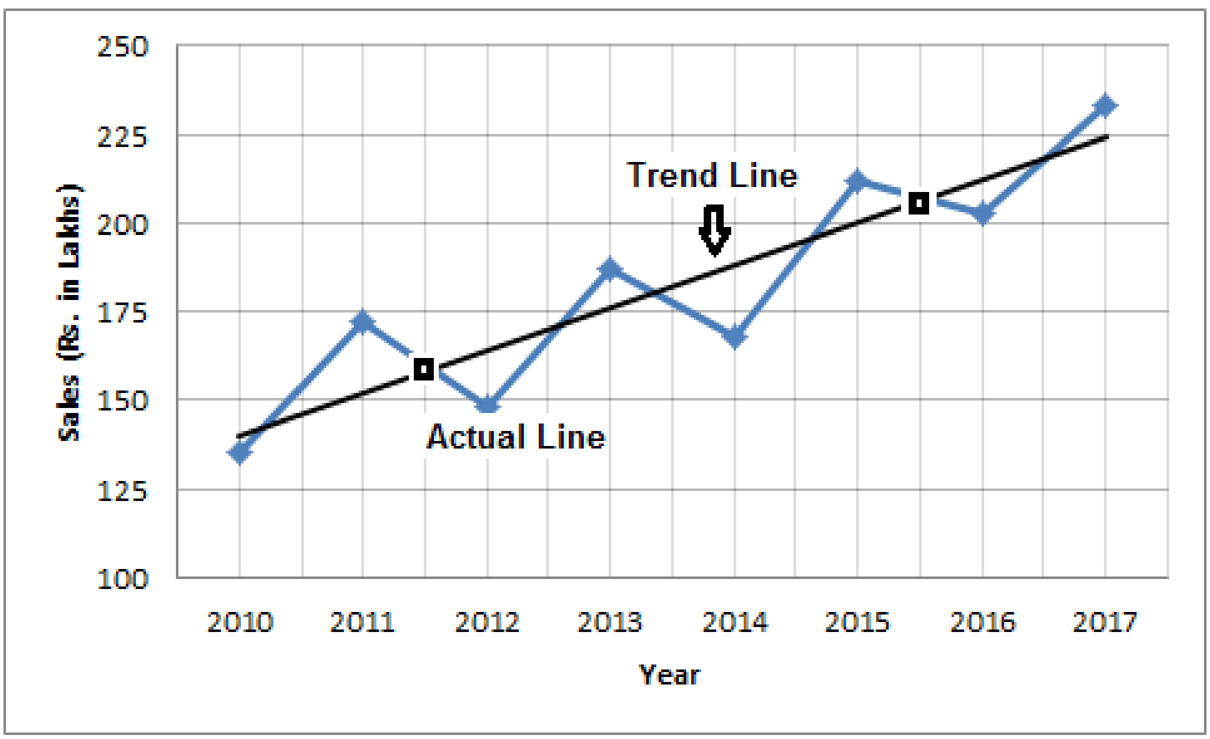
Year	2010	2011	2012	2013	2014	2015	2016	2017
Sales (Rs. in Lakhs)	135	172	148	187	168	212	203	233

Solution:

The number of observations are 'even', i.e., 8. The series must be divided into two equal parts (first 4 and last 4) 1st part will be from 2010 to 2013 and the last part will be from 2014 to 2017. Averages must be calculated for the two parts separately.

Year	Sales (Rs. in Lakhs)	Semi-total	Semi-Average	Middle Year
2010	135			
2011	172	→ 640	$\bar{X}_1$ =642/4 =160.5	→ 2011.5
2012	148			
2013	187			
2014	168			
2015	212	→ 816	$\bar{X}_2$=816/4 = 204	→ 2015.5
2016	203			
2017	233			

The value 160.5 is plotted against the middle of the first four years (2010 to 2013), i.e., 20. The value 204 is plotted against the middle of the last four years (2014 to 2017), i.e., 2015.5. Then both the points are joined by a straight line as under.



(ii) When the number of years is odd

When the number of years in the series are odd like 5, 7, 9, etc. In such a case the dividing the series into two equal parts becomes a problem, in such a case the middle year can be dropped .The following example would make it clear

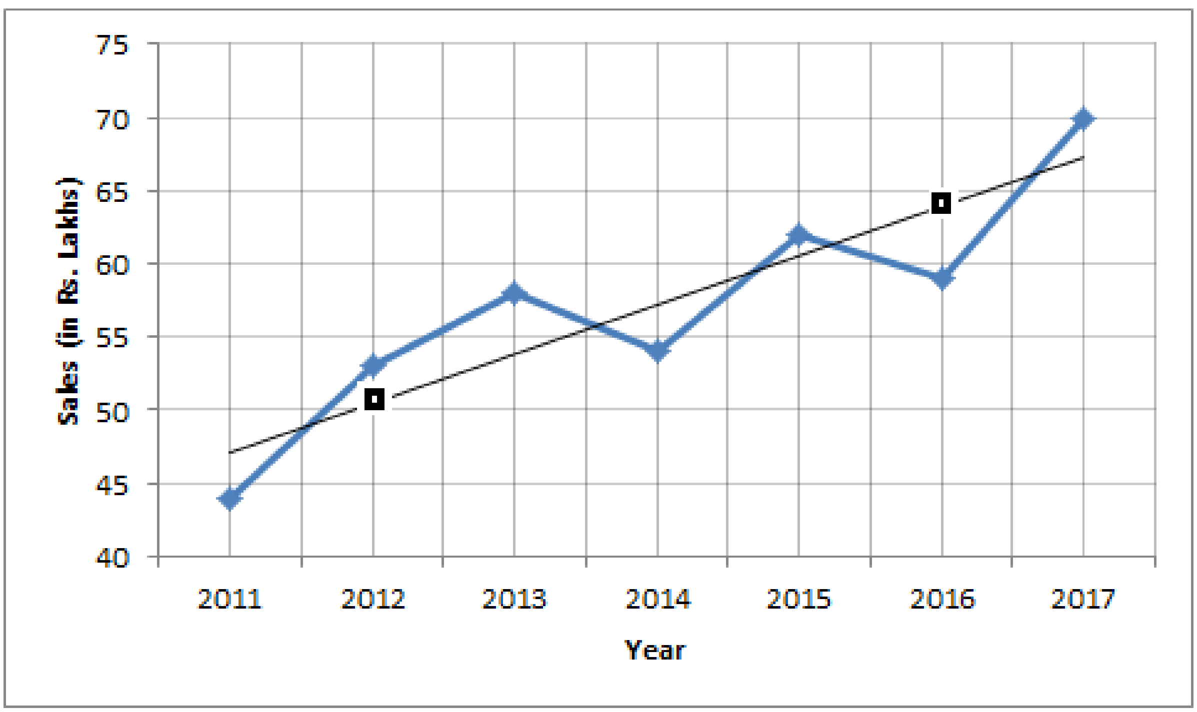
Example 3: Fit a trend line by the method of semi-averages to the data given below:

Year	2011	2012	2013	2014	2015	2016	2017
Sales (in Rs. Lakhs)	44	53	58	54	62	59	70

Solution:

Since there are 7 years , the middle year 1984 will be left out and the arithmetic average of the two parts will be calculated as given below:

Year	Profit	Semi-total	Semi-average	Middle year
2011	44			
2012	53	→155	$\bar{X}_1$ =155/3 =51.67	→2012
2013	58			
2014	54	Omitted		
2015	62			
2016	59	→191	$\bar{X}_2$ =191/3 =63.67	→2016
2017	70			



9.3.2.1 Merits of Semi Average Method

(i) One of the key advantages of this method is its objectivity, as it does not rely on personal judgment. This ensures that all users derive the same trend line and trend values.

(ii) It is simpler to comprehend and implement compared to other trend measurement methods, such as the moving average or least squares method.

(iii) The trend line obtained through this method can be extrapolated in both directions, making it useful for predicting past or future values.

9.3.2.2 Demerits of Semi Average Method

(i) This method operates under the assumption that a linear trend exists in the time series data, which may not always be the case.

(ii) The reliance on the arithmetic mean to calculate semi-averages can be problematic due to its inherent limitations.

Self-Check Exercise 9.2

Q1. Explain Semi Average Methods to fit trend line.

Q2. Fit a trend line of the following data by Semi Average Method

Year	2015	2016	2017	2018	2019	2020	2021	2022
Profit (Rs. in Lakhs)	26	30	24	34	38	32	38	42

Q3. Fit a trend line of the following data by Semi Average Method

Year	2011	2012	2013	2014	2015	2016	2017
Sale (Rs. in Lakhs)	80	82	85	70	89	95	98

9.4 SUMMARY

In this unit, we have learned about the Free Hand Curve Method and Semi Average Method of fitting trend of the time series. In the next unit, we will discuss the Moving Average Method and Least Square Method.

9.5 GLOSSARY

- **Time Series:** A time series is a series of statistical data recorded in accordance with their time of occurrence. It is a set of observations taken at specified times, usually (but not always) at equal intervals. Thus, a set of data depending on the time is called time series.
- **Cyclical Variations:** Cyclical Variations are oscillatory movements in a time series with period of oscillation greater than one year.
- **Irregular Variations:** Irregular Variation or Random movements are such variations which are completely unpredictable in character. These powerful

variations are caused usually by factors which are either wholly unaccountable or caused by such unforeseen events as war, flood, strikes and lockouts, etc.

- **Seasonal Variation:** These fluctuations in a time series occur due to recurring patterns that follow a regular and periodic cycle within a year, typically spanning less than 12 months. These variations tend to repeat in a consistent manner each year and are observed when data is recorded at intervals such as quarterly, monthly, weekly, daily, or even hourly.
- **Secular Trend:** The long-term movement in a time series that shows a consistent pattern of increase, decrease, or stability over an extended period is referred to as the secular trend or simply the trend.

9.6 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 9.1

Ans. Q1. Refer to Section 9.3.1

Ans. Q2. Refer to Section 9.3.1.1 and 9.3.1.2

Ans. Q3. Refer to Section 9.3.1 (Example 1)

Self-Check Exercise 9.2

Ans. Q1. Refer to Section 9.3.2

Ans. Q2. Refer to Section 9.3.2 (Example 2)

Ans. Q3. Refer to Section 9.3.2 (Example 3)

9.7 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.

9.8 TERMINAL QUESTIONS

Q1. Explain the Free Hand Curve Method along with its merits and demerits.

Q2. Explain the Semi Average Method along with its merits and demerits.

Q3. Fit a trend line through Free Hand Curve Method

Year	1982	1983	1984	1985	1986	1987	1988
Sales	110	115	130	140	145	160	180

Q4. Fit a trend line by the method of Semi- Average

Year	2005	2006	2007	2008	2009	2010	2011	2012
Sales	40	45	43	48	52	47	44	48

Q5. Fit a trend line by the method of Semi- Average

Year	2015	2016	2017	2018	2019	2020	2021
Production	40	45	43	48	52	47	44

Unit-10

TIME SERIES: MEASUREMENT OF TRENDS-II

STRUCTURE

10.1 Introduction

10.2 Learning Objectives

10.3 Measurement of Trend

10.3.1 Free Hand Curve Method

10.3.2 Semi Average Method

10.3.3 Moving Average Method

10.3.3.1 Merits of Moving Average Method

10.3.3.2 Demerits of Moving Average Method

Self-Check Exercise 10.1

10.3.4 Least Square Method

10.4 Summary

10.5 Glossary

10.6 Answers to Self-Check Exercise

10.7 References/Suggested Readings

10.8 Terminal Questions

10.1 INTRODUCTION

In the last unit, we have learnt about the first two methods of trend, i.e., Free Hand Curve Method and Semi Average Method. In this unit, we will discuss the third method, i.e., Moving Average Method. Trend values will be estimated by using Moving Average Method with suitable examples. The merits and demerits of Moving Average Method will also be discussed in this unit.

10.2 LEARNING OBJECTIVES

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

- understand the meaning of Moving Average Method,
- computation of trend values under Moving Average Method with suitable examples, and
- discuss the merits & limitations of Moving Average Method.

10.3 MEASUREMENT OF TREND

A time series refers to a collection of statistical data recorded in chronological order, based on their time of occurrence. It consists of observations taken at specific time intervals, which are typically equal but not necessarily so. Any dataset influenced by time is classified as a time series. One of its key components is the secular trend, which represents the long-term movement of data, indicating a general pattern of growth, decline, or stability over an extended period. The trend component in a time series is commonly analyzed and measured using four primary methods.

10.3.1 Free Hand Curve Method

10.3.2 Semi Average Method

10.3.3 Moving Average Method

10.3.4 Least Square Method

We are going to discuss the moving Average Method of measuring trends.

10.3.3 MOVING AVERAGE METHOD

Moving Average Method is normally used method. Under this method, the average value for a number of years is secured, and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average. The effect of the coverage is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend. Moving averages are calculated for 3years or 4years or 5years or 6years or 7 years or 8years or 9 years and so on depending upon the size of the items. The method of calculation of these averages will be as under.

3 years Moving Averages: $\frac{A+B+C}{3}$; $\frac{B+C+D}{3}$; $\frac{C+D+E}{3}$; $\frac{D+E+F}{3}$; $\frac{F+G+H}{3}$

4 years Moving Averages: $\frac{A+B+C+D}{4}$; $\frac{B+C+D+E}{4}$; $\frac{C+D+E+F}{4}$; $\frac{D+E+F+G}{4}$; $\frac{E+F+G+H}{4}$

5 years Moving Averages: $\frac{A+B+C+D+E}{5}$; $\frac{B+C+D+E+F}{5}$; $\frac{C+D+E+F+G}{5}$; $\frac{D+E+F+G+H}{5}$

6 years Moving Averages: $\frac{A+B+C+D+E+F}{6}$; $\frac{B+C+D+E+F+G}{6}$; $\frac{C+D+E+F+G+H}{6}$

Under this method, the trend values can be obtained by employing arithmetic means of the series except at the two ends of the series. The procedure of averaging the 3 items or 4 items or different items simplifies the analysis and removes variations in the values for a period concerned. The moving averages may form a straight line trend or a curve.

The calculation of moving averages in case of odd number of years is very simple as the average can be placed at the centre of the period. In case of even number of years, the averages fall in between the two years or periods. This again requires to be adjusted in such a way that the averages shall be placed against the respective years. So we can have the arithmetic mean of the two averages that shifts the values to the respective years. This process of shifting the figures is "Centering " the averages.

In the case of 3 years moving averages, take the total of first three items (A, B and C) and put the total and the average against the 2^{nd} item (B). Then take the total of next three items (B, C and D) and put the total and average against the third item (C). This procedure is continued till the end. The first and last items remain without the average figures. In the case of 4 years moving averages, take the totals of first four items (A,B,C & D) and put the total and average in between 2nd and 3rd items (between B & C). Then continuing the next process, put the total and averages in between 3rd & 4th items (between C & D). In this way we have the 4 years moving averages which are laying between the two periods. This needs the procedure of shifting or centering the averages. If we simply obtain the means of the two averages, the new averages will be centered exactly against the periods mentioned.

As already mentioned above that moving average is studied under the situations, when the number of years are odd and when the number of years is even we discuss below both of them:

(a) Odd Period Moving Average
(b) Even Period Moving Average

(a) Odd Period Moving Average

The computation procedure of moving average in case of odd years say 3, is illustrated with the following example.

Example 1. Find 3 yearly moving averages from the following time series data

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Exp.	5	8	6	10	11	15	16	18	20	22	25	30

Solution: Calculation of three yearly moving average

Year	Expenditure	3 Yearly Moving Total	3 Yearly Moving Average
1995	5	---	---
1996	8	5+8+6 = 19	19/3=6.33
1997	6	8+6+10 = 24	24/3=8
1998	10	6+10+11=27	27/3=9
1999	11	10+11+15=36	36/3=12
2000	15	11+15+16=42	42/3=14
2001	16	15+16+18=49	49/3=16.33
2002	18	16+18+20=54	54/3=18
2003	20	18+20+22=60	60/3=20
2004	22	20+22+25=67	67/3=22.33
2005	25	22+25+30=77	77/3=25.67
2006	30	---	---

Examples 2: Find 5 yearly moving averages from the following data.

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Exports (in Rs. Crore)	64	76	84	88	92	100	96	88	104	104	108	104

Solution: Calculation of five yearly moving average

Year	Exports (in Rs. Crore)	5 Yearly Moving Total	5 Yearly Moving Average
2001	64	---	---
2002	76	---	---
2003	84	64+76+84+88+92=404	404/5= 80.80
2004	88	76+84+88+92+100=440	440/5= 88.00
2005	92	84+88+92+100+96= 460	460/5= 92.00
2006	100	88+92+100+96+88=464	464/5= 92.80
2007	96	92+100+96+88+104= 480	480/5= 96.00
2008	88	100+96+88+104+104=492	492/5= 98.40
2009	104	96+88+104+104+108=500	500/5= 100.00
2010	104	88+104+104+108+104= 508	508/5= 101.60
2011	108	---	---
2012	104	---	---

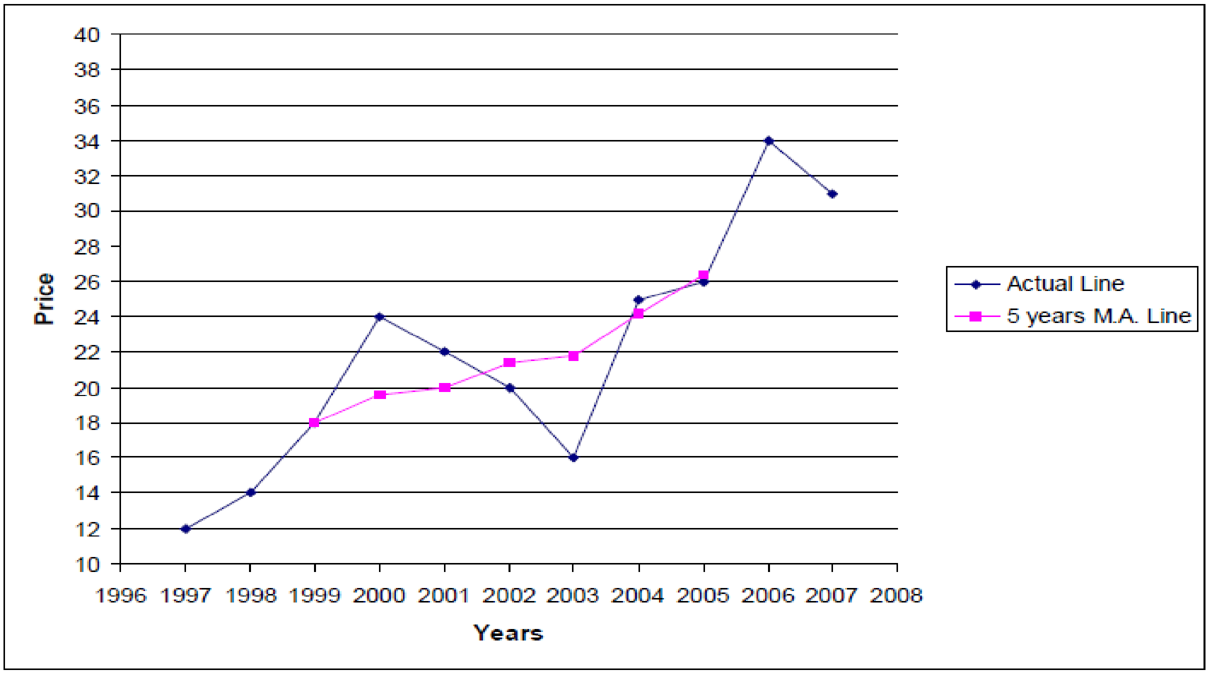
Examples 3: Calculate the trend values by five yearly moving average method and plot the same on a graph from the following.

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Price	12	14	18	24	22	20	16	25	26	34	31

Solution: Calculation of five yearly moving average

Year	Price	5 Yearly Moving Total	5 Yearly Moving Average
1997	12	----	----
1998	14	----	----
1999	18	12+14+18+24+22=90	90/5=18.00
2000	24	14+18+24+22+20=98	98/5=19.60
2001	22	18+24+22+20+16=100	100/5=20.00
2002	20	24+22+20+16+25=107	107/5=21.40
2003	16	22+20+16+25+26=109	109/5=21.80
2004	25	20+16+25+26+34=121	121/5=24.20
2005	26	16+25+26+34+31=132	132/5=26.40
2006	34	----	----
2007	31	----	----

Graph showing Price of product in llYears and trend values of 5 Years.



(b) Even Period Moving Average

The computation procedure of moving average in case of even years say 4, is illustrated with the following example:

Example 4. Find the trend of the following time series by the method of moving average

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Sale	53	79	76	66	69	94	105	88	80	104	98	96	102	106

Solution:

Year	Sale	4-yearly moving total	2 period moving total of 4 yearly moving totals	4-yearly moving average (centered) (Trend Values)
1989	53	---	---	---
1990	79	---	---	---
		53+79+76+66=274		
1991	76		274+290=564	564/8=70.50
		79+76+66+69=290		
1992	66		290+305=595	595/8=74.37
		76+66+69+94=305		
1993	69		305+334=639	639/8=79.87
		66+69+94+105=334		

1994	94		334+356=690	690/8=86.25
		69+94+105+88=356		
1995	105		356+367=723	723/8=90.37
		94+105+88+80=367		
1996	88		367+377=744	744/8=93.00
		105+88+80+104=377		
1997	80		377+370=747	747/8=93.37
		88+80+104+98=370		
1998	104		370+378=748	748/8=93.50
		80+104+98+96=378		
1999	98		378+400=778	778/8=97.25
		104+98+96+102=400		
2000	96		400+402=802	802/8=100.25
		98+96+102+106=402		
2001	102	---	---	---
2002	106	---	---	---

10.3.3.1 Merits of Moving Average Method

(i) This method is simple as compared to the method of least square

(ii) It is a flexible method of measuring trend for the reason that if few more figures are added to the data, the entire calculations are not changed. We only get some more trend values.

(iii) If the period of moving average happens to coincide with the period of cyclical fluctuations in the data, such fluctuations are automatically eliminated.

(iv) The moving average has an advantage that it follows the general movement of data and that its shape is determined by the data rather than the statistician's choice of mathematical function

(v) It is particularly effective if the trend of a series is very irregular.

10.3.3.2 Demerits of Semi Average Method

(i) Trend values cannot be computed for all the years. The longer the period of moving average, the greater the number of years for which trend values cannot be obtained.

(ii) Great Care has to be exercised in selecting the period of moving average.

(iii) Since the moving average is not represented by a mathematical function this method cannot be used in forecasting which is one of the main objectives of trend analysis.

(iv) When the trend situation is not linear, the moving average lies either above or below the true sweep of the data.

Self-Check Exercise 10.1

Q1. Explain the meaning of Moving Average Method.

Q2. What are the merits of Moving Average Method.

Q3. What are the demerits of Moving Average Method.

10.4 SUMMARY

In this unit, we have learned about the Moving Average Method of estimating trend of the time series. Moving average is a simple indicator to understand the average movement of a data set. The worth of moving average as a technical indicator is amplified when identifying areas of support and resistance. Moving averages, though used widely, may not rightfully gauge movement every time. Therefore, it is paramount to understand the various nuances of moving averages and use them along with other technical indicators for a better trade. In the next unit, we will discuss the Least Square Method.

10.5 GLOSSARY

- **Time Series:** A time series is a series of statistical data recorded in accordance with their time of occurrence. It is a set of observations taken at specified times, usually (but not always) at equal intervals. Thus, a set of data depending on the time is called time series.
- **Secular Trend:** It refers to the overall pattern or direction in a time series data, indicating a consistent increase, decrease, or stability over an extended period.
- **Moving Average Method:** A statistical technique used to examine data trends by calculating the average of different subsets within a larger dataset, helping to smooth fluctuations and identify underlying patterns.

10.6 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 10.1

Ans. Q1. Refer to Section 10.3.3

Ans. Q2. Refer to Section 10.3.3.1

Ans. Q3. Refer to Section 10.3.3.2

10.7 SUGGESTED READING

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.

- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor, V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.

10.8 Terminal Questions

Q1. Explain the Moving Average Method along with its merits and demerits.

Q2. Calculate 3 yearly moving average to determine the trend value. Plot the data on a graph paper.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Price	50	52	55	53	58	59	62	67	64	63	58

Q3. Calculate the 4 yearly moving average from the following data and show it on a graph Paper

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2015
Price	45	75	80	96	52	58	76	84	98	103	113

Unit-11

TIME SERIES: MEASUREMENT OF TRENDS-III

STRUCTURE

11.1 Introduction

11.2 Learning Objectives

11.3 Measurement of Trend

11.3.1 Free Hand Curve Method

11.3.2 Semi Average Method

11.3.3 Moving Average Method

11.3.4 Least Square Method

Self-Check Exercise 11.1

11.4 Applications in Economics

Self-Check Exercise 11.2

11.5 Summary

11.6 Glossary

11.7 Answers to Self-Check Exercise

11.8 References/Suggested Readings

11.9 Terminal Questions

11.1 INTRODUCTION

In this preceding units, we learnt about Time Series, utility of Time Series and Components of Time Series. We also studied the different method of measuring trend. In the last unit, we had explained the Free Hand Curve Method, Semi Average Method and Moving Average Method of measuring trend. In this unit we will learn about the last method, i.e., Least Square Method of measuring trend. Applications of Time Series in Economics will also be discussed in this unit.

11.2 LEARNING OBJECTIVES

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

- Compute trend using Least Square Method
- Explain applications of Time Series in Economics
- Solve numerical problems of Time Series Analysis

11.3 MEASUREMENT OF TREND

A time series refers to a collection of statistical data recorded sequentially based on their time of occurrence. It consists of observations taken at specific time intervals, which are typically, but not necessarily, equal. Any dataset that varies over time is classified as a time series. One of its key components is the secular trend, which represents the overall direction of the data over an extended period. This trend may indicate a consistent increase, decrease, or stability in the data over time. The trend component in a time series is typically analyzed and measured using four primary methods.

10.3.1 Free Hand Curve Method

10.3.2 Semi Average Method

10.3.3 Moving Average Method

10.3.4 Least Square Method

We are going to discuss the Least Square Method of measuring trends.

10.3.4 LEAST SQUARE METHOD

This technique is widely utilized in practice as a mathematical approach for fitting a trend line to data. It ensures that the following two conditions are met:

(i) The total sum of deviations between the actual values of Y and the estimated values (Ye) equals zero, i.e., $\sum(Y - Ye) = 0$.

(ii) The sum of the squared deviations between the actual and estimated values is minimized, i.e., $\sum(Y - Ye)^2$ is the least. This minimization property gives the method its name—the method of least squares. The resulting line from this approach is referred to as the line of best fit.

The least squares method can be applied to fit various types of trends, including a straight-line trend, a parabolic trend, or an exponential trend. A straight-line trend follows the equation:

$$Y = a + bX$$

In time series analysis, Y represents the trend values, distinguishing them from the actual observed Y values. The parameter a denotes the Y-intercept, indicating the estimated trend value of Y when X = 0, while b signifies the slope of the trend line, reflecting the rate of change in Y for each unit increase in X. Here, X represents time. When applying the least squares method to fit a straight-line trend, three key aspects must be considered:

(i) Which year was selected as the origin?

(ii) What is the unit of time represented by *X*? Is it half year, one year or five years?

(iii) In what kind of unit is *Y* measured? Is it production in tones, sales in rupees, prices in rupees, employment in thousands of workers?

In order to determine the value of constants *a* and *b* the following two normal equations are to be solved:

$$\sum Y = Na + b\sum X \quad \text{.........(1)}$$
$$\sum XY = a\sum X + b\sum X^2 \quad \text{.........(2)}$$

Where *N* represents the number of years for which the data are given

After determining the equations *Y = a +bX*, we find the trend values related to different years and plot them on the graph paper which show a straight line trend. There are two methods of computing straight line trend by using lest square method:

(i) Direct Method
(ii) Short Cut Method

(i) Direct Method

The procedure to compute straight line trend is as follows:

(i) Any year is taken as the year of origin. Usually the first year or before that is taken as zero, deviations of the other years are marked on 1,2,3…., etc. Time deviations are denoted by X:

(ii) Then ∑X, ∑Y, ∑XY and $\sum X^2$ are computed.

(iii) The values computed are put in the following equations:

$$\sum Y = Na + b\sum X \quad (1)$$
$$\sum XY = a\sum X + b\sum X^2 \quad (2)$$

The values of a and b are determined by solving the above two equations.

(iv) Finally, the calculated values of a and b are put in Y= a+bX and trend values are computed.

The following example will make it more clear:

Example 1. Fit a straight line trend by the method of least square (taking 1978 as year of origin) to the following data:

Year	1979	1980	1981	1982	1983	1984
Production	5	7	9	10	12	17

Solution:

Fitting of Straight Line Trend

Year	Production	Deviations from 1978 (X)	XY	X^2
1979	5	1	5	1
1980	7	2	14	4
1981	9	3	27	9
1982	10	4	40	16
1983	12	5	60	25
1984	17	6	102	36
N=6	**∑Y=60**	**∑X=21**	**∑XY=248**	**$\sum X^2=91$**

The straight line trend is define by the equation:

$$Y=a+bX$$

Two normal equations are

$$\sum Y = Na + b\sum X \quad \text{(i)}$$
$$\sum XY = a\sum X + b\sum X^2 \quad \text{(ii)}$$

Substituting the value, we get

$$60=6a+21b$$

$$248=21a+91b$$

Solving the two equations (i) and (ii)
Multiplying (i) by 7 and (ii) by 2 and then subtracting

$$420=42a+147b$$
$$496=42a+182b$$
$$- \quad - \quad -$$
$$\overline{-76= \quad -35b}$$

$$b=\frac{-76}{-35}=2.17$$

By substituting the value of 'b' in equation (i), we get

$$60=6a+21b$$
$$60=6a+21b(2.17)$$
$$6a=14.43$$
$$a=2.40$$

Hence, the trend equation is
Y=2.40+2.17X; origin= 1978, X unit=1Year.

(ii) Short Cut Method

The process of computation in this method to find straight line trend is as follows:

(i) First of all, middle year is taken as origin with value zero and deviations for other years are computed. Sum of the deviations will be zero, i.e., $\sum X = 0$. Since deviations above and below middle year will balance out.

(ii) $\sum Y$, $\sum XY$, and $\sum X^2$ are computed.

(iii) For computing the values of *a* and *b*, we need not have normal equations but they are found by the following formula:

$$a=\frac{\sum Y}{N}\ ;\ b=\frac{\sum XY}{\sum X^2}$$

(iv) Finally, the calculated values of *a, b* are put in the equation *Y=a+bX* and with its help, trend values are computed.

Short Cut Method is studied in two cases:

(a) When number of years is odd.
(b) When number of years is even.

(a) When number of years is odd.

When number of years is odd like 5,7,9…., etc then the computation of straight line trend can be illustrated with the following examples:

Example 2: Fit a straight line trend for the following data by the method of least square.

Year	1997	1998	1999	2000	2001
Sales (in Rs. Crore)	70	74	80	86	90

Solution:

Years	Sales	***X=(x-1999)***	***XY***	***X^2***	Trend Values
1997	70	-2	-140	4	69.6
1998	74	-1	-74	1	74.8
1999	80	0	0	0	80.0
2000	86	1	86	1	85.2
2001	90	2	180	4	90.4
N=5	**400**	**0**	**52**	**10**	

The equation of the trend line is Y=a+bX

$$\text{Where } a=\frac{\sum Y}{N}=\frac{400}{5}=80$$

$$b=\frac{\sum XY}{\sum X^2}=\frac{52}{10}=5.2$$

The equation of the trend line Y=a+bX

Y=80+5.2 (*x*-1999)

When x=1997,Y=80+5.2(1997-1999)
=80+5.2x(-2)=69.6
x=1998,Y=80+5.2(1998-1999)=74.8
x=1999,Y=80+5.2(1999-1999)=80
x=2000,Y=80+5.2(2000-1999)=85.2
x=2001,Y=80+5.2(2001-1999)=90.4

(b) When the number of years is even.

When the number of years is even (6,10, 12…., etc.) in such the case, the selection of the middle year becomes a problem. In such case, mean of two middle years is taken as year of origin and corresponding, deviations for the other years are found out. Deviations will be -2.5, -1.5, -0.5, 0.5, 1.5 and 2.5 . To simplify the computation process, these deviations are divided by ½ or multiplied by 2. The remaining steps are the same as before.

Example 3: Fit a straight line trend for the following data by the method of least square.

Year	1996	1997	1998	1999	2000	2001
Production	7	9	12	15	18	23

Solution:

Years	**Production**	$X=\frac{x-1998.5}{0.5}$	***XY***	X^2	**Trend Values**
1996	7	-5	-35	25	6.15
1997	9	-3	-27	9	9.29
1998	12	-1	-12	1	12.43
1999	15	1	15	1	15.57
2000	18	3	54	9	18.71
2001	23	5	115	25	21.85
N=6	**84**	**0**	**110**	**70**	

The equation of the trend line is Y=a+bX

When $\sum X=0$

$$a=\frac{\sum Y}{N}=\frac{84}{6}=14$$

$$b=\frac{\sum XY}{\sum X^2}=\frac{110}{70}=1.57$$

The equation of trend line is

$Y=a+bX = Y=14+1.57X=14+1.57\left(\frac{x-1998.5}{0.5}\right)$

When $x=1996, Y=14+1.57\left(\frac{1996-1998.5}{0.5}\right)=6.15$

$x=1997, Y=14+1.57\left(\frac{1997-1998.5}{0.5}\right)=9.29$

$x=1998, Y=14+1.57\left(\frac{1998-1998.5}{0.5}\right)=12.43$

$x=1999, Y=14+1.57\left(\frac{1999-1998.5}{0.5}\right)=15.57$

$x=2000, Y=14+1.57\left(\frac{2000-1998.5}{0.5}\right)=18.71$

$x=2001, Y=14+1.57\left(\frac{2001-1998.5}{0.5}\right)=21.85$

10.3.4.1 Merits of Least Square Method

(i) This mathematical approach to measuring trends ensures objectivity, eliminating any scope for subjectivity.

(ii) The resulting line from this method is known as the line of best fit because it minimizes the sum of squared deviations and ensures that the sum of positive and negative deviations equals zero. Mathematically, this is represented as $\sum(Y - Ye) = 0$, and $\sum(Y - Ye)^2$ is minimized.

(iii) This method allows trend values to be determined for all time periods within the given data series, unlike other techniques such as the moving average, semi-average, or freehand curve methods, which do not offer this capability.

10.3.4.2 Demerits of Least Square Method

(i) It is crucial to carefully choose the appropriate type of trend curve to be fitted, whether linear, parabolic, or another form. Neglecting this aspect may result in misleading conclusions.

(i) This approach is more complex and time-consuming compared to alternative methods.

(ii) Forecasts rely solely on long-term trends, disregarding the influence of cyclical, seasonal, and irregular variations.

(iii) As a mathematical technique, it lacks flexibility—any additional observation requires recalculating the entire process.

(iv) This method is unsuitable for fitting growth curves such as the Gompertz Curve or Logistic Curve.

Self-Check Exercise 11.1

Q1. Explain the method of computation of trend values under Least Squares Method.

Q2. What are the merits of Least Squares Method ?

Q3. What are the demerits of Least Squares Method?

Q4. Find out the trend values under least squares method from the following data.

Year	2003	2004	2005	2006	2007
Sales (in Rs Crore)	12	18	20	23	27

Estimate the sales for the year 2009

10.4 Applications in Economics

The time series analysis is an integral part of every empirical investigation which aims at describing and modeling the evolution over time of a variable or a set of variables in statistically coherent way. Time Series have received lot of attention in economic literature .The components of time series depicts important type of fluctuations in economic data. One of the main assumptions underlying time series analysis is that the regularities observed in the sample period are not specific to that period, but can be extrapolated into the future. This leads to the issue of forecasting which is a major application of time series analysis.

Forecasting is a technique that can aid in future planning. Time Series is an important tool for prediction. In general there are two common ways of forecasting. These are: qualitative forecasting and quantitative forecasting techniques. Executive opinion, panel judgment, Delphi methods, marketing research, and past analogy methods are some important methods of qualitative forecasting. Quantitative method of forecasting can be broadly divided in to Time Series analysis and casual analysis.

Time Series technique is widely used by firms to formulate their sales strategies, investment decisions and interest rates. Economists use time series to forecast future economic conditions.

Although time series analysis has its roots in the natural sciences and in engineering, several techniques are specific to economics. The application of Time Series in Economics has become increasingly important. Time Series analysis is helpful in making predictions about population, national income, capital formation, etc. Time Series is also widely used in stock market indices, weekly sales figures, Monthly

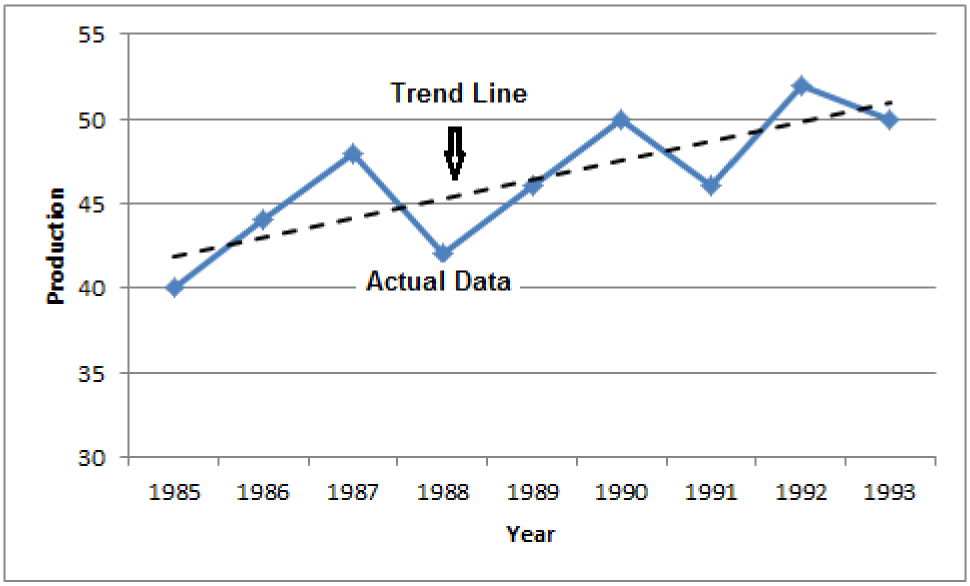
Consumer Price Index, Unemployment rate or the quarterly gross national product. Time Series analysis helps explain change in variables over time by examining the data. It enables us to identify the nature of the phenomenon represented by the sequence of observations and then make predictions or forecast future values. The main objective in analyzing time series is to understand, interpret, and evaluate changes in economic phenomenon in the hope of anticipating the course of future events correctly.

The following examples will illustrate the application of Time Series in Economics.

Example 4. Use the Free Hand Curve Method to fit a trend line to the given data.

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993
Production	40	44	48	42	46	50	46	52	50

Solution:



Example 5: Use the semi-average method to fit a trend line to the given data

Year	1989	1990	1991	1992	1993	1994	1995
Sales of Firm A	112	115	124	120	118	126	122

Solution: Since seven years are given the middle year will be left out and an average of the first three and the last three shall be obtained.

The average of the first three years is:

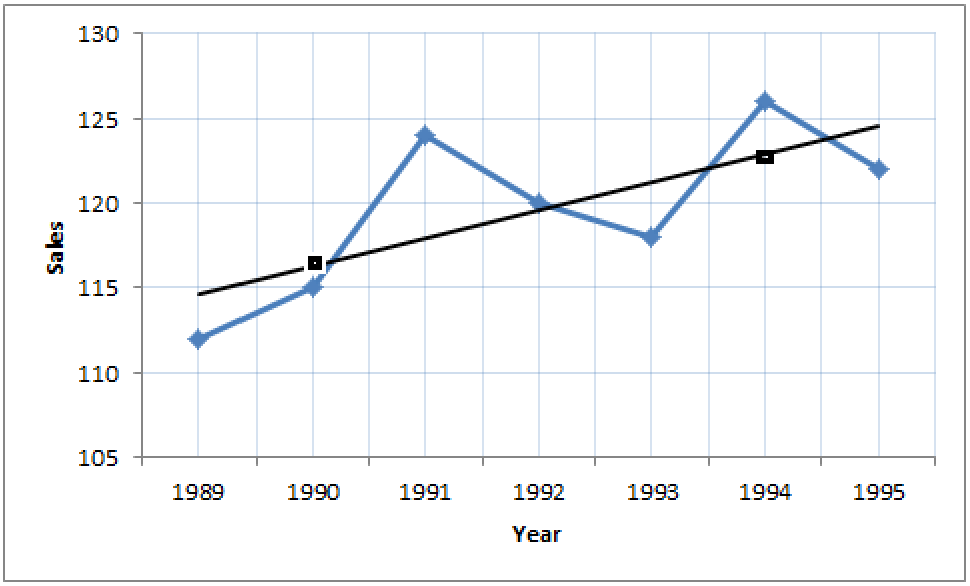
$$\frac{112+115+124}{3} = \frac{351}{3} = 117$$

And the average of the last year is

$$\frac{118+126+1224}{3} = \frac{366}{3} = 122$$

Thus, we get two points 117 and 122 which shall be plotted corresponding to the or respective middle years, i.e., 1990 and 1994. By joining these points we shall obtain the required trend line. The line can be extended and can be used either for prediction or for determining intermediate values.

The actual data and the trend lines are shown in the following graph:



Example 6: Determine the trend values of the following by semi- average method:

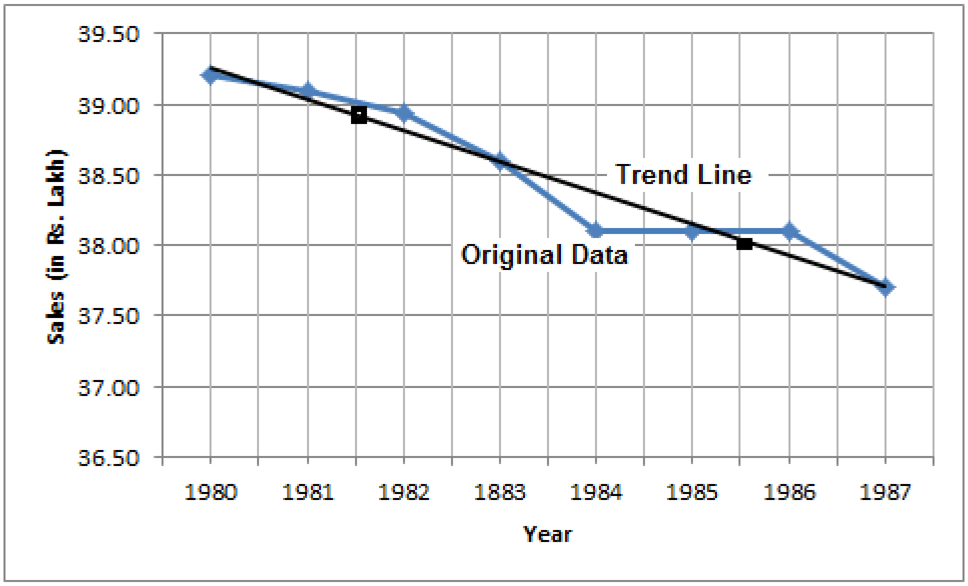
Year	1980	1981	1982	1883	1984	1985	1986	1987
Sales (Lakh Rs.)	39.2	39.1	38.94	38.6	38.1	38.1	38.1	37.7

Solution: Since eight years are given, therefore

Average of first 4 years(1980-1983) $= \frac{39.20+39.10+38.94+38.60}{4} = \frac{155.84}{4}$ = 38.96

Average of last 4 years(1980-1983) $= \frac{38.1+38.1+38.1+37.7}{4} = \frac{152.00}{4}$ = 38.00

These two figures namely, 38.96 and 38, shall be plotted at the middle of their respective period, i.e., (1st July,1981,38.96) and (1st July 1985,38), we shall get the required trend line which describes the given data.



Example 6: From the following Data, calculate trend values using three yearly moving average.

Year	1968	1969	1970	1871	1972	1973	1974	1975	1976	1977
Production in tones	21	22	23	25	24	22	25	26	27	26

Solution:

Three Yearly Moving Average

Year	**Production**	**Three yearly moving total**	**Three yearly moving average**
1968	21	-	-
1969	22	66	22.00
1970	23	70	23.33
1971	25	72	24.00
1972	24	71	23.67
1973	22	71	23.67
1974	25	73	24.33
1975	26	78	26.00
1976	27	79	26.33
1977	26	-	-

Example 7: Calculate trend values using 4-yearly moving average from the following data

Year	1970	1971	1972	1873	1974	1975	1976	1977	1978	1979
Sales (in crore)	7	8	9	11	10	12	8	6	5	10

Solution:

Year	Sales (Crore)	4-yearly moving total	2 period moving Total of 4 yearly moving totals	4-yearly moving average (Trend Values)
1970	7	-		
1971	8	-		
		7+8+9+11=35		
1972	9		35+38=73	73/8=9.125
		8+9+11+10=38		
1973	11		38+42 =80	80/8=10.00
		9+11+10+12=42		
1974	10		42+41=83	83/8=10.375
		11+10+12+8=41		
1975	12		41+36=77	77/8=9.625
		10+12+8+6 =36		
1976	8		36+31=67	67/8=8.375
		12+8+6+5=31		
1977	6		31+29=60	60/8=7.50
		8+6+5+10=29		
1978	5			
1979	10			

Example 8: The sales of a commodity (in'000 of Rs) are given below:

Year	1999	2000	2001	2002	2003	2004	2005
Sales	82	86	81	86	92	90	99

(i) Using the method of least square, fit a straight line trend equation of the data.
(ii) What is the average annual change in the sales?
(iii) Obtain the trend values of the years 1999-2005 and show that the sum of differences between the actual trend and the trend values is equal to zero
(iv) What are the expected sales for the year 2010?

Solution: Let the equation of the straight line be Y_C= a+bX. Since N=7 is odd, therefore, select the middle year, i.e., 2002 as origin.

Fitting a Straight Line Trend by the method of Least Square

Year	*X (Deviation from 2002)*	X^2	*Sales* Y^2	*XY*	*Trend Values* $Y_C=a+bX$	$Y-Y_C$
1999	-3	9	82	-246	80.5	1.5
2000	-2	4	86	-172	83.0	3.0
2001	-1	2	81	-81	85.5	-4.5
2002	0	0	86	0	88.0	-2.0
2003	1	1	92	92	90.5	1.5
2004	2	4	90	180	93.0	-3.0
2005	3	9	99	297	95.5	3.5
N=7	$\sum X=0$	$\sum X^2=28$	$\sum Y=616$	$\sum XY=70$		$\sum(Y-Y_C)=0$

Since $\sum X=0$, therefore the values of $a=\frac{\sum Y}{N}=\frac{70}{28}=2.5$ and $b=\frac{\sum XY}{\sum X^2}=\frac{616}{7}=88$

(i) Hence the straight line trend equation is given by : Y_C=88+2.5X ……………..(I)

[Origin= 2002, *X* unit =1 year. *Y* unit= Annual Sales in '000 of Rs]

(ii) Average annual changes in sales (b)= 2.5

(iii) For *X*=-3, Y_{1999}= 88 + 2.5 (-3) = 88 - 7.5 = 80.5

For *X*=-2, Y_{2000}= 88 + 2.5 (-2) = 88 – 5 = 83

Similarly, by putting *X*= -1,0, 1, 2, 3, in the equation (1), we obtain:

Y_{2001} =88+2.5(-1) = 85.5,

Y_{2002} = 88+2.5 (0),

Y_{2003} =88+2.5(1) =90,

Y_{2004} =88+2.5(2) =93,

Y_{2005} =88 +2.5(3) =95.5.

Since the values of b is constant ,i.e., b=2.5, therefore, after finding the first trend value, adding the values of b to every preceding value, we will get the corresponding trend values of the next years. For example, in the above case for 1999 the calculated value of Y_c is 80.5. For 2000 it will be 80.5 + 2.5= 83 and so on. If b is negative , then instead of adding we will subtract.

(iii) From the table it is clear that $\sum(Y-Y_c)$ =0, i.e., the sum of differences between the actual and the trend values is sequal to zero.

(iv) The expected sales for the year 2010 is obtained by putting *X* = 8 in the equation (I), we obtain: Y_{2010} =88 +2.5(8) =108.

Example 9: Fit a straight line trend by the method of least squares and estimate the sales for 2006.

Year	2000	2001	2002	2003	2004	2005
Sales	12	13	14	15	22	26

Solution: Let the equation of the straight line trend be *Y=a+bX*. Here, n=6 (even).By taking 2002.5, the midpoint of the two middle years 2002 and 2003 as origin and divide by ½.

Fitting a Straight Line Trend by the method of Least Square

Year	*Deviation from 2002.5*	*Deviations Multiplied By 2*	X^2	*Sales (Rs.in lakhs)*	*XY*
2000	-2.5	-5	25	12	-60
2001	-1.5	-3	9	13	-39
2002	-0.5	-1	1	14	-14
2003	0.5	1	1	15	15
2s004	1.5	3	9	22	66
2005	2.5	5	25	26	130
N=6		**∑X=0**	**$\sum X^2$=70**	**∑Y =102**	**∑XY =98**

Since ∑X =0, therefore the values of a = $\frac{\sum Y}{N}$ =$\frac{102}{6}$ = 17 and b = $\frac{\sum XY}{\sum X^2}$ = $\frac{98}{70}$ =1.4

Hence, the annual trend equation is given by : Y_c = *a+bX = 17 +1.4 X*

[Origin = 2005.5 ,X_{unit} = ½ Year, Y_{unit} =Annual sales in lakhs of Rs.]

The estimate sales for the year 2006 is obtained by putting X=7 in equation (1), we obtain:

Y_{2006} = 17+1.4(7) = 17 +9.8 +26.8 lakhs.

Self-Check Exercise 11.2

Q1. Explicate Application of Time Series in Economics?

11.5 SUMMARY

In this unit, we have gone through Least Square Method of measuring trend. We have also discussed the merits and demerits of Least Square Method. We have also understood the applications of Time Series in Economics

11.6 GLOSSARY

- **Time Series:** A time series is a series of statistical data recorded in accordance with their time of occurrence .It is a set of observations taken at specified times, usually(but not always)at equal intervals. Thus a set of data depending on the time is called time series.
- **Secular trend:** The general tendency of the time series data to increase or decrease or stagnate during a long period of time is called the secular trend or simply trend.
- **Least Square Method:** is the process of finding the best-fitting curve or line of best fit for a set of data points by reducing the sum of the squares of the offsets (residual part) of the points from the curve.

11.7 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 11.1

Ans. Q1. Refer to Section 11.3.4
Ans. Q2. Refer to Section 11.3.4.1
Ans. Q3. Refer to Section 11.3.4.2
Ans. Q4. Refer to Section 11.3.4 (Example 1)

Self-Check Exercise 11.2

Ans. Q1. Refer to Section 11.4

11.8 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor, V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.

11.9 TERMINAL QUESTIONS

Q1. Explain the Least Square Method of estimating trend equation. Also discuss its merits and demerits?

Q2. Find out straight line trend values under least squares method. Estimate the values for the year 2010.

Year	2001	2002	2003	2004	2005	2006	2007
Values	100	120	110	140	80	95	115

Q3. Find out the trend values under least squares method

Year	1990	1992	1995	1996	1997	1999	2000
Price (in Rs.)	25	32	40	37	44	50	57

Estimate the price for the year 2003

Unit-12

INDEX NUMBER: AN INTRODUCTION

STRUCTURE

12.1 Introduction

12.2 Learning Objectives

12.3 Index Number

- 12.3.1 Definition of Index Number
- Self-Check Exercise 12.1

12.4 Uses and Limitations of Index Number

- Self-Check Exercise 12.2

12.5 Types of Index Numbers

- 12.5.1 Price Index Number
- 12.5.2 Quantity Index Number
- 12.5.3 Value Index Number
- 12.5.4 Simple and Aggregative Index Number
- 12.5.5 Cost of Living Index Number
- 12.5.6 Special Purpose Index Number
- Self-Check Exercise 12.3

12.6 Problems in the Construction of Index Numbers

- Self-Check Exercise 12.4

12.7 Methods of Constructing Index Number

- 12.7.1 Unweighted or Simple Index Numbers
  - 12.7.1.1 Simple Aggregative Method
  - 12.7.1.2 Simple Average of Price Relatives Method
- Self-Check Exercise 12.5

12.8 Summary

12.9 Glossary

12.10 Answers to Self-Check exercise

12.11 References/Suggested Readings

12.12 Terminal Questions

12.1 INTRODUCTION

Index numbers are designed to analyze changes in factors that cannot be measured directly. As Bowley stated, “Index numbers are used to measure the changes in some quantity which we cannot observe directly.” For instance, while business activity in a country cannot be directly measured, its relative changes can be assessed by examining variations in measurable factors that influence business activity.

Index numbers serve as a widely used statistical tool for assessing overall fluctuations in a group of related variables. When comparing today’s consumer price levels with those from a decade ago, it is not sufficient to analyze a single item; rather, an average price level needs to be considered. Similarly, when comparing agricultural or industrial production over time, all production items must be taken into account, each experiencing varying degrees of increase or decrease. To obtain a consolidated measure of these variations, index numbers are employed. They can be defined as a method for combining the changes in a set of related variables over time to derive a single figure that represents the overall effect of these changes.

Index numbers can be categorized based on the variables they measure. In business and economics, they are commonly used to track (i) price, (ii) quantity, (iii) value, and (iv) business activity. Examples include the wholesale price index, consumer price index, industrial production index, export value index, and business activity index. This discussion will explore the meaning, uses, and limitations of index numbers, their types, and the challenges involved in their construction.

12.2 LEARNING OBJECTIVES

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

- Understand the meaning of index number
- Discuss the uses and limitations of index number
- Know the various methods of constructing index number
- Identify different types of index numbers

12.3 INDEX NUMBER

An index number is a statistical tool used to measure fluctuations in price, quantity, or value of a single item or a group of related items over time, across different locations, or based on other characteristics. Essentially, index numbers serve as a specialized form of rates, ratios, or percentages, providing an overall indication of the relative magnitude of a set of interconnected variables in multiple scenarios.

12.3.1 Definition of Index Number

“Index Numbers are devices for measuring differences in the magnitude of a group of related variables.” ………….Croxton & Cowden

“Index Numbers are a specialized type of averages.”…………. M. Blair

“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.”……………… Spiegel

Self-Check Exercise 12.1

Q1. Define Index Number.

12.4 USES AND LIMITATIONS OF INDEX NUMBER

Uses:

(i) The price index numbers are used to measure changes in a particular group of prices and help us in comparing the movement in prices of one commodity with other.

(ii) A common use of index number is to deflect a future value so that it can be compared with the base period.

(iii) They are also used to forecast business condition of a country.

(iv) It is a statistical device used by the government to revise wages, salaries, pensions, social welfare schemes and design future planning.

(v) They measure the purchasing power of money

(vi) Index numbers of import prices and export prices are used to measure the changes in the trade of a country.

(vii) The index numbers are used to measure seasonal variations and cyclical variations in a time series.

Limitations:

(i) Index numbers measure relative changes indifferent phenomenon. Index numbers are true only on the averages.

(ii) A given type of index number is not suitable for all purposes. Multipurpose index numbers cannot be constructed.

(iii) No consideration is given to the changes taking place in the quality of a commodity while constructing the index numbers.

(iv) Bias in the selection of base year and selection of representatives sometimes leads to misleading results.

(v) Index numbers lack in perfect accuracy because they are mostly constructed on the basis of sample commodities.

Self-Check Exercise 12.2

Q1. What are the uses of Index Number?

Q2. What are the limitations of Index Number?

12.5 TYPES OF INDEX NUMBERS

2.5.1 Price Index Number

Price index**,** numbers measure of relative price changes, consisting of a series of numbers arranged so that a comparison between the values for any two periods or places will show the average change in prices between periods or the average difference in prices between places. These are of two types' wholesale price index and retail price index.

(a) Wholesale Price Index Number:- it measures the changes in the general price level of a commodity.

(b) Retail Price index Number:- it measures the general changes in the retail prices of commodities which are bought and sold in the retail market.

12.5.2 Quantity Index Number

Index numbers are used to assess variations in the physical quantity of goods that are produced, consumed, or sold, whether for a single item or a group of items. This category encompasses indices related to agricultural and industrial production.

12.5.3 Value Index Number

Value index numbers measure the changes in the value of some commodities or group of commodities consumed or purchased in the given period with the reference to base period.

12.5.4 Simple and Aggregative Index Number

On the basis of the number of commodities that enter into the construction of an index, index numbers are classified into two categories:

(a) **Simple Index Number**: When index numbers are constructed for individual commodities, these are termed as simple index numbers.

(b) **Aggregative Index Number**: When index numbers are constructed for a group of commodities, these are known as aggregative index numbers.

12.5.5 Cost of Living Index Number

The cost of living index, also known as the consumer price index, measures the average change in consumer spending and consumption of goods over different time periods. It reflects variations in expenditure patterns for a specific group of consumers.

12.5.6 Special Purpose Index Number

Some index numbers are constructed for some specific purpose .They measure the average change as compared to the base period of any specific purpose.

Self-Check Exercise 12.3

Q1. What is meant by Price Index Number?

Q2. What is meant by Quantity Index Number?

Q3. What is Cost of Living Index Number?

12.6 PROBLEMS IN THE CONSTRUCTION OF INDEX NUMBERS

Several problems/challenges arise when constructing index numbers, and some of the most significant ones are outlined below:

i) Defining the Purpose of the Index Number: The primary challenge in constructing an index number is to clearly define its objective. Establishing a well-defined purpose is crucial, as it determines the type of data to be collected, the statistical methods to be applied, and various other factors, such as the choice of commodities, selection of a base period, and the type of average to be used.

ii) Selection of Commodities or Items: Once the objective of the index number is established, the next challenge is choosing the appropriate commodities or items. The selection process should consider the following factors:

- The commodities chosen should be relevant to the purpose of the index.
- The items should be grouped into relatively homogeneous categories such as food, clothing, and fuel.
- A sufficient number of representative items should be selected from each category.
- The total number of commodities should strike a balance—too few may make the index unrepresentative, while too many may complicate the computation process, making it time-consuming and costly.
- To ensure valid comparisons, selected commodities should maintain consistent quality over time.

iii) Data Collection for Index Numbers: The accuracy of index numbers depends on reliable data regarding the prices and quantities of selected commodities over different periods. Such data should ideally come from credible sources such as standard trade journals, financial newspapers (e.g., *Economic Times*, *The Financial Express*), reputed periodicals, and special reports from producers and exporters. If these sources are unavailable, unbiased field surveys can serve as an alternative. The principles of accuracy, comparability, and adequacy should always be maintained when using secondary data.

iv) Selecting a Base Period: The base period serves as a reference point for comparing changes in the phenomenon over time. The index number for this period is always set to 100. The selection of a base period must consider the following:

- It should represent stable and normal economic conditions, free from irregular disruptions such as wars, famines, or economic crises.
- The base period should not be too far removed from the current period.

- A decision must be made between using a **fixed-base** or **chain-base** index. The fixed-base method keeps the reference year constant, while the chain-base method updates the base periodically.

v) Choosing an Appropriate Average: To represent overall price changes, the values of different commodities must be averaged. Since index numbers function as specialized averages, selecting an appropriate method is essential. Common choices include the Arithmetic Mean (AM) and the Geometric Mean (GM). The Harmonic Mean, Median, and Mode are generally not used in index number calculations.

vi) Assigning Weights to Commodities: Weighting is a critical yet challenging aspect of constructing index numbers. Commodities like food, clothing, and fuel do not contribute equally to overall economic activity, so appropriate weights must be assigned based on their relative importance. There are two types of index numbers based on weighting:

- **Unweighted Index Numbers**, where all commodities are treated equally.
- **Weighted Index Numbers**, where commodities receive different weights based on their significance.

vii) Choosing a Calculation Method: Index numbers can be computed using various techniques, such as the aggregative method and the average of relative method**.** The choice of method depends on the purpose of the index and the availability of data.

By addressing these challenges thoughtfully, an index number can be constructed to accurately reflect economic trends and changes over time.

Self-Check Exercise 12.1

Q1. List out the problems in the construction of Price Index Number.

12.7 METHODS OF CONSTRUCTING INDEX NUMBER

Different methods are used in the construction of index numbers, we discuss them one by one below

12.7.1 Unweighted or Simple Index Numbers

Unweighted or Simple Index Numbers are those in whose construction all commodities are given equal importance. There are two methods to construct them

12.7.1.1 Simple Aggregative Method

12.7.1.2 Simple Average and Price Relative Method

12.7.1.1 Simple Aggregative Method

In this method, the sum of current years is divide by the sum of base year’s price and the quotient is multiplied by 100. The following formula is used:

$$P_{01} = \frac{\sum p_1}{\sum p_0} \times 100$$

Where, $\sum p_1$ =Aggregate of Prices in Current Year

$\sum p_0$ = Aggregate of Prices in Base Year, and P_{01} = Price Index

This method can be illustrated with the following examples:

Example 1: Construct price index number for 1990 based on 1981 using simple aggregative method:

Commodity	Price in 2010 (in Rs)	Price in 2015 (in Rs)
A	50	80
B	40	60
C	10	20
D	5	10
E	2	8

Solution:

Construction of Price Index Number

Commodity	Price in 2010 (p_0)	Price in 2015 (p_1)
A	50	80
B	40	60
C	10	20
D	5	10
E	2	8
	$\sum p_0=107$	$\sum p_1=178$

$$P_{01} = \frac{\sum p_1}{\sum p_0} \text{x}100 \quad = \frac{178}{107} \text{x}100 = 166.36$$

12.7.1.2 Simple Average of Price Relatives Method

In this method first of all, the price relatives of the commodities or items are found out. To compute the price relatives, price in current year(p_1) is divided by the price in the base year(p_0) and then, the quotient is multiplied with 100. In terms of formula,

$$\text{Price Relative} = \frac{\text{Current year's Price}}{\text{Base year's Price}} \text{x}100 \text{ or } P = \frac{p_1}{p_0} \text{x}100$$

After this, using Arithmetic Mean or Geometric Mean or median. we find the averages of price relatives.

(i) When Arithmetic Mean is used, then the following formula is used:

$$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N}$$

Where N=number of items or commodities.

(ii) When Geometric Mean is used then the following formula is used:

$$P_{01} = \text{Antilog}\left(\frac{\sum \log P}{N}\right)$$

$$\text{Where, } P = \frac{p_1}{p_0} x100$$

(iii) When Median is used then the following formula is used:

$$P_{01} = \text{Size of}\left(\frac{N+1}{2}\right)$$

The following example illustrates the procedure of the method:

Example 2: From the data given below construct the price index number for the year 2006 and 2005 as the base year by using (i) Arithmetic Mean (ii) Geometric Mean

Commodity	**A**	**B**	**C**	**D**	**E**
Price in 2005 (Rs)	20	110	80	40	50
Price in 2006 (Rs)	20	120	90	60	70

Solution:

Commodities	**Price in 2005 (p_0)**	**Price in 2006 (p_1)**	**Price Relatives** $P = \frac{p_1}{p_0} x100$	**Log P**
A	20	20	$\frac{20}{20}$X100 =100	2.0000
B	110	120	$\frac{120}{110}$X100 =109.1	2.0378
C	80	90	$\frac{90}{80}$X100 = 112.5	2.0512
D	40	60	$\frac{60}{40}$X100= 150	2.176
E	50	70	$\frac{70}{50}$X100 = 140	2.1461
N=5			$\sum \frac{p_1}{p_0}$ ***x100=611.6***	**∑LogP =10.4112**

(i) Price Index Number based on price relatives using (A.M.) is:

$$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N}$$

$$= \frac{611.6}{5} = 122.32$$

(ii) Price Index Number based on price relatives using (G.M.) is:

$$P_{01} = \text{Antilog}\left(\frac{\sum logP}{N}\right)$$

$$= \text{Antilog}\left(\frac{10.4112}{5}\right)$$

$$= \text{Antilog}\ (2.0822) = 120.9$$

It should be noted that since Arithmetic Mean (A.M.) gives greater weightage to bigger items the index which uses Arithmetic Mean would be higher than index using Geometric Mean.

Merits

(i) Extreme items do not influence the index. Equal importance is given to all items.

(ii) The index is not influenced by the units in which prices are quoted or by the absolute level of individual prices.

Demerits

(i) Difficulty is faced regarding selecting an appropriate average. The use of the arithmetic mean is sometimes questionable because it has an upward bias. The use of geometric mean involves difficulties in comparison.

Self-Check Exercise 12.4

Q1. What are Unweighted or Simple Index Number?

Q2**.** Construct price index number for 1990 based on 1981 using simple aggregative method

Commodity	A	B	C	D	E
Price in 2010 (in Rs)	50	40	10	5	2
Price in 2015 (in Rs)	80	60	20	10	8

12.8 SUMMARY

In this unit, you have been introduced to the concepts and methods involved in the construction of Index Numbers. You have also learnt about the use of index numbers, types of index numbers and problems in construction of Index Number. You have been shown how to use the Unweighted or Simple Index Numbers, Simple Aggregative Method and Simple Average of price relatives Method of calculating numbers.

12.9 GLOSSARY

- **1.Index Number:** is a statistical measure that is used to show changes in price, quantify or value of an item or group of related items with respect to time, place or other characteristics.
- **Base Year:** is a normal year, in terms of variable concerned, base year index is

invariably taken as 100. Current year index is expressed as a percentage of base year index.

- **Price Relative:** for a commodity is the ratio of the current year price to base year price of that commodity
- **Price Index Number:** is a measure of how prices change over a period of time, or in other words, it is a way to measure inflation.
- **Value Index Number:** measures the changes in the value of some commodities or group of commodities consumed or purchased in the given period with the reference to base period.
- **Quantity Index Number:** These index numbers are considered to measure changes in the physical quantity of goods produced, consumed or sold of an item or a group of items. Indices of agriculture and industrial production are included in this category.

12.10 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 12.1

Ans. Q1. Refer to Section 12.3.1

Self-Check Exercise 12.2

Ans. Q1. Refer to Section 12.4

Ans. Q2. Refer to Section 12.4

Self-Check Exercise 12.3

Ans. Q1. Refer to Section 12.5.1

Ans. Q2. Refer to Section 12.5.2

Ans. Q3. Refer to Section 12.5.5

Self-Check Exercise 12.4

Ans. Q1. Refer to Section 12.6

Self-Check Exercise 12.5

Ans. Q1. Refer to Section 12.7.1

Ans. Q2. Refer to Section 12.7.1.1 (Example 1)

12.11 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.

- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor, V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.
- Gupta S.C. (2016). Fundamentals of Statistics, Himalaya Publishing House.

12.12 TERMINAL QUESTIONS

Q1. What is an Index number? Briefly discuss the uses of Index numbers.

Q2. "For constructing index numbers the best method on theoretical grounds is not the best method from practical point of view" Discuss

Q3. What are the difficulties in the construction of Index Numbers. Discuss in detail.

Unit-13

WEIGHTED INDEX NUMBER

STRUCTURE

13.1 Introduction

13.2 Learning Objectives

13.3 Methods of Constructing Index Number

13.3.1 Weighted Index Number

13.3.1.1 Weighted Aggregative Method

(a) Laspeyre's Method

(b) Paasche's Method

(c) Fisher's Method

13.3.1.2 Weighted Average of Price Relative Method

13.3.1.3 Merits of Weighted Index Number

13.3.1.4 Demerits of Weighted Index Number

Self-Check Exercise 13.1

13.4 Tests of Index Numbers

13.4.1 Time Reversal Test

3.4.2 Factor Reversal Test

Self-Check Exercise 13.2

13.5 Summary

13.6 Glossary

13.7 Answers to Self-Check exercise

13.8 References/Suggested Readings

13.9 Terminal Questions

13.1 INTRODUCTION

In this last unit, we have learnt about meaning and uses of Index Numbers. We have also learnt the different types of Index Numbers and the problems that are involved in the construction of Index Numbers. Unweighted or Simple Index Numbers have also been discussed. The present unit will deal with the remaining methods of measuring Index Numbers.

13.2 LEARNING OBJECTIVES

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

- compute the Weighted Price Index Number
- apply the different types of tests of consistency

13.3 METHODS OF CONSTRUCTING INDEX NUMBER

13.3.1 Weighted Index Numbers: Weighted Index Numbers are of two types :

13.3.1.1 Weighted Aggregative Method

13.3.1.2 Weighted Average of Price Relative Method

13.3.1.1 Weighted Aggregative Method

These index numbers are the simple aggregative type with the fundamental difference that weights are assigned to the various items included in the index. There are three methods through which weighted index numbers can be calculated under Weighted Aggregative Method:

(a) Laspeyre's Method

(b) Paasche's Method

(c) Fisher's Method

(a) Laspeyre's Method

Prof. Laspeyre was a French economist who devised a formula in 1817 to construct Index Number. Prof. Laspeyre has assigned weights to the commodities on the basis of base year quantities (q_o). Laspeyre's formula is as follows:

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

b) Paasche's Method

Prof. Paasche was a German statistician who gave the formula in 1874 to construct Index Number. Prof. Paasche has assigned weights to the commodities on the basis of current year quantities (q_1). Paasche's formula is as follows:

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

(c) Fisher's Method

Irving Fisher advocated the geometric mean of Laspeyre's and Paasche's price index number. Fisher has assigned weights to different commodities on the basis of both the base year as well as current year quantities. Fisher's formula is as follows:

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times \mathbf{100}$$

In fact, Fisher's Index is the geometric mean of Laspeyre's and Paasche's Indices, i.e.,

$$P_{01} = \sqrt{L \text{ X } P}$$

Fisher's Index Number as an Ideal Formula

(i) Index number computed by Fisher's method is known as ideal index. This is because:

(ii) It is based in geometric Mean which is the best average considered for computing index numbers.

(iii) This formula provides weighted index number.

(iv) Equal importance is given to the base year and current year prices and quantities.

(v) Fisher's index number satisfies many tests of ideal index like Time Reversal Test and Factor Reversal Test.

Different methods of constructing weighted Index Number can be illustrated with the following example:

Example 1: Compute (i) Laspeyre's; (ii) Paasche's; and (iii)Fisher's price index numbers from the following table:

Items	Price in 2010	Quantity in 2010	Price in 2015	Quantity in 2015
A	10	4	12	6
B	15	6	20	4
C	2	5	5	3
D	4	4	4	4

Solution:

Calculation of Price Index Numbers

Items	p_0	q_o	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	10	4	12	6	40	60	48	72
B	15	6	20	4	90	60	120	80
C	2	5	5	3	10	6	25	15
D	4	4	4	4	16	16	16	16
					$\sum p_0q_0$= 156	$\sum p_0q_1$= 142	$\sum p_1q_0$= 209	$\sum p_1q_1$= 183

(a) Laspeyre's Price Index:

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

$$= \frac{209}{156} \times 100 = 133.97$$

b) Paasche's Price Index:

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

$$= \frac{183}{142} \times 100 = 128.87$$

(c) Fisher's Price Index:

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times \mathbf{100}$$

$$= \sqrt{\frac{209}{156} \times \frac{183}{142}} \times \mathbf{100} \quad \mathbf{= 131.40}$$

13.3.1.2 Weighted Average of Price Relative Method

In weighted Average of relative, the price relatives for the current year are calculated on the basis of the base year price. These price relatives are multiplied by the respective weight of items. These products are added up and divided by the sum of weights. In terms of formula,

Weighted Index Number (p_{01}) = $\frac{\sum PW}{\sum W}$, using A.M.

Where, P = Price Relatives = $\frac{P_1}{P_0}$ x 100 ; W = value in the base year (i.e., p_0q_0) or fixed weights. If the weighted Geometric Mean is used, then the formula is

$$P_{01} = \text{Antilog}\left[\frac{\sum (logP).W}{\sum W}\right]$$

Example 2: Compute Price Index from the following data by applying weighted average of price relative method:

Commodity	Base year price	Base year Quantity	Current year price
A	6.0	40	8.0
B	3.0	80	3.2
C	2.0	20	3.0

Solution:

Commodity	**p_0**	**q_0**	**p_1**	**$W(p_0q_o)$**	$\frac{P_1}{P_0}$ **x 100**	**PW**
A	6.0	40S	8.0	240	$\frac{8}{6}$ x 100= 133.3	31,992
B	3.0	80	3.2	240	$\frac{3.2}{3.0}$x100= 106.7	25,608
C	2.0	20	3.0	40	$\frac{3}{2}$ x 100= 150.0	6,000
				$\sum$W=520		$\sum$PW=63,600

Using Weighted Average of Price Relative Method,

$$P_{01} = \frac{\sum PW}{\sum W} = \frac{63,600}{520} = 122.30$$

13.3.1.3 Merits of Weighted Index Number

(i) **Simplicity and Ease of Computation** – This method is straightforward to understand and simple to calculate.

(ii) **Common Base for Multiple Indices** – It enables the construction of various indices with a shared base in a consolidated manner.

(iii) **Flexibility in Calculation Methods** – Both arithmetic mean and geometric mean can be employed for computations.

(iv) **Appropriate Use of Weights** – When selecting one item from different sub-groups, the values of each sub-group can be assigned as weights, making the weighted average of relatives method the most suitable.

(v) **Adaptability** – The method allows for easy replacement of old commodities with new ones, making it more flexible.

(vi) **Direct Representation of Index Values** – The price or quantity relatives for individual items directly function as index numbers in this method.

13.3.1.4 Demerits of Weighted Index Number

(i) **Complex Calculations** – The computation process becomes cumbersome when the geometric mean is used.

(ii) **Lack of a Standardized Weighting Method** – There is no fixed rule for assigning weights, which may lead to inconsistencies.

(iii) **Fails Key Consistency Tests** – This method does not satisfy the time reversal and factor reversal tests, which are essential for ensuring the consistency of an index formula.

Self-Check Exercise 13.1

Q1. Explain the Laspeyre's formula for computing index number

Q2. What is Fisher's Index Number? Why is it called ideal?

Q3. Write down the merits and demerits of weighted index number

13.4 TESTS OF INDEX NUMBERS

Up to this point, we have explored various formulas used in constructing index numbers. However, none of these formulas can measure price or quantity changes with absolute precision, as each carries some degree of bias. The challenge lies in selecting the most suitable formula for a given scenario. To assess the accuracy of different formulas, several mathematical tests—collectively known as tests of consistency for index number formulas—have been proposed. This section focuses on these tests, which are also referred to as the criteria for determining a reliable index number. Among these, the key tests include the Unit Test, Time Reversal Test, Factor Reversal Test, and Circular Test. However, our discussion will be limited to the Time Reversal Test and the Factor Reversal Test.

13.4.1 Time Reversal Test

According to Professor Fisher, "The test requires that the formula used to compute an index number should yield the same ratio between two points of comparison, regardless of which is taken as the base. In other words, the index number calculated forward should be the reciprocal of the one calculated in reverse."

This means that if an index number is computed using the same formula for two different time periods, but with the base years swapped, the resulting index values should be reciprocals of each other. Mathematically, the Time Reversal Test is satisfied if:

$$P_{01} = \frac{1}{P_{10}} \text{ or } P_{01} \times P_{10} = 1$$

Where P_{01} is the price index for the year 1 with 0 as base and P_{10} is the price index for the year 0 with 1 as base.

The Time Reversal Test is successfully met by Fisher’s Ideal Index, the Marshall-Edgeworth Index, and the Simple Geometric Mean of Price Relatives. However, it is not satisfied by the Simple Arithmetic Mean of Price Relatives, as well as Laspeyres’ and Paasche’s formulas.

(i) Laspeyre’s Formula:

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad \text{……… (i)} \qquad \text{(omitting factor 100)}$$

Interchanging time subscripts, i.e., 0 to 1 and 1 to 0

$$P_{10} = \frac{\sum p_0 q_1}{\sum p_1 q_1} \quad \text{……..(ii)}$$

Multiplying (i) and (ii), we get

$$P_{01} \text{ X } P_{10} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \text{ X } \frac{\sum p_0 q_1}{\sum p_1 q_1} \neq 1$$

Since P_{01} X $P_{10} \neq 1$, the Laspeyre’s formula does not satisfy Time Reversal Test.

(ii) Paasche's Formula:

$$P_{01} = \frac{\sum p_1\, q_1}{\sum p_0\, q_1} \quad \text{……….(i)} \qquad \text{(omitting factor 100)}$$

Interchanging time subscripts, i.e., 0 to 1 and 1 to 0

$$P_{10} = \frac{\sum p_0\, q_0}{\sum p_1\, q_0} \qquad \text{……………(ii)}$$

Multiplying (i) and (ii), we get

$$P_{01} \text{ X } P_{10} = \frac{\sum p_1\, q_1}{\sum p_0\, q_1} \text{ X } \frac{\sum p_0\, q_0}{\sum p_1\, q_0} \neq 1$$

Since P_{01} X $P_{10} \neq 1$, the Paasche's formula does not satisfy Time Reversal Test.

(iii) **Fisher's Formula**:

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \quad \textbf{……………..(i)}$$

Interchanging 0 to 1 and 1 to 0

$$P_{01} = \sqrt{\frac{\sum p_0 q_1}{\sum p_1 q_1} \times \frac{\sum p_0 q_0}{\sum p_1 q_0}} \quad \textbf{……………….(ii)}$$

Multiplying (i) and (ii), we get

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

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

$$= \sqrt{\mathbf{1}} = \mathbf{1}$$

Since, $P_{01} \times P_{01} = 1$, the Fisher's Formula satisfies Time Reversal Test.

(iv) Marshall-Edgeworth Formula:

$$P_{01} = \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \qquad \text{……………..(i)}$$

Interchanging 0 to 1 and 1 to 0

$$P_{10} = \frac{\sum p_0 q_1 + \sum p_0 q_0}{\sum p_1 q_1 + \sum p_1 q_0} \qquad \text{………………(ii)}$$

Multiplying (i) and (ii), we get

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

Since, $P_{01} \times P_{01} = 1$, the Marshall-Edgeworth Formula satisfies Time Reversal Test.

13.4.2 Factor Reversal Test

This is another test introduced by Prof. Irving Fisher. According to him, "Just as our formula should allow the interchange of two time periods without leading to inconsistencies, it should also permit swapping prices and quantities without producing conflicting results. In other words, the product of the two results should yield the true value ratio, apart from a proportional constant."

This means that when price and quantity indices are calculated using the same dataset, base period, and formula, their product (excluding the factor 100) should represent the true value ratio, as total value is derived by multiplying price by quantity. The Factor Reversal Test is considered valid when:

Price Index × Quantity Index= Value Index

Or

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

Similar to the Time Reversal Test, the Factor Reversal Test is not satisfied by Laspeyres', Paasche's, or Marshall-Edgeworth's formulas. However, it holds true for Fisher's Ideal Formula.

(i) **Laspeyre's Formula**:

$$P_{01} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad \text{......... (i)} \qquad \text{(omitting factor 100)}$$

Interchanging factor subscripts, i.e., p to q and q to p

$$Q_{01} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \quad \text{........(ii)}$$

Multiplying (i) and (ii), we get

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

Since $P_{01} \times Q_{01} \neq \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Laspeyre's formula does not satisfy Factor Reversal Test.

(ii) Paasche's Formula:

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad \text{..................(i)} \qquad \text{(omitting factor 100)}$$

Interchanging factor subscripts, i.e., p to q and q to p

$$Q_{01} = \frac{\sum q_1 p_1}{\sum q_0 p_1} \quad \ldots\ldots\ldots\ldots\ldots(ii)$$

Multiplying (i) and (ii), we get

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

Since $P_{01} \times Q_{01} \neq \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Paasche's formula does not satisfy Factor Reversal Test.

(iii) Fisher's Formula:

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(i)$$

Interchanging factor subscripts, i.e., p to q and q to p

$$Q_{01} = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(ii)$$

Multiplying (i) and (ii), we get

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

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

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

Since, Since $P_{01} \times Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Fisher's Formula satisfies Factor Reversal Test.

(iv) Marshall-Edgeworth Formula:

$$P_{01} = \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum p_0 q_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots(i)$$

Interchanging factor subscripts, i.e., p to q and q to p

$$Q_{01} = \frac{\sum q_1 p_0 + \sum q_1 p_1}{\sum q_0 p_0 + \sum q_0 p_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots(ii)$$

Multiplying (i) and (ii), we get

$$P_{01} \times Q_{01} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_0 p_0 + \Sigma q_0 p_1} \neq \frac{\Sigma p_1\, q_1}{\Sigma p_0\, q_0}$$

Since, $P_{01} \times Q_{01} \neq \frac{\Sigma p_1\, q_1}{\Sigma p_0\, q_0}$, the Marshall-Edgeworth Formula does not satisfy Factor Reversal Test.

Example 3: Calculate Laspeyre's, Paasche's and Fisher's price index numbers for the following data:

Commodity	2005		2015	
	Price	Expenditure	Price	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

Which of them satisfies Time Reversal and Factor Reversal Test?

Solution: Since we are given the expenditure and price, we can obtain quantity figure by dividing total expenditure by price for each commodity.

Commodity	p_0	q_0	p_1	q_1	$p_0\, q_0$	$p_0 q_1$	$p_1 q_0$	$p_1 q_1$
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	42	60
E	20	5	25	6	100	120	125	150
					396	**426**	**451**	**476**

(i) Laspeyre's Formula:

$$P_{01} = \frac{\Sigma p_1\, q_0}{\Sigma p_0\, q_0}$$

$$= \frac{451}{396} \times 100 = 113.88$$

Time Reversal Test

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

L.H.S. $= P_{01} \times P_{10}$

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

$$=\frac{451}{396} \times \frac{426}{476} \neq 1 = \text{R.H.S}$$

Since P_{01} X $P_{10} \neq 1$, the Laspeyre's formula does not satisfy Time Reversal Test.

Factor Reversal Test

$$P_{01} \text{ X } Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

L.H.S = P_{01} X Q_{01}

$$= \frac{\sum p_1 q_0}{\sum p_0 q_0} \text{ X } \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

$$=\frac{451}{396} \times \frac{426}{396} \neq \frac{476}{396} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Since P_{01} X $Q_{01} \neq \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Laspeyre's formula does not satisfy Factor Reversal Test.

(ii) Paasche's Formula

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

$$=\frac{476}{426} \text{ x } 100 = 111.74$$

Time Reversal Test

P_{01} X P_{10}= 1

L.H.S. = P_{01} X P_{10}

$$= \frac{\sum p_1 q_1}{\sum p_0 q_1} \text{ X } \frac{\sum p_0 q_0}{\sum p_1 q_0}$$

$$= \frac{476}{426} \times \frac{396}{451} \neq 1 = \text{R.H.S}$$

Since P_{01} X $P_{10} \neq 1$, the Paasche's formula does not satisfy Time Reversal Test.

Factor Reversal Test

$P_{01} \text{ X } Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$

L.H.S = $P_{01} \text{ X } Q_{01}$

$$= \frac{\sum p_1 q_1}{\sum p_0 q_1} \text{ X } \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

$$= \frac{476}{426} \times \frac{476}{451} \neq \frac{476}{396} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Since $P_{01} \text{ X } Q_{01} \neq \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Paasche's formula does not satisfy Factor Reversal Test.

(v) Fisher's Formula:

$$P_{01} = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times \mathbf{100}$$

$$= \sqrt{\frac{451}{396} \times \frac{476}{426}} \times \mathbf{100} = \mathbf{112.80}$$

Time Reversal Test

$P_{01} \text{ X } P_{10} = 1$

L.H.S. $= P_{01} \text{ X } P_{10}$

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

$$= \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{476} \times \frac{396}{451}}$$

$$= \sqrt{\mathbf{1}} = \mathbf{1}$$

Since, $P_{01} \times P_{01} = 1$, the Fisher's Formula satisfies Time Reversal Test.

Factor Reversal Test

$P_{01} \text{ X } Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$

L.H.S = P_{01} X Q_{01}

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

$$= \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{396} \times \frac{476}{451}} = \sqrt{\frac{476}{396} \times \frac{476}{396}}$$

$$= \frac{476}{396} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Since, P_{01} X $Q_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$, the Fisher's Formula satisfies Factor Reversal Test.

Self-Check Exercise 13.2

Q1. Explain Time Reversal Test.

Q2. State the 'Factor Reversal Test'. Show that Fisher's method for constructing index number satisfies it.

Q3 Calculate the Fisher's Ideal Index from the following data. Does it satisfy the Time Reversal And Factor Reversal Tests.

Items	Base Year		Current Year	
	Quantity	Price	Quantity	Price
A	15	4	10	6
B	20	3	25	4
C	10	6	20	5
D	30	5	25	5

Ans (109.4**)**

13.5 SUMMARY

In this unit, we explored various methods for computing price and quantity indices, focusing on three widely used formulae: Laspeyres' Index, Paasche's Index, and Fisher's Ideal Index. Additionally, we examined the fundamental properties and tests associated with index numbers, specifically the Time Reversal Test and the Factor Reversal Test. These tests help assess the consistency and reliability of index numbers when measuring changes in prices and quantities over time. Through this study, we gained a deeper understanding of how different index formulas function and how they contribute to economic and statistical analysis.

13.6 GLOSSARY

- **Index Number**: A statistical tool used to measure variations in price, quantity, or value of a single item or a group of related items over time, across locations, or based on other factors.
- **Base Year**: A reference year that is considered standard for comparison, typically chosen as a normal year concerning the variable being analyzed. The index for the base year is always set at 100, and the index for the current year is expressed as a percentage relative to it.
- **Price Index Number**: A metric used to track changes in price levels over a specific period, essentially serving as an indicator of inflation.
- **Value Index Number**: This index evaluates fluctuations in the total value of a particular commodity or a group of commodities consumed or purchased during a given period, compared to a base period.
- **Time Reversal Test**: A principle stating that if the prices and quantities of two compared periods are swapped, the resulting price index should be the reciprocal of the original index.
- **Factor Reversal Test**: A condition that asserts the product of a price index and a volume index of the same type should correspond to the proportional change in the total value for the given period.

13.7 ANSWERS TO SELF CHECK EXERCISE

Self-Check Exercise 13.1

Ans. Q1. Refer to Section 13.3.1.1 (a)

Ans. Q2. Refer to Section 13.3.1.1 (C)

Ans. Q3. Refer to Sections 13.3.1.3 and 13.3.1.4

Self-Check Exercise 13.2

Ans. Q1. Refer to Section 13.4.1

Ans. Q2. Refer to Section 13.4.2

Ans. Q3. Refer to Section 13.4.2 (Example 3)

13.8 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.

- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum's Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor, V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.
- Gupta S.C. (2016). Fundamentals of Statistics, Himalaya Publishing House.

13.9 TERMINAL QUESTIONS

Q1. Verify whether Laspeyre's Price Index Number and Paasche's Price Index numbers satisfy the time reversal test and Factor Reversal Test.

Q2. Calculate Index Number by (i) Laspeyre's method, (ii)Paasche's method, and (iii) Fisher method from the following:

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

Which of them satisfies Time Reversal and Factor Reversal Test?

Ans: (i) 118.82; (ii) 112.76; (iii) 115.75

SPECIAL PURPOSE INDEX NUMBERS

STRUCTURE

14.1 Introduction
14.2 Learning Objectives
14.3 Chain Base Index Numbers
    Self-Check Exercise 14.1
14.4 Base Shifting
    Self-Check Exercise 14.2
14.5 Splicing of Index Number
    Self-Check Exercise 14.3
14.6 Deflating of Index Number
    Self-Check Exercise 14.4
14.7 Special Purpose Indices
    14.7.1 Cost of Living Index
        14.7.1.1 Aggregate Expenditure Method
        14.7.1.2 Family Budget.
    14.7.2 Wholesale Price Index
    Self-Check Exercise 14.5
14.8 Summary
14.9 Glossary
14.10 Answers to Self-Check Exercise
14.11 References/Suggested Readings
14.12 Terminal Questions

14.1 INTRODUCTION

In the previous units, we explored the concept of index numbers, their significance, and the various purposes they serve in economic and statistical analysis. We also examined the challenges encountered in their construction, the different methodologies used to develop them, and the essential tests that determine the reliability of an ideal index number.

While constructing index numbers, several specific issues may arise, including the formulation of chain indices, shifting the base year, deflating index values to adjust for inflation, and developing the consumer price index (CPI) to measure changes in the cost of living. In this unit, we will delve deeper into these specific challenges, discussing their implications and the approaches used to address them effectively.

14.2 LEARNING OBJECTIVES

By the end of this unit, you will be able to:

- Comprehend the concepts of chain base index, base shifting, splicing, and deflating.
- Calculate the Cost of Living Index, also known as the Consumer Price Index.

14.3 CHAIN BASE INDEX NUMBERS

In the fixed base method, the base year remains constant and unchanged throughout the series. However, over time, certain items may be added or removed from the series, making it difficult to compare current data with past periods accurately. In such cases, updating the base period becomes a more suitable approach.

The chain index number method addresses this issue by expressing each year's figures as a percentage of the preceding year, known as Link Relatives. These values are then linked through successive multiplication to create a chain index. Unlike the fixed base method, the base year in this approach changes annually. For instance, the base year for 2001 would be 2000, for 2002 it would be 2001, and so forth.

Steps in Construction of Chain Base Index

(i) First of all, link relative are computed using the following formula:

$$\text{Link Relative} = \frac{Current\ Year's\ Price}{Previous\ Year's Price} \times 100$$

(ii) The link relatives are then converted into chain base index using the following formula:

$$\text{Chain Base Index} = \frac{Link\ Relatives\ of\ Current\ Year\ \times Chain\ Index\ of\ Previous\ Year}{100}$$

Example 1: Construct the chain base index from the following data

Year	2005	2006	2007	2008	2009	2010
Prices	94	98	102	95	98	100

Solution:

Year	Prices	Link Relatives	Chain Base Index
2005	94	100	100
2006	98	$\frac{98}{94}\times 100 = 104.3$	$\frac{104.3\times 100}{100} = 104.3$
2007	102	$\frac{102}{98}\times 100 = 104.1$	$\frac{104.1\times 104.3}{100} = 108.6$
2008	95	$\frac{95}{102}\times 100 = 93.1$	$\frac{93.1\times 108.6}{100} = 101.1$
2009	98	$\frac{98}{95}\times 100 = 103.2$	$\frac{103.2\times 101.1}{100} = 104.3$
2010	100	$\frac{100}{98}\times 100 = 102$	$\frac{102\times 104.3}{100} = 106.4$

Self-Check Exercise 14.1

Q1. What is Chain Base Index Number.

Q2. What are the Steps involved in construction of Chain Base Index Number

14.4 BASE SHIFTING

Base shifting refers to the process of updating the base period (year) of a given index number series and recalculating it using a more recent base year. This adjustment is necessary in the following situations:

(i) When the existing base year is outdated.

(ii) When different index series need to be compared, but their base years vary.

During base shifting, the index number of the newly chosen base year is set to 100, and the index numbers of all previous years are recalculated accordingly.

Recast Index Number of any Year

$$= \frac{Old\ Index\ Number\ of\ the\ Year}{Index\ Number\ of\ the\ New\ Base\ Year} \text{ X } 100$$

Example 2: The following are the index numbers of prices (1986=100)

Year	**Index**	**Year**	**Index**
1986	100	1991	410
1987	110	1992	400
1988	120	1993	380
1989	200	1994	370
1990	400	1995	340

Shift the base from 1986 to 1992 and recast the index numbers.

Solution:

Year	Index numbers (1986=100)	Index Numbers (1992=100)	Year	Index numbers (1986=100)	Index Numbers (1992=100)
1986	100	$\frac{100}{400}$x100=25	1991	410	$\frac{410}{400}$x100=102.5
1987	110	$\frac{110}{400}$x100=27.5	1992	400	$\frac{400}{400}$x100=100
1988	120	$\frac{120}{400}$x100=30	1993	380	$\frac{380}{400}$x100=95
1989	200	$\frac{200}{400}$x100=50	1994	370	$\frac{370}{400}$x100=92.5
1990	400	$\frac{400}{400}$x100=100	1995	340	$\frac{340}{400}$x100=85

Self-Check Exercise 14.2

Q1. What is meant by base shifting in index number.

14.5 SPLICING OF INDEX NUMBER

At times, an index number series is available for a specific period but later undergoes significant revisions, including changes in the reference period. The technique used to address this is known as splicing, which involves merging two or more overlapping index number series to create a single continuous series. Maintaining this continuity is essential for facilitating meaningful comparisons. Splicing can be carried out using two different methods.

(i) Splicing of new index series to old index series: The following formula is used

Spliced new index series to old index series $= \frac{New\ Index \times Old\ index\ of\ overlapping\ year}{100}$

(ii) Splicing of old index series to new index series: The following formula is used

Spliced old index series with new index series =Old index X $\frac{100}{\text{Old Index of Overlapping year}}$

Example 3: Given below are two price index series. Splice them on the base 2001 = 100

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Index A (2001=100)	100	110	120	125	150	---	---	----	----	----
Index B (2005 = 100)	---	---	----	----	100	115	120	130	140	160

Solution:

Year	**Index A (2001=100)**	**Index B (2005= 100)**	**Spliced index (2001 = 100)**
2001	100	----	100
2002	110	----	110
2003	120	----	120
2004	125	----	125
2005	150	100	$\frac{150}{100}$x100=150
2006	----	115	$\frac{150}{100}$x115=172.5

2007	----	120	$\frac{150}{100}$x120=180
2008	----	130	$\frac{150}{100}$x130=195
2009	----	140	$\frac{150}{100}$x140=210
2010	----	160	$\frac{150}{100}$x160=240

Self-Check Exercise 14.3

Q1. What is meant by splicing of Index Number

14.6 DEFLATING OF INDEX NUMBER

Deflating refers to the process of adjusting or correcting an inflated value. In the context of price index numbers, deflating involves modifying them to account for changes in price levels. This adjustment is particularly important in economies experiencing inflation, where rising prices over time lead to a decline in real income—defined as the purchasing power of money. In such cases, an increase in nominal income (or money income) does not necessarily indicate an actual rise in real income. Therefore, it becomes essential to adjust nominal wages in relation to the corresponding price index to determine real income accurately. The purchasing power is represented by the reciprocal of the price index, and real income (or wages) is calculated by dividing nominal income by the appropriate price index and multiplying the result by 100.

Symbolically,

Real Wages = $\frac{Money\ or\ Nominal\ Wages}{Price\ Index}$ X 100

Real income, also referred to as deflated income, represents income adjusted for inflation. The corresponding real wages can be expressed in the form of index numbers using the following method:

Real Wages Index Number = $\frac{Real\ Wages\ of\ Current\ Year}{Real\ Wage\ of\ Base\ Year}$ X 100

Real Income Index Number = $\frac{Real\ Income\ of\ Current\ Year}{Real\ Wage\ of\ base\ year}$ X100

Example 4: The following data related to the people and the general price index number. Calculate (i) Real Wages (ii) Index Number of Real Wages with 1980 as base.

Year	Wage (in Rs.)	Price Index
1980	800	100
1981	819	105
1982	825	110
1983	876	120
1984	920	125
1985	938	140
1986	924	140

Solution:

Year	**Wage (in Rs.)**	**Price Index**	**Real Wage**	**Real Wage index number (1980=base year)**
1980	800	100	$\frac{800}{100}$x100=800	**100**
1981	819	105	$\frac{819}{105}$x100=780	$\frac{780}{800}$x100=97.5
1982	825	110	$\frac{825}{110}$x100= 750	$\frac{750}{800}$x100=93.5
1983	876	120	$\frac{876}{120}$x100=730	$\frac{730}{800}$x100=91.25
1984	920	125	$\frac{920}{125}$x100=736	$\frac{736}{800}$x100=92
1985	938	140	$\frac{938}{140}$x100=670	$\frac{670}{800}$x100=83.75
1986	924	140	$\frac{924}{140}$x100=660	$\frac{660}{800}$x100=82.5

Self-Check Exercise 14.4

Q1. What is meant by deflating of index number.

14.7 SPECIAL PURPOSE INDICES

14.7.1 Cost of Living Index

The **Cost of Living Index**, also referred to as the **Consumer Price Index (CPI), Retail Price Index,** or **Price of Living Index Numbers**, is formulated to assess the impact of price fluctuations in a specific set of goods and services on the purchasing power of a particular demographic over a given time period (current) in comparison to a predetermined reference period (base). Since price changes influence different groups in diverse ways, the index accounts for variations in consumption patterns based on factors such as social class, income level, and occupation. For instance, government employees, industrial workers, low-income households, high-income groups, laborers, and agricultural workers all exhibit distinct spending behaviors. The cost price index thus serves as a crucial tool in evaluating how price increases or decreases affect various consumer groups across different regions.

Utility of the Cost Price Indices

1. **Regulation of Wages and Allowances:** It plays a key role in determining dearness allowance and formulating bonus policies.
2. **Assessing Purchasing Power and Real Wages:** The index helps measure fluctuations in the purchasing power of money and the real value of wages or income.
3. **Policy Formulation:** Governments rely on it to shape critical policies, including income policy, wage policy, and price regulations.
4. **Market Analysis:** It aids in studying market trends for specific goods and services, helping businesses and policymakers make informed decisions.

Methods of Constructing Cost of Living index

Cost of Living Index can be calculated by using any of the following two methods:

14.7.1.1 Aggregate Expenditure Method (or Weighted Aggregate Method)

In this method, the quantities of various commodities consumed by a particular class of people in the base year are taken as weights. Thus, in usual notations.

Cost of Living Index $=\frac{\sum p1q0}{\sum p0q0}$ X 100 $= \frac{Total\ Expenditure\ in\ the\ Current\ year}{Total\ Expenditure\ in\ the\ Base\ Year}$ X 100

This is in fact the Laspeyre's Price Index. This method is most popular for constructing cost of living index.

14.7.1.2 Family Budget (or Method of Weighted Average of Price Relatives)

The cost of living index using the family budget method is determined by first computing the price relatives for individual commodities. These price relatives (P) are then multiplied by their respective value weights (W). The sum of these products is subsequently divided by the total of all weights. Mathematically, it is expressed as:

Cost of Living Index $=\frac{\sum WP}{\sum W}$, where P= Price Relative $=\frac{P1}{P0}$ X 100

And $W=p_0q_0$ = Value weight being the aggregate expenditure of an average family on various commodities in the base year.

Example 5: Construct cost of living Index Number of 2010 on the basis of 2000 by (i) Aggregate Expenditure Method, and (ii) Family Budget Method:

Items	Quantity Consumed in 2000	Unit	Price in 2000	Price in 2010
Wheat	2 Qtls.	Qtls.	50	100
Rice	1 Qtls.	Qtls.	80	110
Arhar	20 Qtls.	Kg	1.20	2.80
Sugar	0.5 Qtls.	Kg	2.0	3.00
Salt	10 Kg	Qtls.	20	30
Oil	10 Kg	Kg	4	8
Clothing	20 Meter	Meter	3	5
Fuel	4 Qtls.	Qtls.	12	15
Rent	1 House	House	50	75

Solution: Since the unit of price and quantity of two items such as sugar and salt are different, we should convert the unit of quantity in to of price before we apply any method:

(i) Construct Cost of Living Index by Aggregate Expenditure Method

Items	Quantity Consumed in 2000 (q_0)	Unit	Price in 2000 (p_0)	Price in 2010 (p_1)	p_0q_0	p_1q_1
Wheat	2 Qtls.	Qtls.	50	100	100	200
Rice	1 Qtls.	Qtls.	80	110	80	110
Arhar	20 Qtls.	Kg	1.20	2.80	24	56
Sugar	50 Kg	Kg	2.0	3.00	100	150
Salt	0.1Qtls.	Qtls.	20	30	2	3
Oil	10 Kg	Kg	4	8	40	80
Clothing	20 Meter	Meter	3	5	60	100
Fuel	4 Qtls.	Qtls.	12	15	48	60
Rent	1 House	House	50	75	50	75
					$\sum p0q0 = 504$	$\sum p1q1 = 834$

(ii) Construct Cost of Living Index by Family Budget Method

Items	Quantity Consumed in 2000 (q_0)	Unit	Price in 2000 (p_0)	Price in 2010 (p_1)	Price Relatives P = $\frac{P1}{P0} \times 100$	$W=p_0q_0$	PW
Wheat	2 Qtls.	Qtls.	50	100	200	100	20000
Rice	1 Qtls.	Qtls.	80	110	137.5	80	11000
Arhar	20 Qtls.	Kg	1.20	2.80	233.33	24	5600
Sugar	50 Kg	Kg	2.0	3.00	150.00	100	15000
Salt	0.1Qtls.	Qtls.	20	30	150.00	2	300
Oil	10 Kg	Kg	4	8	200	40	8000
Clothing	20 Meter	Meter	3	5	166.67	60	10000
Fuel	4 Qtls.	Qtls.	12	15	125	48	6000
Rent	1 House	House	50	75	150	50	7500
						$\sum \mathbf{W} =$ **504**	$\sum \mathbf{p1q1} =$ **83400**

Cost of Living Index Number (P_{01}) $= \frac{\sum PW}{\sum W}$

$$= \frac{83400}{504} = 165.50$$

14.7.2 Wholesale Price Index

The Wholesale Price Index (WPI) reflects the price of goods at the wholesale level, meaning goods sold in bulk and traded between businesses rather than directly to consumers. It serves as an indicator of price fluctuations in goods before they reach the retail stage. In simpler terms, WPI tracks changes in the prices set by manufacturers and wholesalers. These price variations may occur at different stages, including transactions between manufacturers and wholesalers or wholesalers and retailers, or a combination of both.

WPI is a crucial measure of inflation in several economies, including India, where it significantly influences fiscal and monetary policy decisions. It provides a straightforward method for assessing inflation trends. The inflation rate is determined by calculating the difference in WPI between the start and end of a year. The percentage increase in WPI over this period represents the inflation rate for that year.

Self-Check Exercise 14.5

Q1. What is meant by Cost of Living Index?

Q2. Explain the methods of construction of Cost of Living Index.

Q3. What is meant by Wholesale Price Index

14.8 SUMMARY

This unit has provided an in-depth exploration of various challenges associated with index numbers. Some of the key issues covered include the process of constructing a chain index, the technique of base shifting, the concept and application of deflating, as well as the methodology involved in developing a consumer price index. Each of these topics plays a crucial role in the accurate measurement and analysis of economic and financial data, ensuring meaningful comparisons across different time periods.

14.9 GLOSSARY

- **Chain Index**: An index number that expresses the value of a specific period in relation to the value of the immediately preceding period.
- **Splicing**: A method used to merge two overlapping series of index numbers, creating a continuous and extended series.
- **Deflating**: The process of adjusting or reducing an inflated value. In the context of price index numbers, deflating involves modifying them to account for changes in price levels over time.
- **Cost of Living Index**: Also referred to as the Consumer Price Index, Retail Price Index, or Price of Living Index, this measure evaluates how price fluctuations in a selected set of goods and services impact the purchasing power of a specific group of people during a given period, compared to a fixed base period. Since price changes affect different social groups in varied ways, this index helps assess their economic impact accordingly.

14.10 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise 14.1

Ans. Q1. Refer to Section 14.3

Ans. Q2. Refer to Section 14.3

Self-Check Exercise 14.2

Ans. Q1. Refer to Section 14.4

Self-Check Exercise 14.3

Ans. Q1. Refer to Section 14.4

Self-Check Exercise 14.4

Ans. Q1. Refer to Section 14.6

Self-Check Exercise 14.5

Ans. Q1. Refer to Section 14.7.1

Ans. Q2. Refer to Section 14.7.1.1 and 14.7.1.2

Ans. Q2. Refer to Section 14.7.2

14.11 REFERENCES/SUGGESTED READINGS

- Gupta, S.P. (2018). Statistical Methods, Sultan Chand & Sons, New Delhi.
- Elhance, D.N., Elhance, V. and Aggarwal, B.M. (2018). Fundamentals of Statistics, Kitab Mahal, New Delhi.
- Lind D.A., Marchal, W.G. and Wathen, S.A. (2017). Statistical Techniques in Business and Economics, Tata McGraw Hill, New Delhi.
- Gun, A.M., Gupta, M.K., Dasgupta, B. (1999). Fundamentals of Statistics, Vol.II, World Press, Calcutta.
- Spiegel, M.R. (1967). Theory & Problems of Statistics, Schaum’s Publishing Series.
- Croxton, F.E., Cowden, D.J. and Kelin, S. (1973). Applied General Statistics, Prentice Hall of India.
- Kapoor, V.K. (2000). Statistics Problems and Solutions, Sultan Chand and Sons. New Delhi.
- Gupta S.C. (2016). Fundamentals of Statistics, Himalaya Publishing House.

14.12 TERMINAL QUESTIONS

Q1. What is meant by consumer price index ? Explain its importance and the methods of construction of consumer price index.

Q2. Construct cost of living Index Number of 2000 on the basis of 1999 by (i) Aggregate Expenditure Method, and (ii) Family Budget Method:

Commodity	**Rice**	**Wheat**	**Pulses**	**Ghee**	**Oil**
Weight	40	20	15	20	5
Price in 1999	16.00	4.00	0.50	5.12	2.00
Price in 2000	20.00	60.00	0.50	6.25	1.5

(i) 137.188 (ii) 123.15